Structural Determinants of Bunyavirus Entry and Neutralisation



Robert Stass

Division of Structural Biology

Nuffield Department of Medicine

Brasenose College

University of Oxford

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford

Trinity term 2019

Structural Determinants of Bunyavirus Entry and Neutralisation

Robert Stass, Brasenose College, University of Oxford
Doctor of Philosophy, Division of Structural Biology, Nuffield Department of Medicine.
Trinity term 2019.

Abstract

The order *Bunyavirales* is comprised of numerous emerging zoonotic pathogens which present a global threat to human health. Many bunyaviruses are arboviruses including members of the family *Nairovirdae* such as the tick-borne Crimean-Congo haemorrhagic fever virus that is responsible for deadly outbreaks in Africa, East Asia and Southern Europe. Also among them are members of the family *Hantaviridae* which include Andes and Sin Nombre virus in the American continent and Hantaan and Puumala virus in Asia and Europe respectively. These are a hosted predominantly by rodents and frequently spill over into humans populations leading to hantavirus pulmonary syndrome and haemorrhagic fever with renal syndrome.

Bunyaviruses are often pleomorphic and are enveloped in a membrane harbouring just two surface glycoproteins, Gn and Gc. These glycoproteins are responsible for budding of the virus from the host cell, attachment to new target cells and release of viral RNA by membrane fusion. Furthermore, these glycoproteins are the target of the antibody-mediated host immune response and could form the basis for the development of future vaccines and immunotherapeutic medical interventions. The precise arrangement of these glycoproteins, and how these processes are mediated, are only partially understood for members of *Hantaviridae* and *Nairoviridae*. The goal of this thesis is to contribute to our understanding by addressing these topics using cryogenic electron microscopy (cryo-EM).

An initial description of the nairovirus surface is presented which displays a novel lattice of glycoproteins across its surface. An improvement to our understanding of the hantavirus glycoprotein array is made which enables, in combination with available crystallographic models, a near-complete pseudo-atomic model of the hantavirus surface to be proposed. In addition, the metastable nature of this lattice is demonstrated after exposure to low pH conditions equivalent of the endocytic pathway. A hantaviral virus-like particle (VLP) system is developed to provide a safe platform for structural studies of otherwise dangerous pathogenic species. This is then used to gain insights into hantavirus receptor interactions including Andes virus and the cell surface protein protocadherin-1. The VLP system is also used to study the structural mechanism of antibody-mediated host immune responses to Puumala virus and Hantaan virus. Finally, a number of tools and optimisations to assist and improve the cryo-ET and subtomogram averaging workflows that enable these structural studies are presented.

Word count: ~35,000 excluding abstract, table of contents, figures, references and appendix

Acknowledgments

There are so many people to thank that helped make this DPhil possible. Firstly, I'd like to thank my excellent supervisors Juha Huiskonen and Thomas Bowden who both gave me the freedom to explore and always made themselves available when help was needed. I'd like to give a special thank you to Ilona Rissanen who effectively became my unofficial third supervisor. Ilona not only provided many protein samples and crystallographic models that played a pivotal role in the project, but was also a constant source of advice and support throughout, and has been sorely missed since the move to Helsinki.

I'd like to thank Sai Li, whose patient advice got me started in the field of subtomogram averaging and for moral support through many late nights at the microscope. I also owe lot to Steinar Halldorsson who taught me everything I know in the virus lab as well as providing a constant source of entertainment in the office. I would like to thank many other collaborators including Katie Doores, Jussi Hepojoki, Roger Hewson and Pablo Guardado-Calvo who all provided me with key materials and advice. In addition, I must thank a number of important people who help keep strubi running including; Margaret, Robin, Karl, Tom, Ollie, Dimple, Agata, Mel, Jun, Weixian, Yvonne and Dave.

This DPhil also wouldn't have been nearly as enjoyable without the hugely supportive community of fellow students who have been a great source of advice, conversation, comfort, and laughter throughout. They really helped to make the DPhil much more than just a string of long days in the lab and kept me going when times got tough. To name just a few, these include; Helen, Liz, Rhys, Alice, Vicky, Vojta, Weng, Serban, Max, Sam and Steinar.

Finally I would like to thank my family, and in particular my mother, Tina, my father, Derek and my sister Joanna, who's unwavering support got me to where I am today.

1 Table of Contents

List of Figures....7
List of Tables....8
List of Abbreviations....8
1 Introduction....10
1.1. Hantavirus....15
1.2. Nairovirus....21
1.3. Cryo-electron microscopy....28
1.4. Aims....31
2. Materials and Methods....33
2.1. Live virus growth and purification....33
2.1.1. Tula virus growth....33
2.1.2. Tula virus Purification....33
2.1.3. Tula virus grid preparation....34
2.1.4. Puumala virus growth....34
2.1.5. Hantaviral focus forming unit assay....34
2.1.6. Hazara virus growth....35
2.1.7. Hazara virus purification....36
2.1.8. Hazara virus grid preparation....36
2.1.9. Hazara virus plaque assay....36
2.2. Molecular cloning....37
2.2.1. Plasmids, cDNA and primers....37
2.2.2. Polymerase chain reaction....38
2.2.3. Agarose gel electrophoresis....38
2.2.4. Restriction digestion....38
2.2.5. Ligation of plasmid....39
2.2.6. Transformation of plasmid in E.coli....39
2.2.7. Large scale DNA production....39
2.2.8. Designing N-terminal His tag variants....40
2.3. Mammalian cell protein and VLP expression and purification....40
2.3.1. Tissue culture....40
2.3.2. Small-scale transient protein and VLP expression....40
2.3.3. Large-scale protein expression....41
2.3.4. Protein purification....41
2.3.5. Large scale hantaviral VLP expression....42
2.3.6. Hantaviral VLP purification....42

2.3.7. SDS-PAGE....43
2.3.8. Western and dot blotting....43
2.3.9. ELISA....44
2.3.10. VLP grid preparation....45
2.4. Cryo EM data acquisition....45
2.5. Cryo EM data processing....46
2.5.1. Tomgraphic data preprocessing and tilt series alignment....46
2.5.2. Subtomogram averaging of pleomorphic virus lattice....46
2.5.3. Tula virus pH comparison....47
3. Tools to improve subtomogram averaging....48
3.1. Introduction....48
3.2. TomoPreprocess – a tool for automated preprocessing of tomography data....51
3.3. Dose weighting....51
3.4. Dose weight sharpening....53
3.5. Subtomogram averaging of pleomorphic enveloped viruses and the PatchFinder postprocessing method....54
3.6. Investigating the resolution limits of subtomogram averaging; Subtomogram reconstruct (Subrec)....58
3.7. Discussion and future directions....71
4. Bunyavirus ultrastructure....73
4.1. Introduction....73
4.2. The surface glycoprotein array of *Nairoviridae*....76
4.2.1. Foreword....76
4.2.2. Cryo-ET reconstruction of the Hazara virus surface glycoprotein array....77
4.2.3. Hazara virus sample optimisation....78
4.2.4. Discussion and future directions....82
4.3. The surface glycoprotein array of *Hantaviridae*....83
4.3.1. Sample preparation....83
4.3.2. Handedness of the map....84
4.3.3. A pseudo-atomic model of the Hantavirus glycoprotein spike....85
4.4. Tula virus conformational changes induced by exposure to low pH....89
4.5. Limitations of Tula virus as a surrogate....90
4.6. Puumala live virus production....92
4.7. Hantaviral virus-like particle development....93
4.8. Hantaan VLP ultrastructure....100
4.9. Discussion and future directions....101
5. Hantavirus interactions with cell surface receptors....105
5.1. Introduction....105

5.2. Andes virus and Protocadherin-1 106
5.3. Puumala virus VLPs and αvβ3 integrin 112
5.4. Puumala virus VLPs and gc1qr 113
5.5. Discussion and future directions 114
6. The host immune response to hantaviruses 119
6.1. Introduction 119
6.2. Neutralisation of Puumala virus by 4G2 fab 120
6.3. Neutralisation of Hantaan virus by the novel fab C4 128
6.4. Discussion and future directions 132
7. References 137
8. Appendix 147
8.1. Sequences used 147
8.1.1. PUUV M segment construct 1 147
8.1.2. PUUV M segment construct 2 148
8.1.3. HNTV M segment construct 149
8.1.4. ANDV M segment construct 151
8.1.5. Protocadherin-1 construct 152
8.2. HNTV N-terminal histidine primer design 152
8.3. PUUV FFU assay data 153
8.4. Grid preparation details 153
8.5. Data acquisition parameters 154
8.6. Data processing details 155
8.7. Additional ELISA data 156
8.8. FSC curves 157
8.9. SDS-PAGE 159
8.10. Additional lattice quality data 161
8.11. Dose Weight Sharpening formula derivation 161

List of Figures

Figure 1.1 Phylogeny of negative strand RNA viruses 11
Figure 1.2 Three classes of viral fusion protein 13
Figure 1.3 Model for class II mediated membrane fusion 14
Figure 1.4 Phylogeny of Hantaviridae 16
Figure 1.5 The hantavirus lifecycle 19
Figure 1.6 Phylogeny of Nairoviridae 22
Figure 1.7 The nairovirus lifecycle 25
Figure 1.8 Processing of the nairovirus glycoprotein precursor 27
Figure 1.9 A simplified schematic of an electron microscope 29
Figure 2.1 The pHLSEC vector 38
Figure 3.1 The effects of dose-weight sharpening applied to TULV 54
Figure 3.2 The patch finder script for STA of the nairovirus glycoprotein array. 57
Figure 3.3 Subrec reconstruction of TULV 61
Figure 3.4 The impact of dose weighted data on resolution 62
Figure 3.5 The effect of electron damaged images on resolution 64
Figure 3.6 The effect of electron damaged data on map quality 65
Figure 3.7 3D CTF correction with Subrec 67
Figure 3.8 The effect of Wiener filtered CTF correction on resolution 68
Figure 3.9 The effect of Wiener filtering on the appearance of STA data 69
Figure 3.10 Reconstruction of unaligned subvolumes using Subrec 70
Figure 4.1 Bunyavirus glycoprotein cryo-EM 73
Figure 4.2 Hantavirus glycoprotein ectodomains studied by X-ray crystallography 75
Figure 4.3 The glycoprotein array of Hazara virus 78
Figure 4.4 Hazara virus titre optimisation 79
Figure 4.5 Cryo-EM sample optimisation of Hazara virus 81
Figure 4.6 Optimised TULV preparation for cryo-EM 83
Figure 4.7 The handedness of the TULV map differs from that reported in EMD-3364 84
Figure 4.8 Confirming the handedness of the Tula virus map 85
Figure 4.9 A pseudo-atomic model of the hantaviral surface glycoprotein array 86
Figure 4.10 The hantaviral lattice contains regions of exposed membrane 89
Figure 4.11 Metastability of the Tula virus envelope after exposure to low pH 90
Figure 4.12 The limitation of Tula virus as a surrogate for antibody experiments 91
Figure 4.13 The slow growth of PUUV in VeroE6 cells 93
Figure 4.14 A disordered-tubular form of PUUV observed with a VLP system 95
Figure 4.15 A curious ordered tube from a sample of PUUV VLPs 96
Figure 4.16 Continued development of ANDV and HNTV VLPs 97
Figure 4.17 Testing purification strategies with HNTV VLPs 98
Figure 4.18 Successful production of hantaviral VLP for investigation by cryo-EM 100
Figure 4.19 Pseudo-atomic model of HNTV obtained from a VLP system 101
Figure 4.20 Diverse morphology of class II fusion displaying viruses 104
Figure 5.1 Purification of recombinantly expressed PCDH1 EC1-EC2 107
Figure 5.2 ANDV pseudovirus neutralisation by PCDH1 EC1-EC2 108
Figure 5.3 ANDV VLPs refined with C4 symmetry with and without PCDH1 EC1-EC2 109
Figure 5.4 Putative ANDV VLP interactions with PCDH1 EC1-EC2 111
Figure 5.5 An attempt to observe an interaction with PUUV VLPs and αVβ3 integrin ectodomain 112
Figure 5.6 ELISA to demonstrate binding of gC1qR to PUUV VLPs 113
Figure 5.7 Additional density seen in PUUV VLPs in the presence of gC1qR 114

Figure 6.1 ELISA to demonstrate binding of 4G2 fab to PUUV VLPs 121
Figure 6.2 The interaction of 4G2 fab with PUUV VLPs 123
Figure 6.3 A demonstration of the lattice present on PUUV VLPs in the absence of fab 4G2 125
Figure 6.4 A demonstration of the lattice present on PUUV VLPs in the presence of fab 4G2 126
Figure 6.5 Comparison of the PUUV VLP lattice in the presence an absence of fab 4G2 127
Figure 6.6 Neutralisation of live Hantaan virus by antibody C4 128
Figure 6.7 A demonstration of the lattice present on HNTV VLPs in the presence of fab C4 130
Figure 6.8 The interaction of antibody C4 with HNTV VLPs 132
Figure 6.9 Common hantaviral immune epitopes and the 4G2 Ab mechanism of action 134
Figure 8.1 A positive control ELISA confirming the presence of TULV 156
Figure 8.2 An attempt to monitor binding of ANDV VLPs to PCDH1 EC1-EC2 by ELISA 156
Figure 8.3 ELISA to demonstrate binding of gC1qR to PUUV VLPs with negative control displayed 157
Figure 8.4 TULV FSC curve 157
Figure 8.5 HAZV FSC curve 158
Figure 8.6 HNTV VLP FSC curve 158
Figure 8.7 ANDV + PCDH1 EC1-EC2 FSC curve 158
Figure 8.8 PUUV VLP + gC1qR FSC curve 159
Figure 8.9 SDS PAGE of purified gC1qR 159
Figure 8.10 SDS PAGE of purified αVβ3 integrin ectodomain 160
Figure 8.11 A demonstration of the lattice present on HNTV VLPs in the presence of fab C4 at 21°C 161

List of Tables

Table 5.1 Hantavirus receptor candidates 106
Table 8.1 Primers to introduce an N terminal His tag to HNTV VLPs 153
Table 8.2 PUUV live virus FFU assay data 153
Table 8.3 Cryo-EM sample preparation details 154
Table 8.4 Cryo-EM data acquisition parameters 154
Table 8.5 Data processing details 155

List of Abbreviations

ANDV	Andes virus
BSL	Biological safety level
Cryo-EM	Cryogenic transmission electron microscopy
Cryo-ET	Cryogenic transmission electron tomography
DMEM	Dulbecco's modified eagle medium
DPI	Days post infection

ELISA	Enzyme-linked immunosorbent assay
FBS	Fetal bovine serum
FFU	Focus forming unit
HAZV	Hazara virus
HFRS	Hantavirus hemorrhagic fever with renal syndrome
HNTV	Hantaan virus
HPS	Hantavirus pulmonary syndrome
HRP	Horseradish peroxidase
IMAC	Immobilized metal affinity chromatography
MW	Molecular weight
PBS	Phosphate buffered saline
PCDH1	Protocadherin-1
PCR	Polymerase chain reaction
PEI	Polyethylenimine
PUUV	Puumala virus
RdRP	RNA-dependent RNA polymerase
RT	Room temperature
RVFV	Rift Valley fever virus
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SNR	Signal-to-noise ratio
SNV	Sin Nombre virus
SOC	Super Optimal broth with Catabolite repression
SPR	Single-particle refinement
STA	Subtomogram averaging
TNTC	Too numerous to count
TULV	Tula virus
VLP	Virus-like particle

1 Introduction

Members of the order *Bunyavirales* are negative sense single strand RNA viruses that are divided into a number of families including; *Phenuiviridae, Peribunyaviridae, Arenaviridae, Hantaviridae* and *Nairoviridae* [1], [2]. A phylogenetic tree is presented in Figure 1.1 to demonstrate how bunyaviruses are related to other negative sense RNA viruses. Bunyaviruses are distributed worldwide and are typically hosted by arthropod vectors, although members are also carried by a diverse range of species including rodents, bats, fish and plants [1], [3]. From these reservoir hosts, some bunyaviruses frequently spill over into livestock and human populations leading to often severe and occasionally fatal outbreaks[1], [4]. Some notable examples of bunyaviruses that cause disease in humans are Crimean-Congo haemorrhagic fever virus, Rift Valley fever virus, Severe fever with thrombocytopenia syndrome virus, Lassa fever virus, and Andes Virus. Bunyaviruses have a segmented genome, a membrane envelope, and unlike many other enveloped viruses, no matrix protein (with the exception of *Arenaviridae*) [1], [5]–[7]. Some members exhibit icosahedral symmetry such as Rift Valley Fever virus, defined by the two surface glycoproteins, Gn and Gc, while others are pleomorphic with locally ordered arrays of Gn and Gc at their surface [2], [8], [9]. This thesis focuses on structural studies of hantaviruses and nairoviruses which both include a number of medically important human pathogens [4], [10].

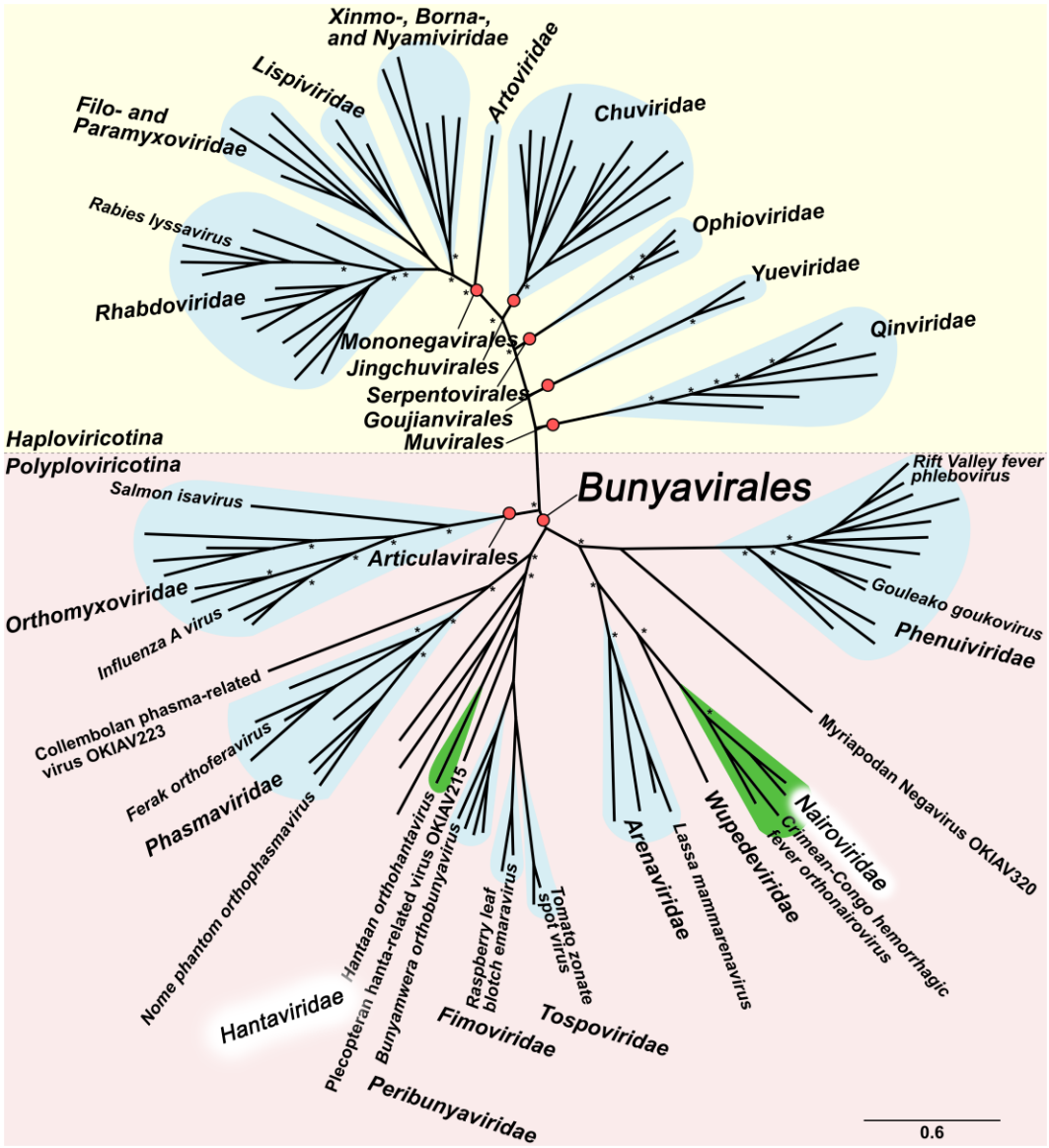


Figure 1.1 Phylogeny of negative strand RNA viruses
A phylogenetic analysis of 234 viral genome sequences computed with the MrBayes algorithm [11]. Red dots signify virus orders and blue/green clouds signify virus families. Selected virus species are displayed and the virus families studied in this thesis are highlighted and coloured green. Figure adapted from [12].

Like all enveloped animal viruses, the lifecycle of a bunyavirus includes attachment to a target cell via a cell surface receptor followed by membrane fusion with a host cell membrane to enable the release of the viral genome into the cell for replication. While some viruses fuse directly at the cell surface, many, including bunyaviruses, are first taken into the cell via endocytosis where fusion is required to escape an intracellular compartment [13]. The process of fusing two lipid bilayers is an energetically unfavourable process because it requires the exclusion of water from around the hydrophilic heads of the lipids as well as inducing curvature in the bilayer itself [14]. To overcome this energetic barrier, enveloped

viruses employ specialised membrane fusion proteins which undergo a transition from a pre-fusion conformation to a more stable post-fusion conformation [15]. It is unproductive for a virus to induce this conformational change until it is ideally situated to enter the target cell, so many viruses rely on a host specific mechanism to trigger the fusion protein. This can be induced by direct binding to a receptor protein as is the case for retroviruses, or via binding to a separate viral protein which intern binds to a cell surface receptor such as seen in paramyxoviruses. A common strategy used by many viruses, including bunyaviruses, is to use the low pH environment of the endocytic pathway to trigger membrane fusion [13], [16].

Despite, their varied evolutionary lineage, viral fusion proteins can be broadly classified into three classes according to their structural organisation (Figure 1.2). Class I fusion proteins have a characteristic α helical core and examples are found in orthomyxoviruses, paramyxoviruses, retroviruses, filoviruses, arenaviruses, and coronaviruses. Class II fusion proteins have a predominantly β sheet composition arranged in a classic three domain architecture and examples are found in flaviviruses, alphaviruses, rubiviruses and bunyaviruses. Finally the class III fusion proteins found in herpesviruses, baculoviruses and rhabdoviruses are comprised of a mixture of α helical and β sheet [13]. Despite their structural variability, there are common features found in all of these proteins. First, they each have hydrophobic loops termed *fusion loops* that embed into the target membrane. These are shielded from the surrounding solvent in the pre-fusion state either by being buried within the pre-fusion complex or shielded by an external chaperone protein or the viral membrane itself. Remarkably, despite existing in a number of different pre-fusion oligomeric states, all these proteins finish the fusion process as a post fusion trimer [13].

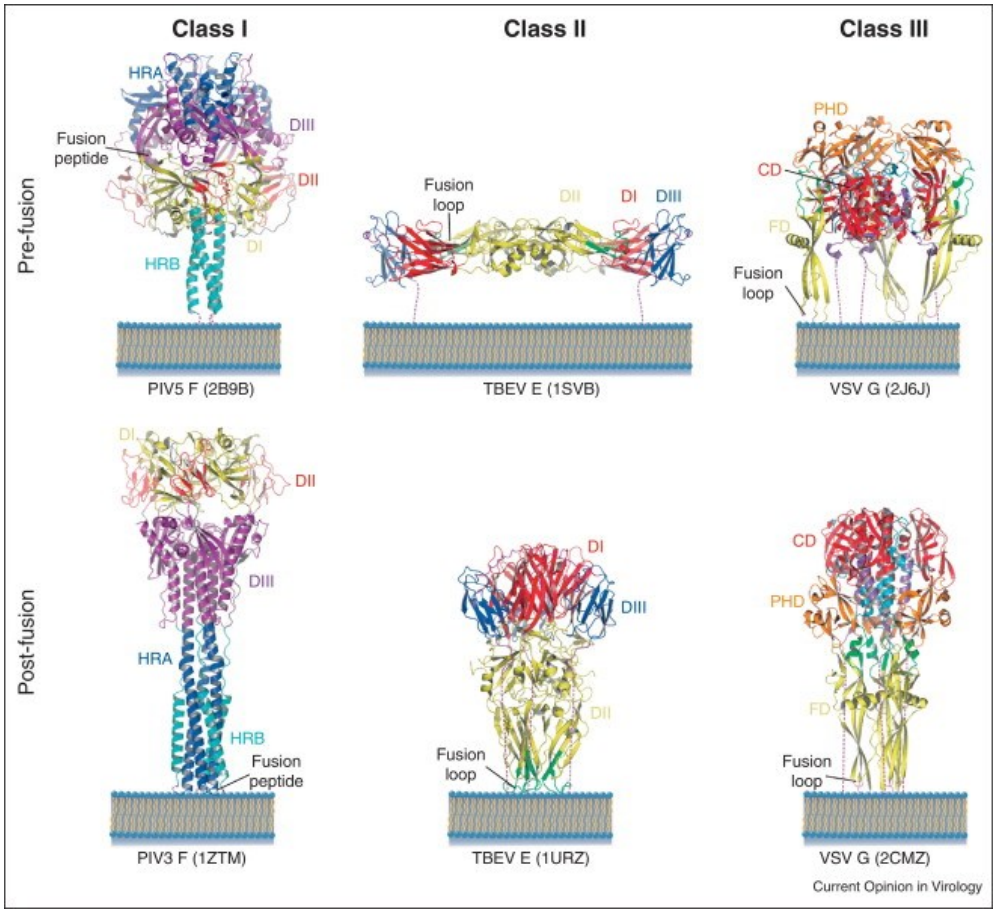


Figure 1.2 Three classes of viral fusion protein
Illustrative ribbon representations of the three classes of virus fusion proteins in pre-fusion and post-fusion conformations. The class I protein displayed is a paramyxoviral F protein from human parainfluenza virus 3 (PIV3) which exhibits a compact trimeric pre-fusion form with the fusion peptide buried inside the globular head. The post-fusion form has an extended α-helical stalk with the fusion peptide inserted into the membrane. The class II protein is the flaviviral E protein from Tick-borne encephalitis virus (TBEV). This is arranged as a head-to-tail dimer on the surface on the virion in its pre-fusion form and converts to a post fusion trimer with its fusion loops inserted into the membrane. The class II fusion protein domains i, ii and iii are coloured red, yellow and dark blue respectively. The class III fusion protein is the rhabdoviral G protein from Vesicular Stomatitis Virus (VSV). Like, the class I proteins, this protein exhibits a compact trimeric pre-fusion conformation and an extended trimeric post-fusion conformation. PDB accession codes are displayed in parentheses. Figure adopted from [17].

The precise mechanism by which each class of fusion protein carries out membrane fusion varies but they all follow the same basic steps. As an example, Figure 1.3 displays a model proposed for a class II fusion protein which was first described by Modis et al, 2004 for the dengue virus E protein and recently updated for the *Bunyavirales* member, Rift Valley fever virus. The class II fusion protein, named Gc for bunyaviruses, is composed of three domains which by convention are coloured, red, yellow and blue for domains I,II and III respectively. In the pre-fusion conformation, the fusion loops at the tip of domain II are shielded from the solvent environment by the chaperone protein Gn. Additional chaperone

proteins are common for class II fusion proteins though there are exceptions such as dengue virus, in which the fusion loop is protected by the formation of a head to tail homodimer [18]. After exposure to the low pH environment of the endocytic pathway, a conformational change is induced in the complex forcing the fusion loops upwards and freeing them from the Gn to embed into the target membrane. The Gc glycoprotein contains a number of histidine residues which are thought to play a role in this pH sensing ability. Following this it is expected that the Gn dissociates from the Gc allowing the Gc to redistribute on the surface of the virus [8]. The next step has been difficult to directly observe but by comparison with the post-fusion conformation it is clear that the protein folds back extending domain III away from the membrane and bringing the tip of domain II, along with the target membrane downwards towards the viral membrane. When two lipid bilayers are brought into close contact they first transition to a hemi-fusion state in which the outer leaflets remain intact while the inner leaflets become merged together [14]. This is followed by the formation of a small pore as the membranes become fully fused, which then widens to enable the passage of the viral genome into the cytoplasm. The Gc ends this processes as a post-fusion trimer in which the fusion loops are facing downwards and embedded in the membrane [8]

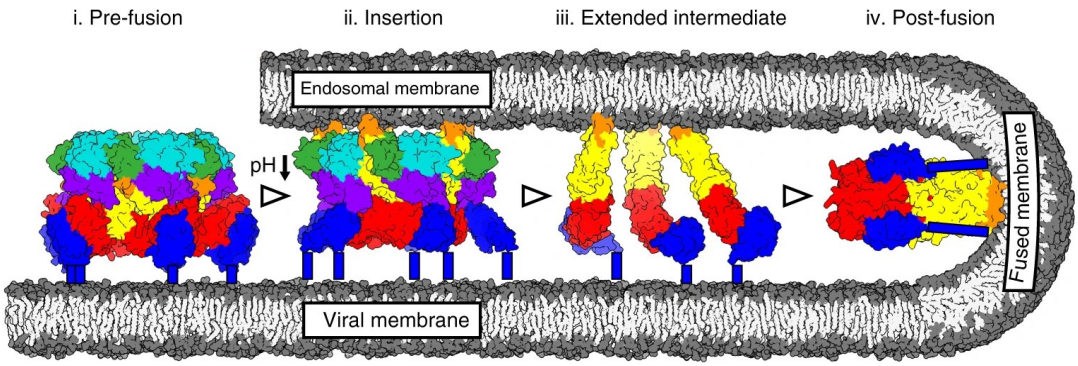


Figure 1.3 Model for class II mediated membrane fusion
A general model for class II mediated membrane fusion using the phleboviral Gc protein of Rift Valley fever virus as an example. The Gc has the classic three domain architecture coloured red, yellow and dark blue with the fusion loops coloured orange. The accessory protein (Gn) is coloured light blue, green, and purple corresponding to domains A,B and C respectively. i) In the pre-fusion state the Gn and Gc proteins form a icosahedral glycoprotein shell on the surface of the virion. For clarity a single pentamer comprised of five Gn-Gc heterodimers is displayed. In this conformation the fusion loops of the Gc are shielded by the Gn. ii) Under low pH conditions a conformational change occurs allowing the fusion loop of the Gc to embed into the target membrane. iii) The Gn dissociate from the Gc which then redistribute on the viral membrane. iv) Following fusion the Gc adopts a folded back trimeric conformation with the fusion loops inserted into the membrane. Figure adopted from [8].

1.1. Hantavirus

Hantaviruses can be broadly categorized into Old World and New World hantaviruses depending on the geographical location of their host species. The host varies with species but typically they are rodent-borne viruses, such as Puumala which infects *Myodes glareolus* (Bank vole), and there are also examples of shrew, mole and bat-borne hantaviruses [3]. A phylogeny of hantavirus species is displayed in Figure 1.4 which shows how viral sequence diversity is related to their host species. Hantavirus outbreaks in humans are commonly associated with various environmental conditions that lead to a boost in the population of their rodent hosts [19]. Transmission to humans occurs via inhalation of aerosols of these animal's excreta which is common among agricultural workers. Transmission is particularly likely in dry, dusty and poorly ventilated buildings that have become infested with rodents [19]. Humans are typically thought of as a dead-end host for hantaviruses with only a few reports of human to human transmission reported for Andes virus only [4], [20] . The first recorded outbreak of hantavirus disease in humans was during the Korean war (1950-1953) in which many soldiers fell ill with what is now referred to as haemorrhagic fever with renal syndrome (HFRS) [21]. It took until 1978 before the first hantavirus (Hantaan virus) was shown to be the causative agent [22]. Since then, other old world hantaviruses, including Dobrava virus, Seoul virus and Puumala virus have been shown to cause HFRS with varying degrees of severity. Puumala virus, which is endemic to northern European countries, causes a milder form of HFRS commonly named nephropathia epidemica [23]. Mortality rates for outbreaks of HFRS vary from <1 - 12% and there are greater than 150,000 cases of HFRS reported globally each year[4]. The majority of HFRS cases are reported in China caused by HNTV infection [4].

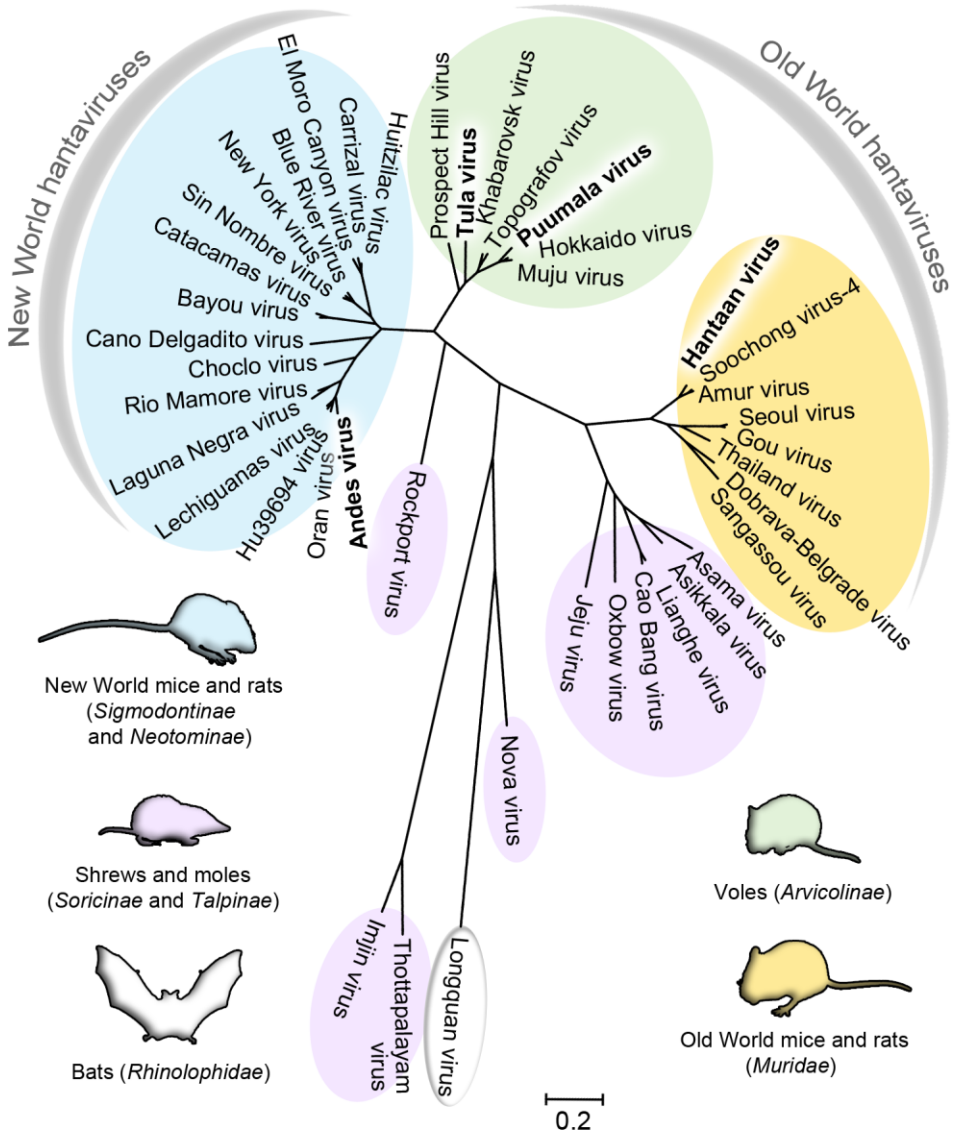


Figure 1.4 Phylogeny of Hantaviridae
A maximum likelihood phylogeny of 42 hantaviral Gn glycoprotein sequences. The clades are coloured according to the host species as shown. Four species studied in this thesis are highlighted. Figure adopted from [24].

New World hantaviruses cause a similar disease in humans named hantavirus pulmonary syndrome (HPS). This is caused by viruses such as Sin Nombre virus (SNV) and Andes virus (ANDV) which can have significantly higher mortality rates of up to 60% [21]. Thankfully, the incidence of HPS is lower than that of HFRS. For example, for the most prevalent disease causing hantavirus in south America, ANDV, there were 786 and 1243 cases reported in Chile between 1995 and 2012 and Argentina between 1995 and 2017, respectively [4].

Following exposure to aerosolised rodent excreta, hantaviruses are able to first infect lung epithelial cells before reaching their main site of replication, endothelial cells, which leads to increased capillary leakage [25]. The conventional view of these diseases, is that HFRS-causing hantaviruses target endothelial cells of the kidney, leading to associated renal complications, while HPS-causing hantaviruses target endothelial cells of the lungs, resulting in pulmonary oedema. In fact, hantavirus pathology is observed in both the lungs and kidneys in both diseases, leading some to suggest they should be referred to as varying manifestations of a single disease [3]. The symptoms of hantavirus disease are highly variable and include fever, headache, cough, myalgia, abdominal pain, nausea and diarrhoea [3]. Common signs of both HPS and HFRS are capillary leakage and thrombocytopenia. In severe HPS, infection leads to pulmonary oedema and cardiogenic shock resulting in respiratory failure that requires mechanical ventilation. Common additional symptoms observed in HFRS are proteinuria and haematuria owing to acute renal disease [19]. Outside the clinical setting, the study of HPS has traditionally utilised ANDV infection of Syrian hamsters (*Mesocricetus auratus*) which results in a lethal disease with similar pathogenesis to that observed in humans. Infection of the HFRS causing hantaviruses, such as HNTV, results in asymptomatic infection in hamsters and adult mice. Instead, newborn mice or immunocompromised severe combined immunodeficient (SCID) mice have been used which both experience lethal disease [26].

Replication of hantaviruses in endothelial cells *in vitro* typically results in little cytopathic effect suggesting alternative causes of disease pathology [3]. For example, increased vascular permeability may result from downregulation of VE-cadherin caused by upregulation of vascular endothelial growth factor A (VEGFA) [27]. Host immune responses likely play a large role in hantavirus disease pathology. Various cytokines are upregulated in response to infection including type I interferons following recognition of viral RNA [28]. This leads the recruitment of cytotoxic T cells and other immune cells to the site of infection which in turn leads to further cytokine release known as a cytokine storm [3]. For example,

interleukin-6 levels in plasma are increased in PUUV patients and correlate to disease severity [29].

There are currently no FDA-approved vaccines or treatments available for any hantaviruses though there are a numerous candidates in various stages of development [4]. Most notably, there is a widespread vaccination program in China based on live inactivated HNTV and SEOV which has had some success [4]. Treatments post infection have been less successful. One candidate, ribavirin, has been shown to be effective at reducing morbidity and mortality of HFRS but only when applied before clinical symptoms are presented [4]. A potential avenue for the development of antiviral therapeutics, are monoclonal antibodies which have shown promise for many other viruses [30]. There are a number of studies that have reported monoclonal antibodies that are able to neutralize hantavirus infection *in vitro* [31]–[33]. In addition there are a few studies that report successful protection from hantavirus infection in animal models. These include mAb HCO2 against HNTV in hamsters and more recently two mAbs, JL16 and MIB22, were shown to prevent ANDV infection in hamsters [31], [32].

An overview of the hantavirus lifecycle is displayed in Figure 1.5. The virus first attaches to the target cell via a host receptor. There are a number of candidate receptor proteins for hantaviruses, as will be discussed in Section 5.1, but a detailed structural description of these interactions is yet to be determined. Following attachment, the virus is taken into the cell by utilising a range of endocytic machinery. Multiple studies demonstrate that the inhibition of components involved in clathrin-mediated endocytosis, including dynamin, impede the entry of old and new world hantaviruses [34]–[36]. In addition there is evidence that ANDV utilises micropinocytosis [37]. Multiple studies also demonstrate a clear dependence of cell entry on cholesterol suggesting that entry may occur via localisation to lipid rafts [37], [38]. Once inside the cell, the virus transports through the endocytic pathway before pH induced membrane fusion mediated by the class II fusion protein Gc. The pH at

which this occurs may vary for different hantaviral species with a pH of 6.3 reported for HNTV and a pH of 5.8 for ANDV [39], [40].

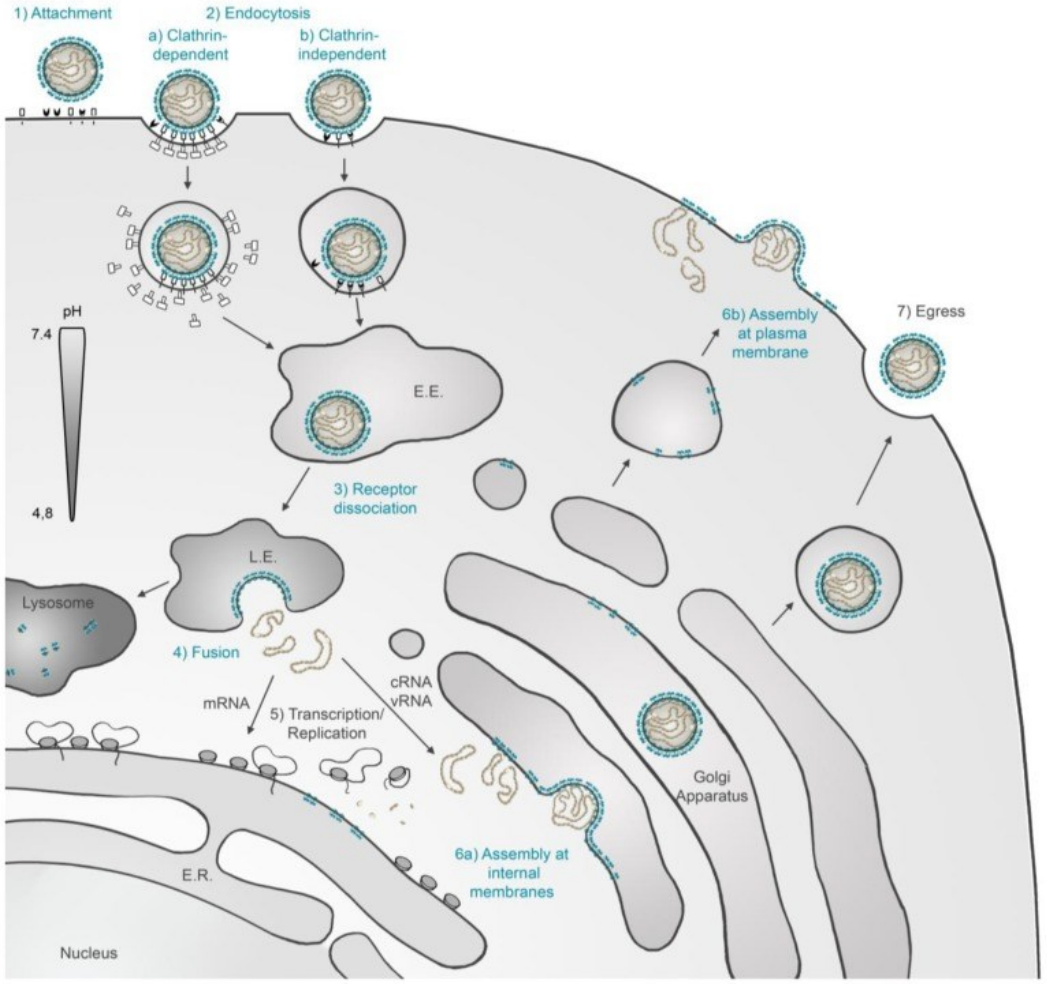


Figure 1.5 The hantavirus lifecycle
1) Attachment of the virus via cell surface receptors. 2) Endocytosis. Hantaviruses have been observed to utilise both clathrin dependent and clathrin independent pathways. 3) Receptor dissociation. 4) Fusion of the of the viral and host membranes in the late endosome (L.E) to allow the release of the virus genome into the cell. 5) Replication of the virus genome and transcription to generate positive sense mRNA transcripts which are both mediated by the viral RdRp. Translation of the surface glycoproteins as a single glycoprotein precursor, occurs at the ER membrane which then proceed through the secretory pathway. The Gn and Gc glycoproteins are formed by proteolytic cleavage of this precursor at the WAASA cleavage site. 6a) Assembly and budding occurs in the golgi. 6b) Assembly and budding has also been reported at the cell surface. 7) Egress of virus particles. Figure adopted from [36]

Membrane fusion leads to the release of the hantavirus genome into the cytoplasm which consists of three negative sense single strand RNA segments; S (small), M (medium) and L (large) [3]. The L segment encodes a viral RNA-dependent RNA polymerase (L protein). This is responsible for the replication of the genome via a cDNA intermediate and for transcription of the viral genome to produce viral mRNA. This processes utilises a “cap-

snatching" mechanism by which the 5' cap from cellular mRNA molecules is cleaved and used as a primer for the generation of viral mRNA [41]. The S segment encodes the nucleoprotein (N) which protects the RNA genome from cellular nucleases by packaging it into a helical assembly [42]. The M segment encodes the surface glycoproteins (Gn and Gc) which are expressed as a single glycoprotein precursor (GPC). This multi-pass membrane protein is cleaved in the endoplasmic reticulum by a cellular signal peptidase complex at the WAASA cleavage site, to produce two glycoproteins, Gn and Gc [43]. The Gn and Gc proceed through the secretory pathway and receive several N-linked glycosylations, of which some appear to be required for protein folding. For example, the prototype hantavirus, Hantaan virus, has four N-linked glycosylation sites on the Gn (N134, -235, -347, and -399) and one on the Gc (N928). Mutations in N134 and also a combination of N235 and N399 were shown to result in misfolding of the Gn [44].

In the golgi, the glycoproteins assemble together and recruit the internal viral components to package inside the virion. Hantaviruses do not have a matrix protein which in other viruses helps to coordinate interactions with internal viral components. The cytoplasmic tails of Gn and Gc both appear to interact directly with the N protein though further study is required to provide a detailed description of how this packaging process is coordinated [45]. The matrix protein in other viruses also provides the force required for membrane curvature during viral budding [46]. For hantaviruses, in the absence of any matrix protein, this force is presumably provided by the Gn and Gc glycoproteins themselves. It has been shown that transfection of the GPC in mammalian cells results in self-assembly of virus-like particles [43]. The budding process appears to occur in the golgi but there are also reports of budding directly at the plasma membrane [36], [47], [48]. To complete budding, a pinching-off reaction is required which many viruses achieve by recruiting cellular components such as the ESCRT (Endosomal Sorting Complexes Required for Transport) machinery. There is little direct evidence of the use of ESCRT machinery

during hantavirus infection but it has been noticed that the cytoplasmic tails of the Gn contain a YxxL motif that is used by other viruses to recruit the ESCRT [49].

Hantavirus virions are roughly spherical with a size of approximately 100 nm but do not exhibit any icosahedral symmetry [50]. Instead the viral glycoproteins form spikes that further create discontinuous locally-ordered patches across the viral surface [9], [50]. This information comes from the study of Tula virus (TULV), a non-human pathogenic hantavirus that is closely related to Puumala (PUUV). The Gc glycoprotein is a class II fusion protein that mediates cell entry as previously described [51]. The role of the Gn is less clear but could act to protect the fusion loop of the Gc as well as to mediate cell surface receptor binding as has been observed in other virus families [8], [52]. A more detailed summary of the structural information known about hantaviruses is included in Section 4.1.

1.2. Nairovirus

Nairoviruses are tick-borne enveloped viruses that have historically been categorised into seven serogroups [10], [53]. As more viruses have been discovered and genome sequencing technology has improved, this has now changed to a classification based on phylogenetic similarity rather than serological reactivity [54], [55]. In total there are now approximately 40 known nairoviruses that have been classified into 14 species. These species broadly spit into those viruses that are carried by soft ticks (family *Argasidae*) and those carried by hard ticks (family Ixodidae) (Figure 1.6). Among these viruses are Nairobi sheep disease (NSD) virus, which is endemic to East Africa and causes often fatal disease in sheep and goats and Crimean-Congo haemorrhagic fever virus (CCHFV), which is the most notable human disease causing nairovirus, with case fatality rates of 10 - 50% [10], [56].

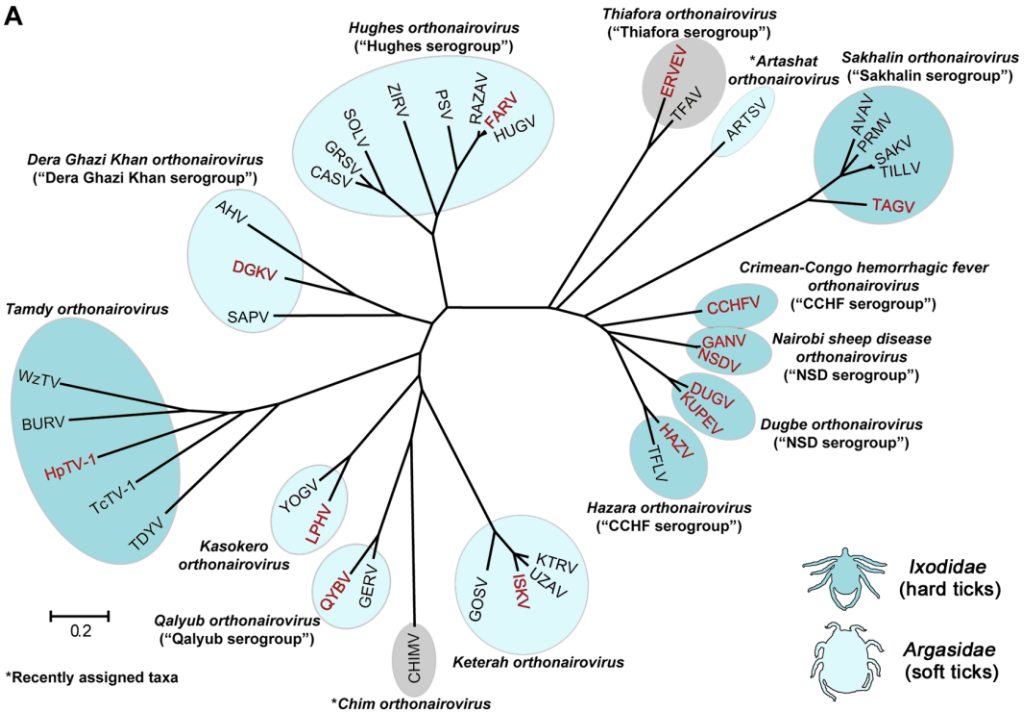


Figure 1.6 Phylogeny of Nairoviridae
A phylogenetic tree of 40 CLUSTALW aligned nairovirus sequences constructed using the Jones-Thornton-Taylor model in the MEGA7 program [57]. The sequence used for comparison is the ovarian tumor (OTU) domain of the nairovirus L protein. Coloured ovals denote the 14 currently assigned species. Where applicable the classical serogroup definitions are included in parentheses. Where known, these species are coloured according to whether they are hosted by *Ixodidae* or *Argasidae* ticks. Figure adapted from [55].

The first recognised outbreak of CCHFV was in 1944 in the Crimean peninsula but has since been found to have a wide ranging geographical distribution covering Southern Asia, the Middle East and most of Africa [10]. CCHFV persists in these regions in *Hyalomma* tick vectors which occasionally transfer the infection, via biting, to a range of vertebrate species including livestock and humans. Curiously the symptoms of CCHFV in animal populations as well as in some humans presents only mild or no symptoms. The reason why only some people develop severe and sometimes fatal haemorrhagic fever is unknown [58]. Human disease can also be transmitted by blood contact with infected livestock and human-to-human transmission has been observed among healthcare workers during outbreaks [59].

CCHFV infection in ticks persists for their lifetime and transfers from parent to child, so they are classified as the reservoir species, while infection in animal populations is transient but may have an amplifying effect which intern raises the chances of human

disease [10]. Human outbreaks of the disease appear to be linked to a broad range of factors that influence the abundance of tick and animal populations in a given area. A good example of this is the first documented outbreak of the disease in the Crimean peninsula in 1944, in which Soviet soldiers and agricultural workers had returned to farmland that had been abandoned under German occupation during the war. As juveniles, the ticks that carry CCHFV target small vertebrate hosts for blood meals but as adults they seek larger hosts such as livestock and humans. The abandonment of farmland during the war, led to an increase in the population of hares and other small vertebrates that are favoured by young ticks, but there was a reduction in the numbers of large livestock. It is thought that this caused an increase in the overall tick population and in particular resulted in many adult ticks that were highly driven to seek out large animal hosts, which resulted in 200 people becoming ill when they returned to the region. Similar outbreaks, concurrent with dramatic changes to the tick ecosystem have also been reported elsewhere [10].

The geographic distribution of the virus appears to be linked to the distribution of the *Hyalomma* tick reservoir species which prefer warm climates, so there are fears that the disease may spread north as a result of global climate change [60]. Recently, CCHFV was detected for the first time in ticks in Spain and the first human case of the disease was reported there, though it is not known how long the virus has been present in the region [60], [61]. Outbreaks of CCHFV are sporadic, with certain countries such as Turkey and Iran, typically reporting over 50 cases per year though the true number of cases is expected to be much higher due to unreported mild or asymptomatic disease [10]. Despite many outbreaks of human CCHFV disease each year, there are limited specific treatments available and no widely used vaccines [58]. This has partly been impeded by a lack of suitable animal models for CCHFV disease. Many commonly used small animal model organisms such as mice, rats, guinea pigs and hamsters display no symptoms under CCHFV challenge. Instead neonatal mice and more recently mice lacking the type I interferon receptor have been used as a lethal model of the disease [62], [63]. Also in the CCHFV serogroup is a non-human

pathogenic virus named Hazara virus that causes haemorrhagic fever in mice [64]. CCHFV is a BSL-4 pathogen so Hazara (BSL-2) is often used as a surrogate for labs without access to BSL-4 facilities [65].

While many people infected with CCHFV remain completely asymptomatic, others develop a non-specific illness categorised by fever, weakness, myalgia, headache and nausea which lasts approximately 3 days [66]. In severe cases this is followed by the onset of widespread haemorrhagic symptoms which can result in visible bruising of the skin [10], [66]. Additional clinical signs of this disease include thrombocytopenia, leukopenia and increased liver enzymes [66]. During this phase of the disease the virus replicates initially in the blood, liver and spleen and later spreads to the lung, kidney and brain. The virus replicates in endothelial cells which results in capillary leakage though it not clear to what extent this is due to the virus itself or the indirect action of the immune system. In response to infection, many pro-inflammatory cytokines are upregulated including IL-6, and TNF-α which correlate to disease severity and can alone increase vascular permeability [67]. In addition to endothelial cells, CCHFV is able to interfere with the immune system by directly infecting macrophages and dendritic cells. It has been demonstrated that patients who die from CCHFV infection mount little, if any, antibody response to the virus [68]. In addition to this, IL-10 is upregulated in response to infection which is known to inhibit cell-mediated immunity [68]. In fatal cases, the combination of a pro-inflammatory cytokine storm, a weak antibody response and uncontrolled virus replication lead to severe vascular leakage and eventually organ failure [10], [66].

The Nairovirus lifecycle is broadly similar to that of other bunyaviruses (summarised in Figure 1.7). Like hantaviruses, nairoviruses first gain access to the cell by binding a cell surface receptor. Currently the cell surface receptor used by nairoviruses is unknown. There is a report that nucleolin may act as a receptor for CCHFV which was found by co-immunoprecipitation with a soluble fragment of Gc [69]. Nucleolin is found mostly in the nucleus but is also observed at the cell surface and has been implicated in viral entry of

Parainfluenza virus type 3 [70], HIV [71] and coxsackie B viruses [72]. In addition it does appear that the Gc glycoprotein is particularly important for nairoviral entry as a panel of antibodies against Gn were not able to neutralise CCHFV infection in vitro but many anti-Gc antibodies did neutralise [73]. However, further study, with live CCHFV is required to confirm nucleolin's role in infection. Following attachment the virus is taken into the cell which appears to be predominantly by clathrin-mediated endocytosis and progresses through the endocytic pathway to finally reach multi-vesicular bodies [74], [75]. Like other bunyaviruses the nairovirus virion is a lipid envelope decorated with two glycoproteins Gn and Gc. The Gc is expected to be a class II fusion protein based on sequence conservation and therefore is likely to be responsible for mediating pH induced membrane fusion [10]. This enables release of the viral genome into the cell for replication and transcription of viral components.

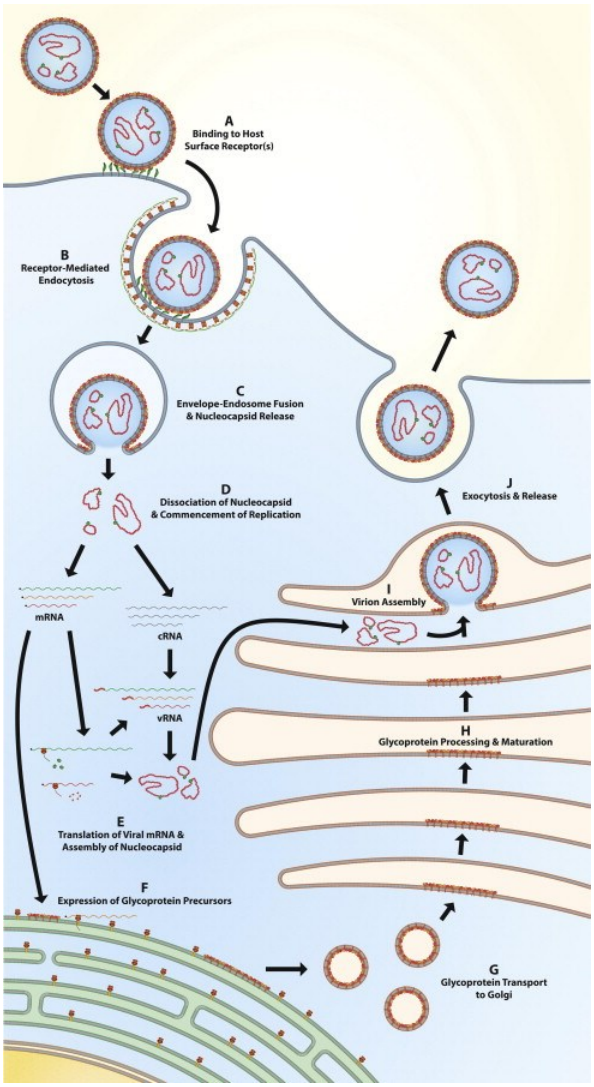


Figure 1.7 The nairovirus lifecycle

A) Attachment via an as yet unidentified cell surface receptor. B) Clathrin mediated endocytosis. C) Fusion of virus and host membranes to enable the release of the virus genome into the cytosol. D) Replication of the viral genome via a cRNA intermediate. Transcription also occurs to generate mRNA transcripts. These processes are mediated by the nairovirus L protein. E) Translation of mRNA transcripts. The surface glycoproteins are translated at the ER membrane as a single glycoprotein precursor chain which then proceed through the secretory pathway (G). H) Further processing of the surface glycoproteins occurs in the golgi (see Figure 1.8 for details). I) Virion assembly occurs in the trans golgi. J) Egress of the virion. Figure adopted from [10].

The Nairovirus genome is split into three segments (S,M, and L) which encode similar proteins as described for the hantaviruses. The S segment encodes the nucleoprotein N, which packages the RNA, and the L segment encodes the L protein which is an RdRp responsible for viral RNA replication and transcription [10]. Like hantaviruses, the M segment encodes the surface glycoproteins Gn and Gc but also some non-structural proteins; NSm, Mucin and gp38 [10], [76]. The full length glycoprotein precursor is co-translationally inserted into the ER membrane and then a number of different proteases are involved in its sequential processing as summarised in Figure 1.8. Cleavage is initiated at several sites by signal peptidase in the ER before further processing by SKI-1 and SKI-1-like proteases in the *cis* golgi. There is later further cleavage between the mucin and gp38 proteins by furin protease [10], [76]. Also in the golgi, several post-translational modifications are applied including N-linked glycosylation of the Gn and Gc as well as extensive O-linked glycosylation of the mucin and gp38 proteins [76]. In the golgi, Gn and Gc associate and recruit the intraviral components to be packaged inside. In the absence of a matrix protein this step is mediated by the internal tails of the glycoproteins [49], [77]. Finally the Gn and Gc mediate budding into the golgi and the complete virion is eventually released at the cell surface. The mucin and gp38 proteins are likely also secreted into the extracellular space but are not directly associated with the virion [76]. There is a large variation in the size and sequence of these regions among different nairovirus species and the function of these proteins is unknown [54]. It is known that similar mucin-like domains are required for the attachment and cytotoxicity of herpes simplex virus type 1 and Ebola virus, respectively, but mucin-like domains are not seen in other bunyaviruses [54], [73], [78], [79].

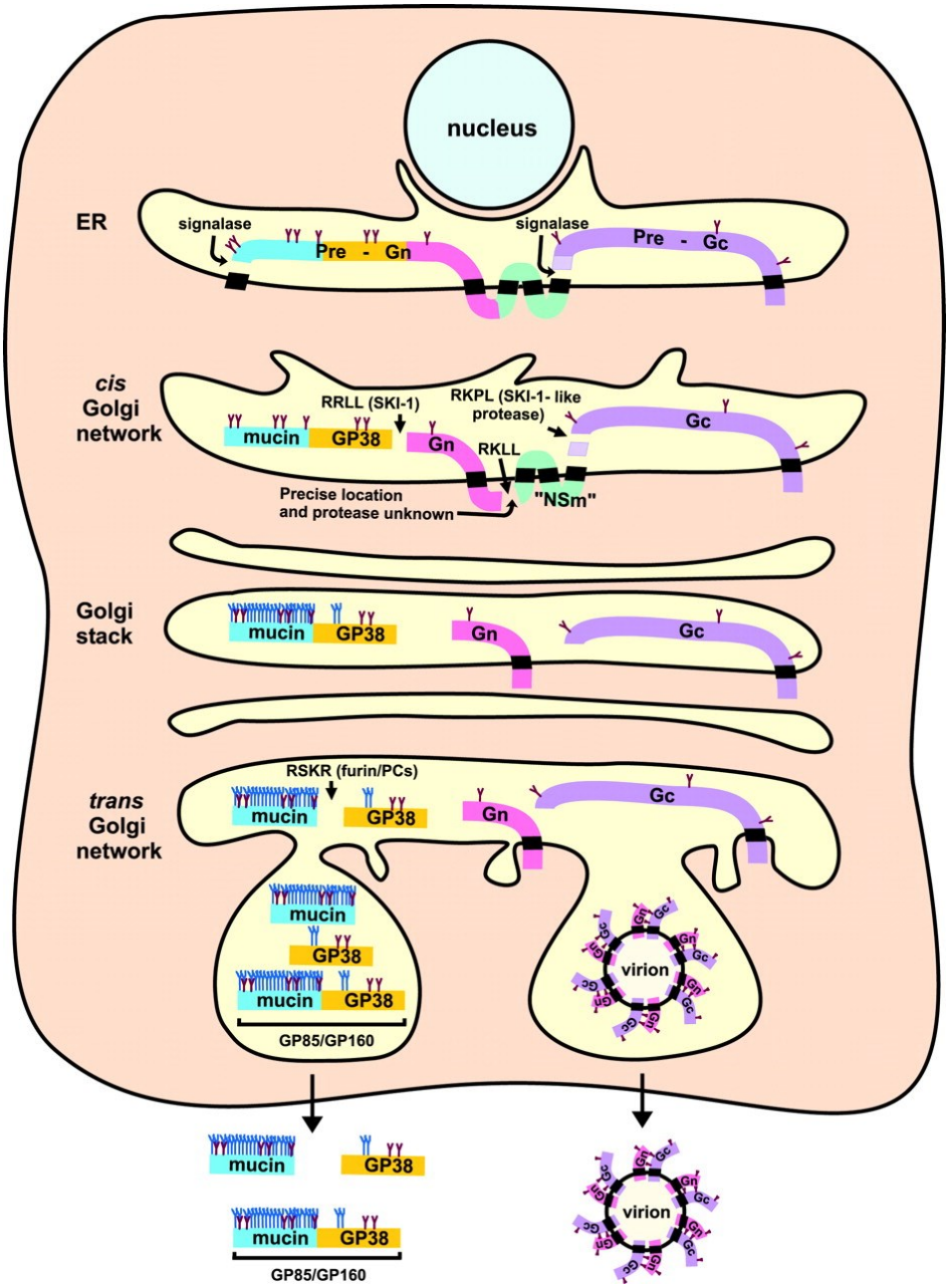


Figure 1.8 Processing of the nairovirus glycoprotein precursor
A schematic detailing the cleavage processing steps of the CCHFV glycoprotein precursor as described in [76]. N-linked and O-linked glycosylations displayed in red and blue respectively. Initially signal peptidase (signalase) cleaves the Pre-Gc from the other proteins and the tip of the Pre-Gn in the ER. Further cleavage occurs in the cis-golgi by SKI-1 and SKI-1-like proteases to produce the mature Gn and Gc. The remaining mucin and GP38 proteins are no-longer membrane anchored and are cleaved by furin protease. Virion assembly and budding occurs in the trans golgi. Following this the virus exits the cell and it is presumed that GP38 and mucin proteins are also secreted but not virion associated. Figure adopted from [76]

Other than a handful of negative stain and plastic section microscopy experiments which confirm it's pleomorphic nature, there are no ectodomain structures of the surface glycoproteins and, at the start of this project, there are no detailed descriptions of how these glycoproteins are organised [80]–[82].

1.3. Cryo-electron microscopy

A powerful tool to improve our understanding of how viruses function, as well as many other biological systems, is structural biology. This includes a number of techniques suited to a range of different targets including x-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and transmission electron microscopy. X-ray crystallography relies on first growing crystals of the molecule of interest which is used to scatter a beam of X-rays producing diffraction patterns that can be used to deduce the structure of the molecule. This technique can routinely produce sufficiently high resolution electron density maps to accurately build atomic models of the molecule of interest. However, the technique requires a crystal of the molecule of interest to be grown, which is not possible for some samples, and only provides a single snapshot which may not reflect the conformational diversity of the protein in its native environment [83]. In contrast, NMR spectroscopy can be used to capture an ensemble of conformations of a protein in its solvent environment but is currently limited by protein size [84]. Transmission electron microscopy is an alternative technique which has been gradually developing alongside these other methods for many years and has recently matured to a point where high resolution structures of a large range of samples is becoming routine [85].

Transmission electron microscopy works by passing a beam of electrons through the sample of interest which is focussed by a lens system to form an image on a detector [86]. A basic schematic of an electron microscope is displayed in Figure 1.9. Electron microscopes require a strong vacuum for the electron beam to pass through so in the early days of this technique, the first obstacle to overcome for biological samples, was to find a way to stabilise the hydrated specimen to protect it from this vacuum. The simplest method to achieve this is to incubate the sample with a heavy-metal stain such as uranyl acetate, blot to a thin film on a carbon coated grid, and allow to dry. This method is known as negative staining because the heavy metal stain strongly scatters electrons providing good contrast against the weakly scattering biological sample of interest. While this is a useful way to

quickly assess the quality of a biological sample, the staining and drying process can introduce conformational changes in the sample that can both limit the attainable resolution and may not reflect the state of the sample in its native environment. In addition, the achievable resolution of negative stain microscopy is limited by the grain size of the stain used and since most contrast is provided by the stain itself, rather than the sample, it mostly provides information about the external surface of the sample [87].

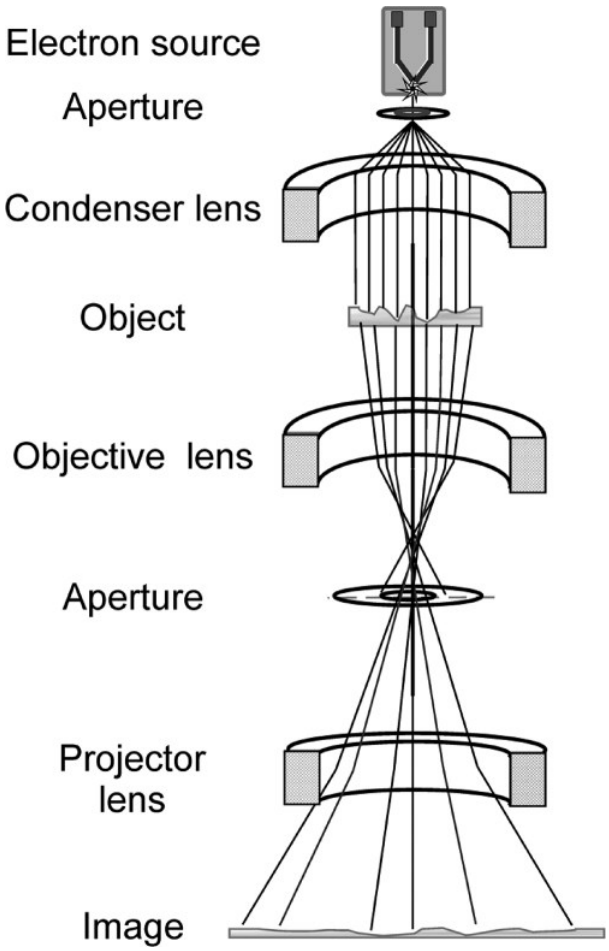


Figure 1.9 A simplified schematic of an electron microscope
A divergent electron beam is first focused onto the sample with a condenser lens. The beam passes through the sample and is focussed through an aperture by the objective lens. Further projection lenses help to magnify the final image that is recorded on a detector. Figure adopted from [86].

An alternative stain-free method was discovered in which a biological sample, suspended in a thin film of water, is rapidly frozen by plunging into a cryogen such as liquid ethane. The speed of this freezing leads to the formation of glass-like vitreous ice instead of crystalline ice which would damage the sample and interfere with the imaging process. The

frozen sample, if maintained below approximately -170°C, is stable in the vacuum of the microscope and therefore provides a general solution for imaging biological systems in a native hydrated state [88]. This is known as cryogenic transmission electron microscopy (cryo-EM) and is a powerful technique for studying purified suspensions of proteins, viruses and even whole cells. The main limitation imposed on the type of sample that can be imaged is the thickness of the sample which must be low enough to allow sufficient electrons through the sample to reach the detector [86], [89].

Obtaining useful images from any microscope relies on the ability to generate contrast. Part of this contrast, termed amplitude contrast, results from electrons that are scattered out of the path of the column by the sample and do not get recorded in the final image. In cryo-EM the biological material we image is predominantly composed of low molecular weight elements (hydrogen, oxygen, nitrogen, and carbon) which weakly scatter electrons resulting in very low amplitude contrast (approximately 7%) [86], [90]. Instead the main source of contrast in cryo-EM is phase contrast which results from the interference between the unscattered electron beam and the elastically scattered electrons. As the electron beam passes through the sample, many electrons pass through unaffected but others interact with the sample becoming either inelastically or elastically scattered. Inelastically scattered electrons deposit energy into the sample, resulting in radiation damage which can limit the attainable resolution if the dose is too high. Elastically scattered electrons become deflected by the sample without a loss of energy and become phase shifted by 90° with respect to the unscattered beam. A further frequency-dependent phase shift is caused by the different path length taken by the scattered electrons as they are focussed by the objective lens onto the detector. Here they constructively and destructively interfere with the unscattered beam in a spatial-frequency dependent manner [86], [90].

These modulations in the phase of the scattered electrons relative to the unscattered electrons are described by the contrast transfer function (CTF) which is dependent on the spherical aberration of the objective lens, the accelerating voltage and the defocus of the

microscope. This function can be seen in the fourier transform of the recorded image as a series of rings named Thon rings [86]. When an image is recorded near focus, the CTF curve becomes spread out resulting in very little phase contrast being transferred into the final image. This leaves noisy images with only very faint features visible. For this reason, cryo-EM images are typically recorded slightly defocussed by several microns which shifts the CTF curve to the centre. This amplifies the low frequency contrast in the image which makes the sample much more visible. Recording defocused images does cause the high frequency information in the image to become scrambled by the oscillations of the CTF curve though this can be partially corrected for computationally by CTF correction [91]. More recently phase plates have become available that induce an additional phase shift to the scattered electrons that enable high contrast images to be obtained near focus [89].

The final step of the imaging process is to record the image on a detector. Detectors have developed substantially from the early days of film to CCD cameras and more recently direct electron detectors. These modern direct detectors are capable of recording fast movies of multiple images that have made it possible to correct for motion in the specimen induced by the electron beam [92]. The images recorded by an electron microscope can be informative by themselves, but the real power comes from being able to computationally process these images to reconstruct high resolution 3D structures. This is done using techniques such as single-particle refinement, cryo-electron tomography and subtomogram averaging which is discussed in more detail in section 3.1.

1.4. Aims

- The initial aim of this thesis was to improve upon the existing structural description of the hantavirus glycoprotein surface. A higher resolution reconstruction obtained by cryo-electron tomography and subtomogram averaging will enable unambiguous fitting of the available hantavirus crystal structures to provide a better description of how the two surface glycoproteins interact and how they combine to create the lattice observed on the surface of the virus.

- A secondary aim was to gain the first structural description of the nairovirus glycoprotein surface which is currently unknown.
- Alongside these aims, I planned to develop improved and optimised strategies for processing cryo-electron tomography data to enable the reconstruction of viral lattices efficiently and at high resolution.
- Beyond simply describing the arrangement of the glycoproteins on the hantavirus surface, I sought to gain insights into how these proteins interact with cell surface receptors to gain access to the cell and how they change in response to the low pH environment of the endocytic pathway.
- Finally, I sought to structurally categorise the ways in which these viruses interact with the host adaptive immune response. In particular, to visualise how antibodies are able to neutralise infection, to provide a structural rationale for the development of future vaccines and therapeutics.

2. Materials and Methods

2.1. Live virus growth and purification

2.1.1. Tula virus growth

Vero E6 cells (ATCC 94 CRL-1586) were cultivated in 8x T175 cm² tissue culture flasks (Griener) in growth medium (DMEM, Sigma) supplemented with 10% Fetal bovine serum (FBS), MEM non-essential amino acids and L-glutamine (Life technologies). Cells were washed in 10ml PBS (Life Technologies) and then infected with virus in 3-5ml of non-supplemented DMEM at a MOI of 0.007. The initial TULV sample (strain Moravia) was a kind gift from Dr Hepojoki, University of Helsinki. After incubation at 37ºC for 40-60 minutes the flasks were supplemented with 20ml of medium supplemented with 10% FBS. After 4dpi this media was discarded and replaced with 15ml of media supplemented with 3% FBS. At 5, 6 and 7 days post infection (dpi) the virus-containing medium was collected and replaced and clarified by centrifugation (3,000 × g for 15 min) to remove large cell debris. Some small aliquots of this were stored at -80ºC to use as stocks for further infections.

2.1.2. Tula virus Purification

TULV containing medium was concentrated approximately 100-fold using a centrifugal 100-kDa cut-off filter (amicon) and overlaid on top of a 25-65% sucrose density gradient in standard buffer (20 mM Tris and 100 mM NaCl pH 7.5). The virus was banded by ultracentrifugation (Beckman SW32 Ti rotor, 25,000 rpm, 4°C, approx. 16hr). The gradient was prepared using a Gradient Master (BioComp Instruments, Canada) and fractionated using a Piston Gradient Fractionator (BioComp Instruments Inc., Canada). Virus containing fractions were pooled, diluted 1:1 in standard buffer before pelleting by ultracentrifugation (Beckman Coulter TLS 55 rotor, 50,000 rpm, 4°C, 2 h). The virus pellet was re-suspended in 40µl standard buffer and stored at 4°C.

2.1.3. Tula virus grid preparation

An aliquot (3 µl) of sample was applied to 1.2-µm hole carbon grids (C-flat, Protochips) that had been glow discharged in a plasma glow discharger for 15 seconds (Harrick). Grids were then floated on a droplet of standard buffer for approximately 5 seconds prior to plunge freezing to remove residual sucrose. For the low pH test (in section 4.4) the sample was floated on either pH 7 or pH 5 buffer for approximately 1 minute prior to plunge freezing. The pH 7 buffer used was the standard buffer and the pH 5 buffer was SPG (succinic acid, sodium dihydrogen phosphate and glycine in the molar ratio 2:7:7) adjusted to pH 5 with HCl. Plunge freezing was performed using a semi-automated vitrification apparatus (CryoPlunge 3; Gatan, Pleasanton, CA). An aliquot (3µl) of 6-nm gold fiducial markers were added to enable tilt series alignment (Aurion). The grids were blotted for 3 s and plunged into an ethane/propane mixture.

2.1.4. Puumala virus growth

The initial PUUV virus sample was a kind gift from Dr. Hepojoki, University of Helsinki. Puumala virus infection was performed identically to that described for TULV virus above. In this case virus containing media was collected and replaced with growth medium supplemented with 3% FBS at 4, 5 ,6 ,7 ,10, 12 and 14 dpi to account for its slower growth.

2.1.5. Hantaviral focus forming unit assay

To determine the titre of hantavirus stocks, an antibody-based focus forming unit assay is required as these viruses do not cause visible cytopathic effect required for a conventional plaque assay. Vero E6 cells were cultivated as previously described in a 24-well plate (Greiner) and then washed in 1ml PBS (Life Technologies). A serial dilution of TULV or PUUV in supplement free DMEM was then used to infect the cells with 200µl/well for 40 min at 37ºC. 1 ml of DMEM supplemented with 10% FBS and 0.8% w/v carboxymethylcellulose (CMC, Sigma) was then overlayed onto the cells which were subsequently incubated at 37 degrees for 5 days (or longer if required). For the detection of focus-forming units the cells were first washed 2x with 1ml PBS/well (Life Technologies) before fixation with 500µl 4%

formaldehyde (Sigma) in PBS buffer for 20 minutes. Following further 3x PBS washes the cells were then permeabilized with 800µl 1% v/v Triton-X-100 (sigma) for 15 minutes and then washed 3x with PBS buffer. The cells were then blocked with 500µl 1% bovine serum albumin (BSA, Life Technologies) for 1 hour followed by a further 3 washes in PBS. Following this, 200µl of an anti-PUUV Gn polyclonal rabbit serum (a kind gift from Dr Hepojoki) was used at a dilution of 1:1000 and incubated for 40-60 minutes. This was followed by 4 further PBS washes before incubation with a fluorescently tagged anti-rabbit antibody at the manufacturers recommended dilution (goat anti-rabbit H+L alexa fluor 488, Thermo Fisher). The plate was then washed a further 3 times before imaging on a fluorescent microscope (Nikon Eclipse TE2000U with Hamamatsu C4742-96 camera) for counting focus-forming units.

2.1.6. Hazara virus growth

SW-13 cells or BSRT7 cells were cultivated in growth medium (DMEM, Sigma) supplemented with 10% Fetal bovine serum (FBS), MEM non-essential amino acids and L-glutamine (Life technologies). Cells were grown in either small scale in a 6 well plate (Greiner), T175 cm² flasks (Greiner) or in large scale in three 5-layer 875 cm² flasks (Falcon). Cells were washed in 10ml or supplement free DMEM (Life Technologies) and then infected with virus in 5ml (small scale 1ml and large scale 25ml) of non-supplemented DMEM at a defined MOI. The initial HAZV sample was a kind gift from Prof Hewson, Public Health England. After incubation at 37ºC for 40–60 minutes the flasks were supplemented with 20ml of medium supplemented with 2% FBS (small scale; 2ml. large scale; 100ml). After 2 or 3 dpi the virus containing media was collected and replaced with 15ml of media supplemented with 2% FBS. Collected media was clarified by centrifugation (3,000 × g for 15 min) to remove large cell debris. Small aliquots were stored at –80ºC for further infections.

2.1.7. Hazara virus purification

A range of purification methods were attempted with Hazara as will be discussed in later chapters. Early purification attempts were performed using PBS buffer. Later attempts used a buffer containing either Tris 10mM and NaCl 20mM or Tris 10mM and NaCl 70mM. To pellet the virus, ultracentrifugation was performed with approximately 23ml of virus containing media through a cushion of 20% w/v sucrose in buffer (Beckman SW32 Ti rotor, 25,000 rpm, 4°C, approx. 2hr). The pellet was re-suspended in 80µl in buffer. Alternatively, virus containing media was concentrated in 100kDa cut-off filters. For small volumes, 15ml centrifugal filters were used (Amicon) and for large volumes a pump powered system (Vivaflow, Sartorius). Where gradient centrifugation was used, a 10-60% w/v sucrose gradient was prepared with a Gradient Master (BioComp Instruments, Canada) and fractionated manually where bands were visible. In order to increase the concentration, the gradient bands were pooled and diluted in buffer to approximately 18ml and pelleted at 28,000 rpm at 4°C in a SW32.1 Beckman rotor. The virus pellet was re-suspended in 60µl standard buffer and stored at 4°C.

2.1.8. Hazara virus grid preparation

3µl of sample was applied to 2-µm hole carbon grids (C-flat, Protochips) that had been glow discharged in a plasma glow discharger for 15 seconds (Harrick). Plunge freezing was performed using a semi-automated vitrification apparatus (CryoPlunge 3; Gatan, Pleasanton, CA). 3 µl of 6-nm gold fiducial markers (Aurion) were added to enable tilt series alignment. The grids were blotted for 3 s and plunged into an ethane/propane mixture.

2.1.9. Hazara virus plaque assay

SW-13 cells were cultivated as previously described in a 24-well plate (Greiner) and then washed in 1ml of supplement free DMEM (Life Technologies). A serial dilution of HAZV in supplement free DMEM was then used to infect the cells with 200µl/well for 40mins at 37ºC. 1ml of DMEM supplemented with 10% FBS and 0.8% w/v carboxymethylcellulose (CMC, Sigma) was then overlaid onto the cells which were subsequently incubated at 37 degrees

for 6 days. For the detection of plaques, the cells were first washed 2x with 1ml PBS/well (Life Technologies) before fixation with 500µl 4% formaldehyde (Sigma) in PBS buffer for 20 minutes. Following further 3x PBS washes the cells were then stained with crystal violet solution (10% ethanol and 2% crystal violet). Cells were finally washed 2x with PBS and plaques counted in those wells with countable numbers of plaques.

2.2. Molecular cloning

2.2.1. Plasmids, cDNA and primers

The pHLsec plasmid vector was used for recombinant protein (PCDH-1) and initial VLP expression (Figure 2.1) [93]. The sequences used were human-codon optimised sequences derived from; Puumala virus M segment construct 1 (Uniprot accession no. Q9WJ31), Puumala virus M segment contruct 2 (ps7 construct as described in [94]), Andes virus M segment (Uniprot accession no. Q80DP7), Hantaan virus accession no. A0A077D153) (See section 8.1). Sequences were synthesised by GeneArt (Life Technologies) and primers synthesised by Eurofins Genomics. Primers were designed to have overlapping regions of sequence at the start and end of the gene which give a melting tempterature (Tm) of approximately 72ºC for PCR. The forward primer included an EcoRI restriction sites to insert the complete Puumala virus M segment including the native signal peptide and also a kozac sequence before the first codon. The use of the KpnI restriction site in the reverse primer allowed the inclusion of the vector derived 6-histidine tag at the end of the gene. Note that initial ANDV and HNTV constructs were synthesised with an included mTurq tag for fluorescent detection and the cloning work for these was performed by Dr Rissanen. After optimisation of VLP production, the constructs were switched to His-tag and fluorescent protein free versions in the pCaggs vector. These constructs were based on the same sequences described above and were a kind gift from Dr Doores, King's College, London.



Figure 2.1 The pHLSEC vector
Key restriction sites in the pHLSEC vector.Figure adopted from [93]

2.2.2. Polymerase chain reaction

Polymerase chain reaction mixtures were performed with the following components; forward primer (500nM), reverse primer (500nM), template cDNA (100-300 pg), and Phusion Master Mix (New England Biolabs). The final volume of the mixture was made up to 50µl with distilled H_2O. The PCR was then performed according to the following scheme; 98°C for 30s, 30 cycles of 98°C for 10 seconds followed by 72°C for 55 seconds, then 72°C for 10 minutes before reducing to 4°C. The resulting PCR products were then visualised by agarose gel electrophoresis. PCR products were purified using the QIAquick PCR purification kit (Qiagen) according to the included protocol.

2.2.3. Agarose gel electrophoresis

Agarose gels were prepared containing 1 % agarose in a TAE (Tris acetate EDTA) buffer and 1x SYBR Safe DNA gel stain (Invitrogen). The DNA was combined with 6x Gel Loading Dye (New England Biolabs) and Hyperladder 1 (Bioline) was run alongside to estimate the construct length. Gels were run at 100 V for 35-45 minutes and imaged using the ChemiDoc XRS+ system (Bio-Rad Laboratories).

2.2.4. Restriction digestion

Purified PCR products and expression vector were digested separately using EcoRI and KpnI endonucleases (New England Biolabs) with the included CutSmart buffer according to the manufacturer's guidelines. The reaction was stopped by the addition of 6x Gel Loading Dye (New England Biolabs) and the samples separated by agarose gel electrophoresis as

previously described. The appropriate bands were cut from the gel and purified using the QIAquick gel extraction kit (Qiagen) according to the manufacturers guidelines.

2.2.5. Ligation of plasmid

Ligation of the digested vector and insert was carried out using the NEB Quick Ligation Kit (New England Biolabs). The molar ratio of vector to insert was calculated using the manufacturer's online ligation calculator. This mixture was incubated with 1 µl of Quick Ligase and 10 µl of Quick Ligase buffer for 15 min at RT.

2.2.6. Transformation of plasmid in E.coli

4µl of the ligation reaction was added to 30µl of competent library efficiency DH5-α *E. coli* cells (Invitrogen) on ice. Heat shock was then performed by incubation at 42°C for 30-40s before cooling on ice followed by the addition of 200µl of SOC medium (Invitrogen). This mixture was then plated on LB-agar plates containing 100µg/mL ampicillin to select for clones containing the vector.

After overnight incubation at 37°C, single colonies were picked and used to inoculate 5ml of LB containing 100 µg/ml ampicillin and grown overnight in a 37°C shaker. The resulting turbid mixture of cells was pelleted and DNA isolated using the QIAprep Spin Miniprep kit (Qiagen) according to the included protocol. The quantity of DNA isolated using this method is typically sufficient for small scale expression tests in mammalian cell culture. An analytical restriction digestion was performed using AgeI and KpnI to check that the vector and insert are of the desired size. Sequencing of the constructs was performed by commercially available service provided by GATC Biotech/Eurofins.

2.2.7. Large scale DNA production

To provide larger quantities of DNA for large scale protein/VLP expression, an initial LB-agar plate was prepared by transformation of 30µl of competent library efficiency DH5-α cells as described above. Colonies picked from this plate were used to inoculate flasks containing LB and 100µg/ml of ampicillin which were incubated for approximately 18 hours in a 37°C

shaker. The isolation of DNA from these cells was then performed using either the Plasmid Giga kit (from 2L of starting culture) or the Plasmid Mega kit (from 1L of starting culture) (Qiagen). The DNA pellet was resuspended in sterile H_2O at RT.

2.2.8. Designing N-terminal His tag variants

The aim of this section was to incorporate an N-terminal His tag into the HNTV M-segment and remove any C-terminal tags at the same time. This involved re-cloning the vector as described above but with the desired modifications made to the primers. The strategy utilised the AgeI cloning site of the pHLSEC vector to include the vector-derived signal peptide into the construct so the primer was designed in such a way to exclude the native signal peptide of the HNTV gene. The reverse primer used at the opposite end of the gene incorporated the XhoI restriction site to allow the resulting insert to be incorporated without including the vector derived C-terminal His tag. It was therefore necessary to include 2 stop codons in this primer sequence. Primers were designed with 6 to 16 Histidine residues and with linker peptides separating the Histdines from the HNTV sequence by 3 or 5 residues. The full sequences used are shown in supplementary Table 8.1. Note that Construct E was built using Construct A as a template.

2.3. Mammalian cell protein and VLP expression and purification

2.3.1. Tissue culture

HEK293T cells were propagated at 37°C in the presence of 5% CO_2. The cell culture medium used was Dulbecco's modified Eagle media (DMEM, Sigma) supplemented with 10% FBS, L-glutamine and non-essential amino acids (Life Technologies).

2.3.2. Small-scale transient protein and VLP expression

To test the viability of a construct for large scale production, a small scale expression test was performed using HEK293T cells in a 6-well plate (Greiner). 4µg of plasmid DNA was added to 250µL of FBS-free DMEM and separately 10µL of lipofectamine (Thermo Fischer Scientific) or 10µg polyethylenimine (PEI) added to 250µl of FBS-free DMEM. After 5

minutes the two solutions were mixed and allowed to combine for a further 20 minutes. The cell culture media was exchanged for DMEM supplemented with 2% FBS and the DNA:lipofectamine mixture added to the cells which are then incubated for 2-3 days. An aliquot of the cell supernatant was assessed by Western or Dot blotting.

2.3.3. Large-scale protein expression

For the production of a large quantity of the hantaviral receptor candidate PCDH-1, a large scale protein expression was performed in a roller bottle system using a similar strategy as described in small scale. HEK293T cells were cultivated in 12 roller bottles (Greiner) with 200mL of DMEM supplemented with 10% FBS in each. 0.5mg of plasmid DNA was complexed with 1mg of PEI in 50ml of FBS-free DMEM per roller bottle and allowed to combine for 20-30 minutes. Meanwhile, the media in the roller bottle was replaced with DMEM supplemented with 2% FBS. The DNA mixture was added and the roller bottles were then incubated for 4-5 days at 37°C. The resulting medium was clarified at 4000xg for 30 minutes to remove cell debris and filtered through a 0.22µm filter prior to purification.

2.3.4. Protein purification

The cell supernatant obtained from roller bottle transfection was first dialysed into the buffer used for protein purification (10mM Tris-HCL pH 8.0, 150mM NaCl) using an Äkta Flux (GE Healthcare) with a 10kDa cut-off filter which also concentrated the volume down to approximately 100mL. This solution was filtered through a 0.22-µm filter before further purification using immobilised metal affinity chromatography (IMAC). The media was passed over a nickel containing HisTrap column (GE Healthcare) to bind the His Tag. The column was washed with several volumes of the purification buffer followed by several rounds of purification buffer containing successively higher concentrations of imidazole (10mL of 20mM, 60mL of 30mM, 60mL of 40mM, 60mL of 40mM, 60mL of 50mM). The protein was then eluted from the column with 250mM imidazole. Protein containing fractions were pooled and concentrated down to approximately 400µl using a 10kDa cut-off centrifugal filter (Amicon) before filtering through a 0.22-µm filter. Size exclusion chromatography was

performed using a Superdex 200 10/30 column (GE Life Sciences) in which the buffer was also exchanged to PBS. The column was pre-equilibrated with PBS buffer and the sample injected onto the column which was fed at a rate of 0.65 ml/min. Protein containing fractions were monitored by their UV signal and analysed by SDS-PAGE. The protein sample was snap-frozen in small aliquots in liquid nitrogen and stored at -80°C for further use.

2.3.5. Large scale hantaviral VLP expression

Transient expression of hantaviral VLPs was performed in HEK293T cells using a similar strategy to that described for small scale protein expression. The following protocol describes the final purification strategy from six 5-layer 875 cm² flasks (Falcon) but note that earlier tests were carried out in T157 cm² flasks (Greiner) using 1/5th the volumes detailed here. 100µg of plasmid DNA was complexed with 500µg of PEI in 25ml of FBS-free DMEM per flask and allowed to combine for 20-30 minutes. Meanwhile, the media in each flask was replaced with 100mL of DMEM supplemented with 2% FBS before the addition of the DNA mixture. The flasks were then incubated for 3 days at 37°C. The resulting medium was clarified at 3000xg for 20 minutes to remove cell debris and filtered through a 0.45µm filter prior to purification.

2.3.6. Hantaviral VLP purification

Multiple strategies were initially attempted for the purification of hantaviral VLPs as will be described in more detail later. The final VLP samples that were used for data collection and subtomogram averaging were purified using the following strategy. The virus containing medium was concentrated down to approximately 30mL using a pump powered 100kDa cutoff filter (Vivaflow, Sartorius). This mixture was then dialysed into an excess of buffer containing 10mM TrisHCl pH 8.0 and 150 mM NaCl through a 1 MDa cut-off dialysis membrane (Biotech CE Tubing, Spectrum Chemical). The high molecular weight cut-off enabled smaller protein impurities to be removed while retaining the large VLPs. The buffer was exchanged once per day over the course of 3-4 days. The resulting solution was then further concentrated down to approximately 3mL using a centrifugal 100kDa cut-off filter.

This solution was then layered onto a 20-60% w/v sucrose density gradient in PBS buffer. The gradient was prepared using a Gradient Master (BioComp Instruments, Canada) and the VLP banded by ultracentrifugation in a SW32 Beckman rotor at 4°C for 4 hours at 25,000 rpm. The diffuse band was collected manually in a volume of approximately 3-4 mL. This solution was then diluted to 20mL in PBS and pelleted through a cushion of 10% sucrose in PBS to further clean and concentrate the sample in a SW32 Beckman centrifuge rotor at 25,000rpm at 4°C for 2 hours. The final pellet was resuspended in 60µl of PBS and stored at 4°C.

2.3.7. SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed to assess sample purity. Approximately 10-20µl samples were prepared by mixing with a 5x loading buffer (0.25 M Tris-HCl pH 6.8, 15% w/v SDS, 50 % v/v glycerol, 25% v/v 2-mercapoethanol, 0.01 % w/v bromophenol blue) followed by boiling at 100°C for 1 minute to ensure denaturation. Samples were then ran on a NuPAGE 4-12% mini poly-acrylamide gel (Thermo Fisher Scientific) at 165 V for 45 minutes. Included in the run is 5µl of a Benchmark protein ladder (Thermo Fisher Scientific) and for Western blots an additional lane included for a Bechmark Pre-Stained protein ladder (Thermo Fisher Scientific). The bands were visualised by incubating the gel in InstantBlue stain (Expedeon) before de-staining in deionized H_2O.

2.3.8. Western and dot blotting

After running an SDS-PAGE, the gel was placed onto a nitrocellulose membrane and protein transferred into the membrane either using the wet-blot NuPAGE Mini Gel apparatus (Thermo Fisher Scientific) with the NuPAGE Transfer buffer (Thermo Fisher Scientific) or with an iBlot transfer apparatus (Thermo Fisher). The membrane was blocked with 5% w/v milk powder in PBS (PBS:milk) for 1 hour and followed by incubation with the mouse-derived anti-PentaHis antibody (Qiagen) using a 1/1000 dilution in PBS:milk. After incubation for 40-60 minutes the membrane was washed 3x with PBS containing 0.5% Tween-20

(PBS:TWEEN, sigma). The secondary antibody used was a horseradish peroxidase conjugated anti-mouse IgG Fc antibody (Sigma) with a 1/2000 dilution in PBS:milk and incubated for 30-40 minutes. Excess antibody was removed with 5x washes in PBS:TWEEN prior to detection with 3ml of Clarity Western ECL substrate (BioRad) and a ChemiDoc XRS+ imager.

For a higher throughput and sensitive method for screening HNTV VLP sample concentration, a dot blot was used in which 2µl of a 2x dilution series of the sample was spotted directly onto the nitrocellulose membrane. For this method, a non-neutralising recombinantly expressed rabbit-derived anti-HNTV antibody named G2 (kindly provided by Dr Doores, King's College London) was used as the primary antibody as this was highly specific for HNTV VLPs with little background signal. The secondary antibody used in this case was a HRP conjugated goat-derived anti-rabbit antibody (ab97080, Abcam). Beyond these differences, the method used was identical to that described for the western blot.

2.3.9. ELISA

ELISA was performed to monitor binding of various proteins to immobilised non-tagged virus/VLP. Accurate estimations of the virus/VLP concentration are difficult to obtain so a constant dilution is used from a highly concentrated and purified sample to absorb onto a 96 well plate at 4°C overnight (Nunc-Immuno 96 MicroWell, PolySorp, Sigma). For example, Tula virus was used at a 1:100 dilution. The excess sample was removed from the wells and the wells washed with PBS. In the Tula virus ELISA, all washing steps were performed with PBS containing 0.05% Tween20 (Sigma). The wells were then blocked with PBS:Milk (PBS(1x):Milk(5%)) solution for 1 hour before removal and further washes. A serial dilution of the analyte protein was prepared in PBS:Milk and 50µl of each concentration applied to the wells in duplicate, and allowed to incubate for 1 hour. The wells were then washed 5 times and incubated with the primary antibody at the appropriate dilution in PBS:Milk. The TULV:fab experiment is a special case because a HRP conjugated anti-mouse antibody was used in a 1:500 dilution to detect the bank vole fabs directly, avoiding the need for a further

secondary antibody (A0168, Sigma). For the detection of His-tagged proteins a rabbit-derived anti-His antibody was used at a 1:5000 dilution (ab9108, Abcam). After incubation for 1 hour the wells were washed again 5x and a HRP conjugated anti-rabbit antibody applied at a 1:1000 dilution in PBS:milk (PI-1000, Vector labs). This was again allowed to incubate for 1 hour and the cells washed a final 5 times. For detection, 50µl/well of a HRP detection kit was used according to the manufacturer's instructions. For the TULV:fab ELISA this was a ABTS Peroxidase (HRP) Substrate Kit (Vector labs) and for the other ELISAs a TMP Substrate kit (Thermo Scientific). The HRP reaction was halted by the addition of 50µl of 2M sulphuric acid and the plate imaged on a Tecan infinite f200 plate reader at 405nm wavelength.

2.3.10. VLP grid preparation

Where protein samples were mixed with VLP, the mixture was prepared before and left to incubate for 1 hour prior to plunge freezing. For more detail on specific experiments see supplementary Table 8.3 the appendix. 3µl of sample was applied to holey carbon grids (2µm hole diameter, C-flat, Protochips) that had been glow discharged in a plasma glow discharger for 15 seconds (Harrick). Plunge freezing was performed using a Vitrobot cryo plunger (Thermo Scientific). 3 µl of 6-nm gold fiducial markers were added to enable tilt series alignment (Aurion). The grids were blotted for 3-4 s and plunged into an ethane/propane mixture.

2.4. Cryo EM data acquisition

Throughout the project, many datasets have been collected using a variety of microscopes, acquisition software and imaging conditions. Details of these are summarised in supplementary Table 8.4.

2.5. Cryo EM data processing

2.5.1. Tomgraphic data preprocessing and tilt series alignment

Tomographic data pre-processing steps were carried out using a custom script named "TomoPreprocess" as described in section 3.2. This includes; motion correction with motioncor2 [95], CTF estimation using CTFFIND4 [91] and a dose-weighting filter was applied to the images according to their accumulated electron dose as described previously [96]. Tilt images were then aligned using gold fiducial markers in etomo, images ctf corrected by phase flipping and tomograms reconstructed. Tomograms were built with a bin factor of 4 initially and if required rebuilt at a bin factor of 2. Where dose weighting is employed, cropped particles were filtered using "dose-weight sharpening" as described in section 3.4. Where relion subtomogram averaging was employed, tomograms were built without CTF correction or dose-weighting as these are handled internally by the software [97].

2.5.2. Subtomogram averaging of pleomorphic virus lattice

A detailed description of the subtomogram averaging workflow is included in section 3.5 and the details of each particular sample included in supplementary Table 8.5. Subtomogram averaging was performed in Dynamo [98] often starting with data that has been reduced by a bin factor of 4. Initial particles for this refinement were created by modelling the surface of the membrane using Dynamo's tomoview program and defining their initial orientation normal to this surface. These initial "seed" particles were more densely populated than the true spikes on the surface. Initial refinement is performed by searching only in the Z-axis with rotationally averaged particles to accurately locate the membrane. The dataset is then split into two halves, on a per virion basis for gold standard refinement with a reference filtered to 50Å to avoid model bias. Subsequent refinements were then performed to allow lateral movement of particles across the surface to locate each target. Overlapping particles were removed based on cross-correlation after each iteration in the refinement. A custom script named patch finder is employed to restrict the data further (see section 3.5). This refinement

utilizes a wide alignment mask encompassing multiple neighbouring spikes and the viral envelope. FSC determination is performed by the use of a soft spherical mask with the base removed to exclude the lipid membrane. The refinement utilizes the adaptive bandpass feature in Dynamo with a pushback of 2 (Fourier pixels). Subsequent, refinement was then carried out with a finer search range and narrower masks. If needed, the refined coordinates were converted to a bin factor of two and then fine refinement repeated under these conditions. For hantaviral samples, these initial steps were carried out assuming C4 symmetry. Where necessary, the refined portions were ported into relion's subtomogram averaging workflow for 3D classification [97]. Here classification was performed without allowing image alignment. The data could then be ported back to dynamo for final reconstruction and occasionally further refinement.

Final maps are depicted as isosurfaces filtered to the resolution determined by FSC. The threshold is determined according to the molecular weight of the viral ectodomain proteins (and any addition proteins bound) assuming a protein density of 0.81 Dalton/$Å^3$.

2.5.3. Tula virus pH comparison

2D class averaging of the virions after exposure to pH 5 and 7 was performed in Relion [99]. Contrast transfer function parameters were estimated using CTFFIND4 [91] and images were corrected by phase flipping. Particles were picked in an evenly spaced circle around the edge of each virus. The images were then binned by a factor of two and subjected to 2D classification with a restricted angular search around the original angle calculated from the near circular geometry of the particle. 226 (pH 7) and 278 (pH 5) particles contributed to each final 2D average. A density profile along the direction perpendicular to the membrane was calculated in Bsoft [100].

3. Tools to improve subtomogram averaging

3.1. Introduction

For many years cryo-EM was limited to resolutions worse than 10 Å, but in recent years the field has undergone a dramatic improvement to the point where the acquisition of near atomic resolution structures of isolated protein samples (of sufficient molecular weight) are now reasonably routine. This development was driven by a number of factors including improved microscope technology, data processing tools and in particular the advent of direct electron detectors to record the images [92], [101], [102].

For cryo-EM, there are a number of different modalities that can be employed to obtain structural information of biological systems, with the vast majority of the near atomic resolution structures utilising “single-particle” cryo-EM. Briefly, this involves the acquisition of 2D projection images of isolated biological molecules suspended in ice. The projection images of each particle are then picked out, often automatically, and then a plethora of software solutions exist to align these projections and reconstruct them into a 3D volume [99], [103]. To achieve high resolution these software solutions also have to account for and partially correct additional factors such as the contrast transfer function (CTF) which describes defects in the image introduced during the imaging process, as well as beam induced sample movement and radiation damage [91], [99]. Single particle cryo-EM works well for many samples including large virus capsids that exhibit icosahedral symmetry and isolated protein samples greater than approximately 100kDa, although it is notable that advances are being made to achieve reconstructions of smaller and smaller molecules [104].

Single particle analysis typically fails to correctly orient the target molecule if it is imaged while overlapping with other molecules at different heights along the imaging axis. For certain samples, this overlap problem is unavoidable such as for pleomorphic viruses like hantaviruses and nairoviruses [105]. Unlike symmetric virus capsids, these viruses exhibit a diverse array of morphologies so cannot simply be aligned and averaged on a per

virion basis. This is also the case for some symmetric viruses that exhibit significant deviation from icosahedral symmetry [8]. Instead, refinement has to focus on the individual glycoproteins on their surface but these are randomly orientated and overlap with each other in the projection images. Another example where a single particle approach is unsuitable is for imaging biological molecules inside living cells. Here the target molecule will be overlapping with other copies of itself and also with all the other molecules and structures that make up the crowded environment inside the cell. For these samples, a different approach is required, namely cryogenic electron tomography (Cryo-ET) [89]. Here the sample is physically rotated inside the microscope and multiple projection images are acquired of the target sample at a range angles. Typically this angular range is limited to ±60 degrees as the thickness of the specimen becomes too great for the electron beam to penetrate the sample, resulting in a "missing-wedge" of information in Fourier space. The projection images are then aligned to each other, often with the help of sample embedded fiducial markers, and reconstructed into a 3D volume called a tomogram. These steps can be carried out using software such as ETomo (from IMOD) which first aligns the images to an "aligned stack" and then reconstructs this into the tomogram [106], [107]. The added signal-to-noise and 3D dimensional information arising from the collection of multiple images in a tilt series mean that often much can be learned simply from visual inspection of the tomogram itself. However to obtain high resolution information, and to account for the "missing wedge" in Fourier space, multiple copies of a target molecule can be picked from the tomogram and aligned in a process called subtomogram averaging (STA) which is analogous to the process used in 2D for a single-particle project. Multiple software packages are available for STA including Dynamo, PEET, Relion and EMClarity [98], [102], [108], [109]

The resolution achieved by STA is typically lower than in single particle projects [102]. The reasons for this are numerous and small improvements are continually being made to whole STA pipeline, from sample preparation to refinement. Firstly, due to the slow nature of tomogram data acquisition, STA projects typically begin with less data which can often limit

the attainable resolution. Secondly, the high accumulated dose applied to the sample over the course of tilt series acquisition causes considerable damage to the specimen which effectively attenuates the high resolution information acquired [96]. It is also well documented that the electron beam also causes considerable movement and warping of the specimen which can blur the sample [95], [110]. This affects all cryo-EM data acquisition and can be partially alleviated by motion correction of rapidly acquired movie frames using software such as motioncor, relion and unblur [95], [96], [99]. However this movement has an added impact on cryo-ET because the warping of the grid affects the tilt series alignment which can be difficult, although perhaps not impossible, to correct [111]. In addition, the geometry of a tilted specimen means that the specimen is thicker at higher tilts than at zero tilt which reduces the SNR of the images and makes CTF estimation more challenging. Finally, factors such as the low SNR, the missing wedge and the aforementioned tilt series alignment problems can all lead to inaccuracies in the alignment of subtomograms which leads to a blurring of the final reconstruction. The STA refinement algorithms can also be prone to errors such as overfitting associated with inappropriately averaged random noise [112]. This is by no means an exhaustive list of all the issues associated with STA and solutions to all of these problems remain an active area of research [102], [109], [111], [113], [114].

Despite these limitations, cryo-ET combined with STA can be a powerful technique for understanding the structure of biological molecules. There are now multiple examples of sub-nanometer resolution STA reconstructions, including the relatively high resolution reconstruction of the immature HIV-1 Gag particle [97], [109], [114]. For certain samples, such as pleomorphic viruses and *in cellula* experiments, STA remains the best way of obtaining detailed structural information and even at more modest resolutions, can help to answer many interesting biological questions.

3.2. TomoPreprocess – a tool for automated preprocessing of tomography data

In recent years, there has been a dramatic improvement in microscope hardware and software which means that the collection of large single particle and tomography datasets is now becoming routine. However, while there are many tools for quickly pre-processing single particle data, including motion correction, CTF estimation and particle picking in real time as the data is acquired, the software availability for tomography datasets is less optimised [103], [115]. After a tomographic tilt series collection, the acquisition software will produce a separate movie file for each tilt in the tilt series. Depending on the tilt scheme used, these may be incorrectly ordered and must be motion corrected with a software such as motioncor2 to account for electron beam induced movement of the specimen [95]. After this stage, the software package ETomo, a part of IMOD, is an excellent option for aligning the tilt series and building the tomogram [107].

In order to speed up this pre-processing I developed a tool named TomoPreprocess to perform the following functions for multiple tilt series with a single command;

- Motion-correction of movie frames into micrographs to account for beam induced movement of the specimen using the software MotionCor2 [95]
- Sort the micrographs into the correct order based on the tilt scheme used
- CTF estimation using Relion which is itself a wrapper for CTFFIND and Gctf [91], [99], [116]. This was included because the tools for single-particle CTF estimation are well developed and can be more user friendly than other methods.
- Dose weighting of the micrographs according to dose applied to mitigate the effects of electron dose damage.

3.3. Dose weighting

One of the factors that limits the resolution of subtomogram averaging projects, is electron damage induced by the electron beam, due to the large accumulated dose received during

the tilt series [96]. One possible way to avoid this could be to use a lower dose for each micrograph or collect a smaller range of tilt angles. However, both these options will reduce the signal-to-noise ratio (SNR) in the subtomograms and make it difficult to accurately align them during refinement. Instead, it may be better to use higher electron doses to aid the alignment but to include a method for preserving the high resolution, un-damaged information that is present in the earlier tilts only.

In a study using rotavirus as a test subject, a range of doses were applied, and the resolution assessed using single particle refinement [96]. This allowed the authors to determine the relationship between electron dose and resolution which produced experimentally derived values for the constants a, b and c in the following equation:

$$W = e^{\frac{-d}{2(af^b+c)}}$$

Where

f = spatial frequency

d = dose

and a, b and c are experimentally derived constants from Grant and Gregorieff 2015 [96] (a = 0.245, b = −1.665, c = 2.81)

This equation describes the dampening that should be applied to limit the contribution of high frequency information in each micrograph that is damaged by the electron beam, while preserving the low frequency information to be used for alignment. This equation is used in single particle projects and is built into MotionCor2, UNLBLUR and Relion during the averaging of movie frames into micrographs [95], [96], [99]. However, these programs are built for single particle projects and do not account for multiple images, acquired in a tilt series. Dose weighting has been used previously to achieve sub nanometer resolution subtomogram average reconstruction using a custom dose-weighting script [117]. This script was kindly provided by Dr. Briggs (University of Cambridge) and I re-implemented this into TomoPreprocess. It should be noted that similar dose weighting solutions have now been implemented into some STA software packages [102], [109].

3.4. Dose weight sharpening

With dose weighting applied, the resulting maps appear unsharp (Figure 3.1A) This is due to the fact that low resolution information is now amplified relative to the high resolution information because a lot of the high resolution information has been dampened. The extent to which the low resolution information in the reconstruction is amplified is described by the average of all the dose weighting functions applied to each micrograph in the tilt series. This allowed the derivation of a new function that is the average of a series of dose weighting functions (See 8.11 for the derivation).

The artificial weighting that is applied to the Fourier components of a tomogram that has been reconstructed from dose weighted images is described by:

$$A = \frac{\sum_{n=1}^{N} e^{ndt+pt}}{N} = \frac{e^{t(d+p)}\left(1 - e^{dtN}\right)}{N(1 - e^{dt})}$$

Where

d = dose per tilt

p = pre applied dose

N = the total number of images

And

$$t = \frac{-1}{2(af^{b} + c)}$$

Where

f = spatial frequency

a, b and c are constants as defined by Grant and Gregorieff 2015 [96] (a = 0.245, b = −1.665, c = 2.81)

The inverse of this function can be applied to the final reconstruction to re-sharpen the map to even out the contribution from all spatial frequencies. This has a similar effect to B-factor sharpening that is applied to reconstructions so will be referred to here as “dose-weight sharpening” [118]. Unlike B-factor sharpening, dose-weight sharpening is independent of the resolution of the map and so can be applied to the original tomogram (or subvolumes) before any refinement is performed. The same affect can be approximated by

applying a B-factor to the maps; however, this has two main disadvantages. First, the B-factor functions vary with the square of spacial frequency whereas the dose weighting function varies with spacial frequency raised to the power of the experimentally derived constant b. This means that a B-factor that perfectly describes the dampening does not exist. Second, for reconstructions lower than 10Å, an appropriate Bfactor cannot be reliably estimated from a Guinier plot so choosing a Bfactor to apply is an arbitrary decision [118].Figure 3.1B shows a reconstruction with the dose weight sharpening applied to restore an even spacial frequency. Similar dose-dependent corrections have been implemented in other software such as relion's subtomogram averaging workflow but this is a general solution to use in any subtomogram averaging software such as Dynamo [102], [109].

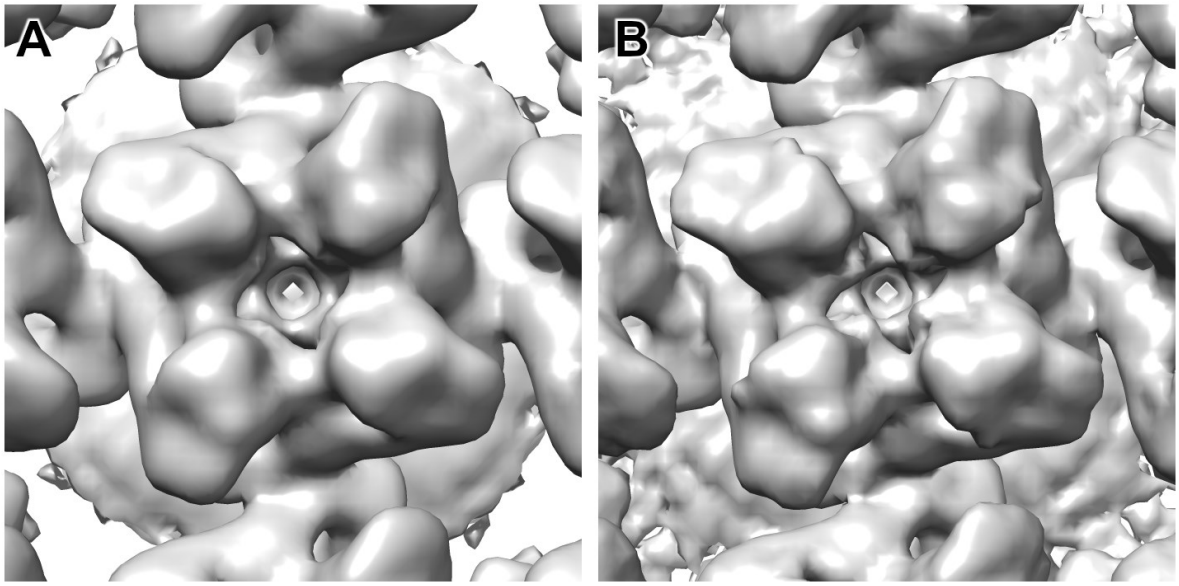


Figure 3.1 The effects of dose-weight sharpening applied to TULV
A demonstration of the effect of the dose weight sharpening function. A) A 11.4 Å TULV reconstruction obtained through subtomogram averaging of dose-weighted data in which the Fourier components have been artificially weighted as a result of the dose weighting applied. B) The same reconstruction in which this weighting has been corrected for using the dose-weight sharpening function.

3.5. Subtomogram averaging of pleomorphic enveloped viruses and the PatchFinder postprocessing method

The strategy used to refine the structure of a protein lattice on membrane-bound pleomorphic viruses can be split into two main stages. The process begins by generating a rough model of the shape of the virus which is inturn used to define the initial particles for

refinement named “seeds”. These seeds should ideally be at least 2x as dense as the true lattice to give the best chance of locating all the particles. The refinement begins with “Stage 1” which uses radially averaged particles and only allows the particles to move in the z direction in order to find the membrane accurately. The subsequent “Stage 2” refinement then allows for full rotation of the particles and a wide image shift search in the x and y directions to locate each spike. During this stage, if any two particles become too close to each other, defined by the geometry of the lattice, the one with the lower cross-correlation score is deleted. After approximately six iterations many of the particles have been located successfully.

However, one problem with this method is that even after multiple iterations, due to the large number of starting “seed” particles, the resulting dataset can still contain many “bad particles” that contribute only noise to the final reconstruction. These particles are often particularly evident at the top and bottom surfaces of the viruses where the refinement typically struggles to locate spikes. This is in part due “missing wedge” effect resulting in a weaker membrane signal to aid alignment. Viruses and VLPs also tend to stick to the air-water interface, which can denature proteins, destroying the lattice, so good top views are frequently absent. In order to obtain higher resolution and more detailed reconstructions it is important to remove the “bad particles” from the dataset but distinguishing “good particles” from bad ones can be difficult.

For membrane-bound viruses with an ordered lattice of glycoproteins on the surface, “good particles” can be distinguished by the fact that they form locally ordered patches, with neighbouring particles that all share similar orientations. The rationale for “patch finder” is to locate these particles by checking the position and orientation of the neighbours for each particle. The script first finds the n nearest particles for each particle, as defined by the user. It then iterates through each neighbour and checks whether the neighbour is located correctly and whether the orientation of the neighbour matches the parent particle to within a defined angular tolerance. The locational restriction is defined by a range of user defined

ellipsoids that define the area in which a neighbouring particle would be expected. If a given number of neighbours satisfy these conditions (typically 3), the particle is selected or else it is removed from the dataset. A further option can be included that also selects particles that were found to be neighbours of a successful parent particle, even if that particle itself does not satisfy the criteria. This avoids deleting particles that are on the edges of an ordered patch and therefore do not have a complete set of neighbours. The parameters for this script can be adjusted to suit any lattice and includes options to account for the symmetry applied to the particles.

Figure 3.2 B-C shows a representative example of a virus lattice (HAZV) with the selected particles coloured blue. Figure 3.2 E-F displays a reconstruction before and after this restriction. A visible validation of this approach is the observation of greater density in the map from neighbouring particles that are outside the alignment mask used during refinement. These neighbours are absent from the original reconstruction due to the “bad particles” that contribute only noise resulting in template bias within the masked region. This template bias is not present outside the alignment mask (as it was not included in the refinement) so displays reduced coherently averaged density. After applying the patch-based restriction, the density outside the mask increases as it comprised of more “real” particles that have neighbours.

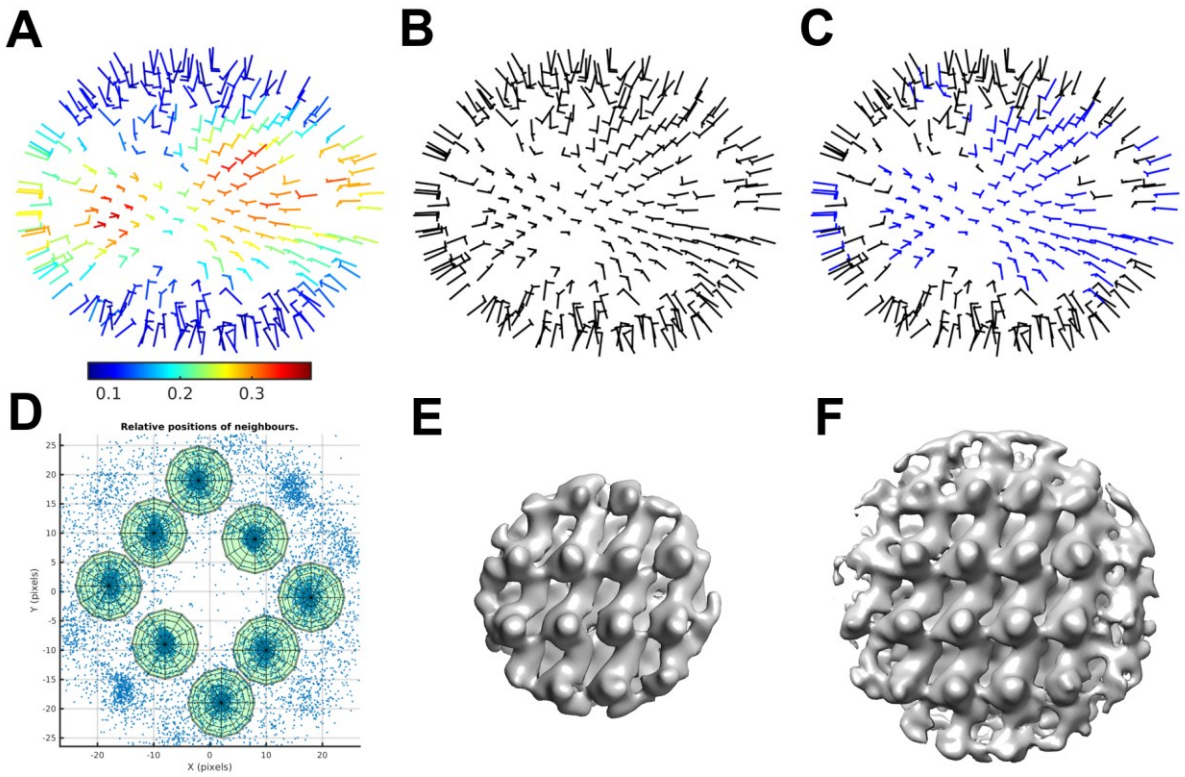


Figure 3.2 The patch finder script for STA of the nairovirus glycoprotein array.
A-C) The refined 3D coordinates of a HAZV virion. A) Coloured by cross correlation. B) Before patch finding. C) The particles belonging to patches are displayed in blue. D) A plot displaying the relative positions of the 8 nearest neighbouring particles to each particle. The ellipsoids that form part of the restriction applied are displayed in green. E-F) A representative reconstruction of the HAZV glycoprotein array before (E) and after (F) the patch based restriction is applied.

An alternative approach to this problem is to restrict data according to its cross-correlation score as good particles tend to correlate better to the template than noise. However the particular value to choose is somewhat arbitrary and can vary between data sets. This therefore requires constant attention to check whether too many or too few particles are being included. The lattice based approach presented here can be applied once and then built into an automated workflow.

The patch finder script is a computationally cheap method for removing bad particles from STA datasets and has now been built into a standard lattice refinement workflow in Dynamo. This workflow starts with “stage1” and “stage2” as described previously followed by patch finder to remove bad particles. A third “stage3” refinement is then carried out with finer search parameters and a narrower alignment mask to improve the alignment of the particles. A final restriction by patch finder can then be applied to produce a reasonable reconstruction

of the lattice. This refinement begins with many particles so can be computationally intensive. It is therefore common to start this with data that has been binned by a factor of 4. Subsequent refinements can then be carried out at lower binning levels if the resolution achieved reaches the Nyquist frequency. As well as this, 3D classification can then be performed on the data to account for conformational heterogeneity in the sample.

3.6. Investigating the resolution limits of subtomogram averaging; Subtomogram reconstruct (Subrec)

At the end of a standard subtomogram averaging project, the location and orientation of the subvolumes have been aligned and the subvolumes are averaged together to create a final map. The resolution achieved in subtomogram averaging projects is generally low compared to single-particle data so there is almost always a desire to maximise the resolution and detail that can be seen in the map to answer the investigator's important biological questions. A number of optimisations that could impact the resolution of the reconstruction rely on making changes to the original images. These include but are not limited to; dose-weighted versus not dose-weighted data, testing an improved CTF estimation of the images, testing improved strategies for the correction of the CTF, testing data reduced by a lower binning factor and removing later electron-damaged tilts in the tilt series. The problem with these potential optimisations is that they require the tomograms to be rebuilt, and particles may need to be re-cropped which can be particularly time-consuming as well as consuming large amounts of computing resources and disk space. As microscope technology improves, the field is moving towards larger and larger datasets with increasing resolution to compound this problem. The rationale behind the set of scripts named "Subtomogram reconstruct" or simply "Subrec" was therefore to bypass this inconvenient step and reconstruct the final map directly from the original 2D data. The script takes the refined positions and orientations of the subvolumes, as well as the corresponding locations of the Etomo project files, and calculates the positions and angles of the subvolumes on both the aligned image stack, and the original images. The 2D particles can then be cropped and reconstructed using the

relion_reconstruct program from Relion [99]. This approach comes with some other added benefits. Firstly, the standard subtomogram averaging workflow contains multiple image interpolation steps as the images are first transformed to the aligned stack, then reconstructed into a tomogram, and finally as the subtomograms are aligned and averaged into the final map. The effect of this is likely small, but will be expected to limit resolution near the Nyquist frequency. Secondly, the use of tools from the single-particle field brings with it access to more modern and mature software features such as Wiener filtered CTF correction, phase plate CTF correction, and Ewald sphere correction [119]. Although it should be noted that phase plate correction is now possible in Etomo and Ewald sphere correction is likely to only impact the resolution of large particles at very high resolution [119]. It should also be noted that the purpose of this tool is not to facilitate any further refinement of the 2D particles using relion, although it is certainly possible to try. The low SNR and overlapping density from neighbouring particles associated with tomographic data will likely cause this to fail and if it does not, it may be better use a standard single-particle approach when collecting the data rather than STA.

There is one technical issue to resolve when using this approach related to 2D particle normalisation. In a single-particle project using Relion, particles are extracted from micrographs and the background solvent is normalised to an average of 0 and standard deviation of 1. This is done by applying a spherical mask over the particle in the centre of the box of a user defined size and then normalising the data according to the solvent containing regions outside of the particle only. The problem with using this approach for tomographic data is that in 2D, particles are almost always overlapping with other particles or contaminating proteins which is often the reason for choosing subtomogram averaging to begin with. The current solution to this problem is to manually pick out the locations of several particle-free solvent regions from all the images in each tilt series. The script then bases the normalisation used for each “real” particle on the calculated average and standard deviation of the nearest “solvent” particle from that image.

In summary, Subrec can be viewed as a helpful tool-kit with which to rapidly test various approaches to subvolume reconstruction. The remainder of this section will use my reconstruction of Tula virus (determined to 11.4-Å resolution) which was obtained using the standard procedure for subtomogram averaging described previously from 18,064 particles with dose-weighted data that was binned by a factor of 2 (see sections 3.5 and 4.3.3) for more details). The FSC is calculated using a soft spherical mask over the central spike with the base cut-off to remove the virus membrane from the calculation. Where dose-weighted micrographs are used the dose-weight sharpening factor described previously is applied to maps (note that this does not alter the FSC curves). It should be noted that some optimisations may only have an effect at resolutions higher than 11.4-Å so further validation may be needed with a higher resolution test subject in the future.

Firstly, Figure 3.3 demonstrates the reproduction of the Tula virus map to an approximately equivalent resolution as obtained by conventional methods. This resolution is the same whether using data binned by a factor of 2 (as was used in the refinement) or with unbinned data. This suggests that the resolution limit is far enough away from the Nyquist frequency of 8 Å to not require unbinned data. The unmasked, masked and phase randomised curves are also displayed in addition to the final corrected FSC curve. This curve experiences a small dip around the phase randomisation boundary but this does not appear to impact the resolution at the threshold of 0.143. It is notable that the map does appear slightly less featured than the initial map obtained from Dynamo. The precise cause of this is unclear but may arise from the different reconstruction algorithms used between IMOD and Relion. These features can however be recovered as will be discussed later in this section.

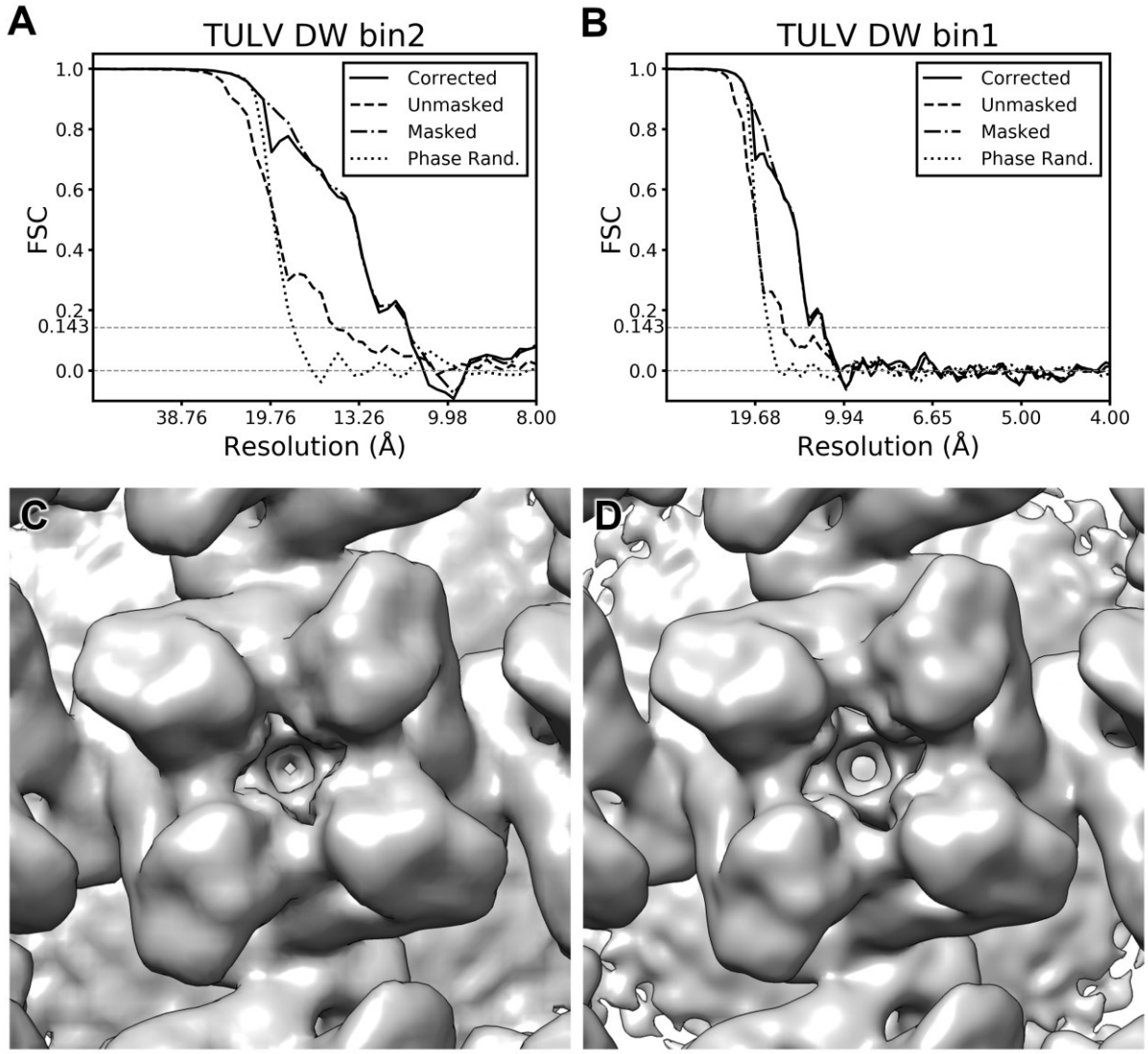


Figure 3.3 Subrec reconstruction of TULV
A) An FSC curve for TULV built with Subrec at bin 2. B) An FSC curve for TULV built with Subrec at bin 1. C) An isosurface depiction of the map from A. Bin factor = 2. D) An isosurface depiction of the map from B. Unbinned.

One thing I was interested to investigate was the effect that dose-weighting would have on resolution. The FSC curves for dose weighted data plotted against non-dose weighted data are shown in Figure 3.4. There is perhaps a very small increase in resolution from the dose-weighted data but this is appears to be mostly from the appearance of a small shoulder in the FSC curve around the threshold value of 0.143 so should perhaps be viewed with caution. Ultimately this demonstrates very little change from the use of dose-weighted images on the final reconstruction. This however does not necessarily mean that dose-weighting does not have a beneficial effect. This is because dose-weighting the data is

expected to help improve the accuracy of the alignment of the subtomograms by not using the electron damaged high resolution information. Note that the reconstruction shown here was aligned using dose-weighted data. The minimal difference in resolution shown here may therefore be because the electron-damaged high resolution information being non-coherently averaged resulting in very little impact on the final map.

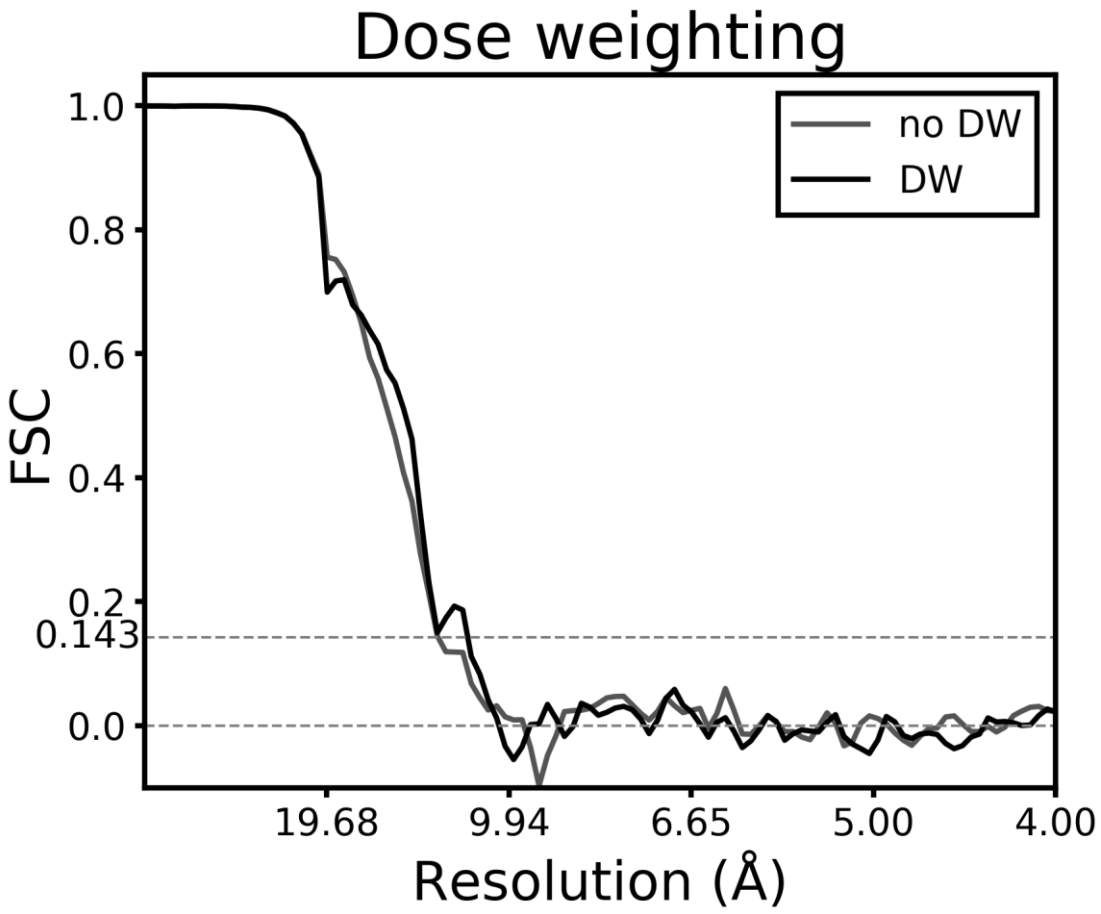


Figure 3.4 The impact of dose weighted data on resolution
An FSC curve demonstrating the effect of dose weighting on the TULV map

Now that the subvolumes have been aligned, there is little reason to include many of the later tilt images in the tilt series as these have received such high doses that all the high resolution structural information has been destroyed. A good example of the power of Subrec is the ability to quickly reconstruct and monitor the resolution of the final map in which each successive tilt image is included in the order in which they were acquired. Tilt

images can of course be removed from tomograms using conventional approaches but this requires the whole tomogram to be rebuilt and particles cropped which can be laborious to test for every tilt image. A script is included with Subrec to carry out this task with a single command to produce the plots in Figure 3.5 A-B. This makes it possible to find the minimum number of tilt images that can be included which still produce the highest resolution map possible. As would be expected the resolution initially improves as more data is added until a minimal value is reached which in this case occurs around the 7th tilt image. The inclusion of more data does not improve the resolution further which eventually slightly drops as more electron-damaged data is included. In addition to this, the fact that higher tilt angle images may be more poorly aligned, partly due to beam induced warping of the specimen, and have thicker ice, may also contribute to the slight fall off in resolution. The use of dose weighted data prevents this small drop off in resolution as previously discussed (see Figure 3.4). One concern associated with dose-weighting images could be that the function used, as derived from a single experiment in [96], may be too harsh, limiting the resolution of the final reconstruction. The evidence presented here demonstrates that this does not appear to be the case for these data.

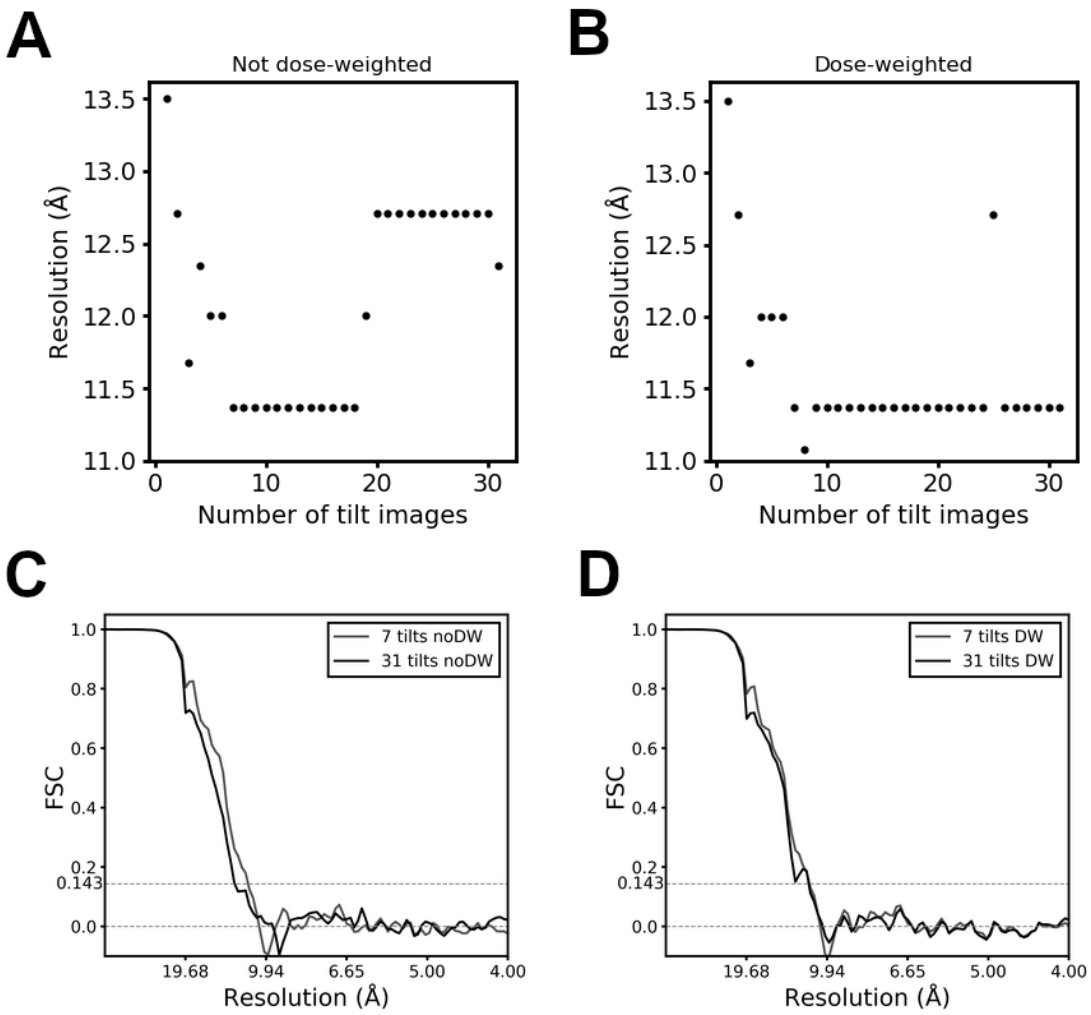


Figure 3.5 The effect of electron damaged images on resolution
A-B) A plot of measured resolution against the number of tilt images included (in the order they were acquired). A) No dose weighting applied. B) Dose weighting applied.C-D) FSC curves with all tilt images and the first 7 tilt images displayed. C) No dose weighting applied. D) Dose weighting applied.

Despite the fairly modest differences in the quoted resolution, the effect of removing later tilt images can be observed in the final reconstruction (Figure 3.6). The maps displayed in Figure 3.6 have all been filtered to 12.3 Å to allow for a fair comparison between those produced from 31 tilts as from 7 tilts. More features can be seen in the reconstructions originating from fewer tilt images which while not conclusive on its own, does suggest the quality of the reconstruction is improved. This does highlight that for low resolution cryo-EM reconstructions, small improvements can be made that are not necessarily captured by a FSC curve. In turn, this shows that low resolution cryo-EM reconstructions should always be

viewed with some caution until they can be validated by comparison to high resolution atomic models.

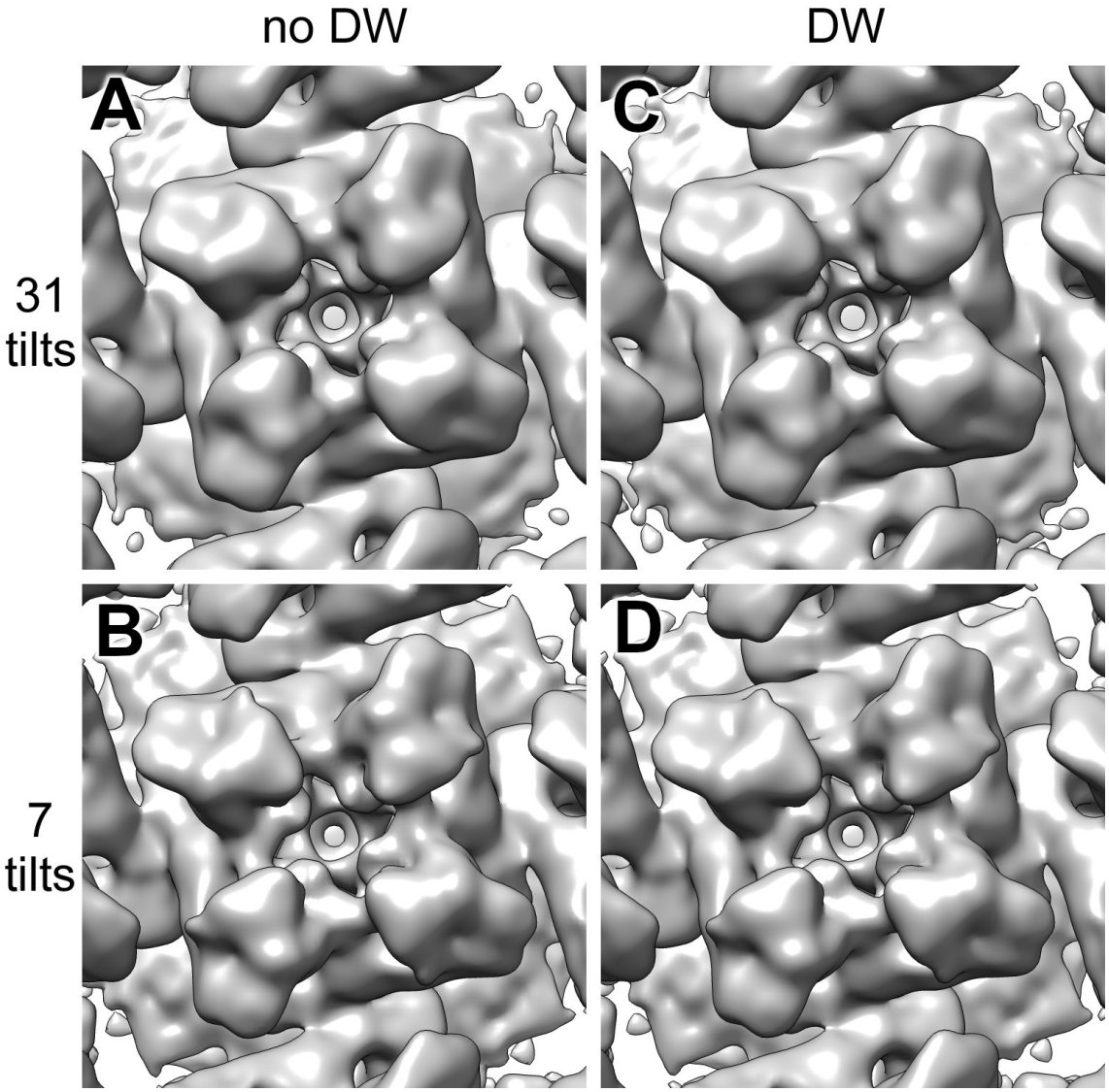


Figure 3.6 The effect of electron damaged data on map quality
Isosurface depictions of TULV with all 31 tilt images and just the first 7 tilt images, with dose weighted data and un-dose weighted data. All maps are filtered to the same resolution for comparison. Dose weighted maps are corrected using dose weight sharpening as previously described. A) All tilt images. Not dose weighted. B) The first 7 tilt images. Not dose weighted. C) All tilt images. Dose weighted. D) The first 7 tilt images. Dose weighted.

Another benefit from this approach is it allows for alternative ways to correct for tilt angle related defocus offsets. Cryo-EM projection images are modulated by the contrast transfer function (CTF) which is dependent on the defocus used when imaging and the z height of the structure of interest in the sample [114]. When a specimen is tilted inside the microscope, the variation in z height of the specimen results in a gradient of defocus across

the image which should be corrected for. The average defocus across the whole image can be estimated using CTF estimation software such as ctfplotter (IMOD), CTFFIND, or gCTF [91], [107], [116]. Following this, in the standard Etomo workflow, the defocus gradient is accounted for by splitting the aligned stack into strips perpendicular to the tilt axis and correcting each with a defocus that has been offset based on the perpendicular distance of the strip to the tilt axis multiplied by the tangent of the tilt angle [107]. However, while this approach does account for the defocus gradient in the image, it does not account for the variation resulting from the thickness of the specimen. Subvolumes that are higher up in a thick tilted specimen experience a different defocus to those directly below, yet both are projected onto the same location in the image and therefore are corrected for identically by the strip-based approach described above. A perfectly accurate approach to correcting the CTF of tomographic data can therefore not be achieved by a global correction of the 2D aligned image stack prior to tomogram reconstruction. One answer to this problem was recently published which incorporated this “3D CTF correction” scheme into the tomogram reconstruction algorithm named NovaCTF [114]. In this study, they demonstrated that this approach was able to improve the resolution of their ultra-high resolution STA structure of the immature HIV-1 CA-SP1 lattice to 3.4 Å. The improvement with respect to the “2D CTF correction” method was ~0.5 Å, which is significant at this resolution range. With Subrec, the 2D images for each particle can be corrected with a unique CTF so this 3D CTF scheme is simple to implement. Figure 3.7 displays the FSC for the 2D vs the 3D correction method applied in Subrec for the TULV reconstruction. At the current resolution of this map, no difference can be observed which is consistent with the reports in [114] that NovaCTF has no effect on reconstructions below a resolution of 1nm.

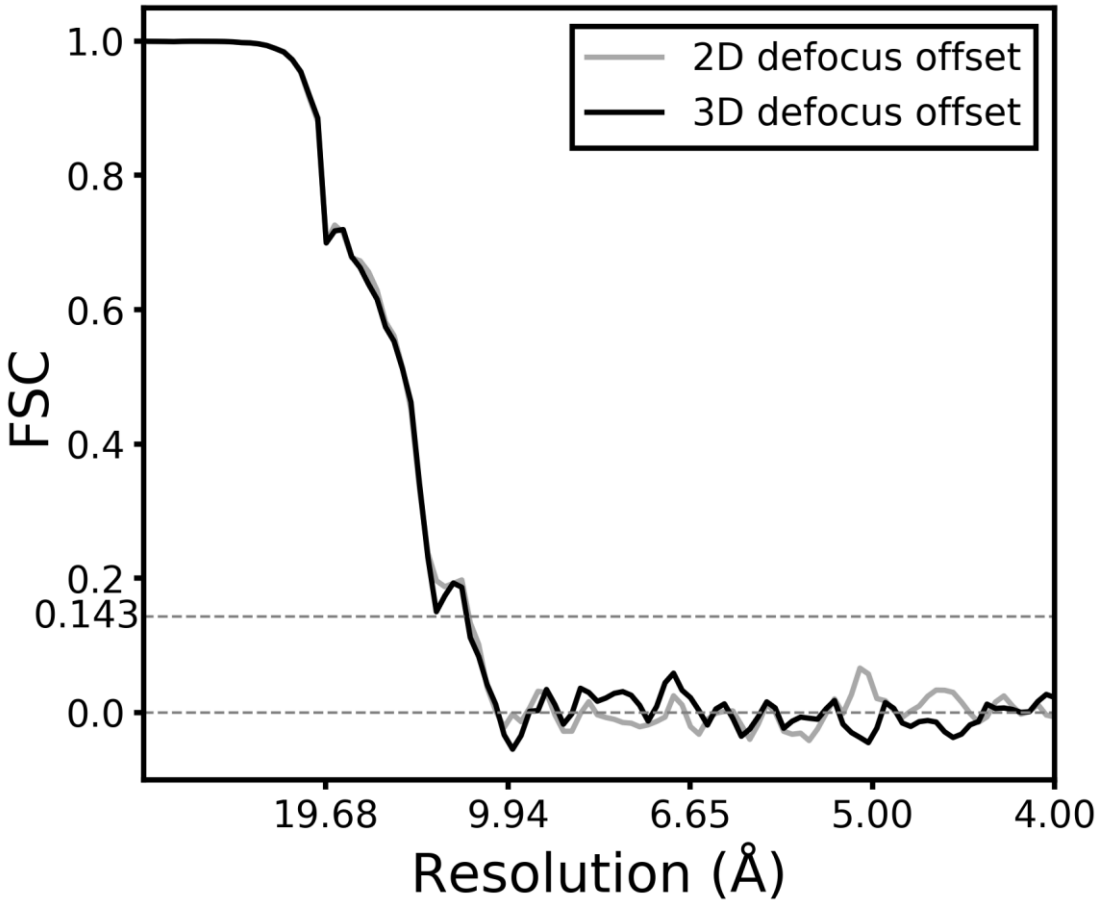


Figure 3.7 3D CTF correction with Subrec
An FSC curve demonstrating the effect of 3D CTF correction on an 11.4 Å TULV map

Next, I explored different CTF correction schemes for sub-tomogram averages. All the reconstructions presented so far have been corrected using “phase flipping” CTF correction which is also used by the ctfphaseflip program within Etomo [106]. An alternative method for CTF correction is built in to relion which utilises Wiener filtered CTF correction to correct for the amplitudes of the CTF as well as simply flipping the phases only [99], [120]. The resolution achieved with Wiener filtered CTF correction is approximately equivalent to that for phase flipping for the TULV map as shown in Figure 3.8.

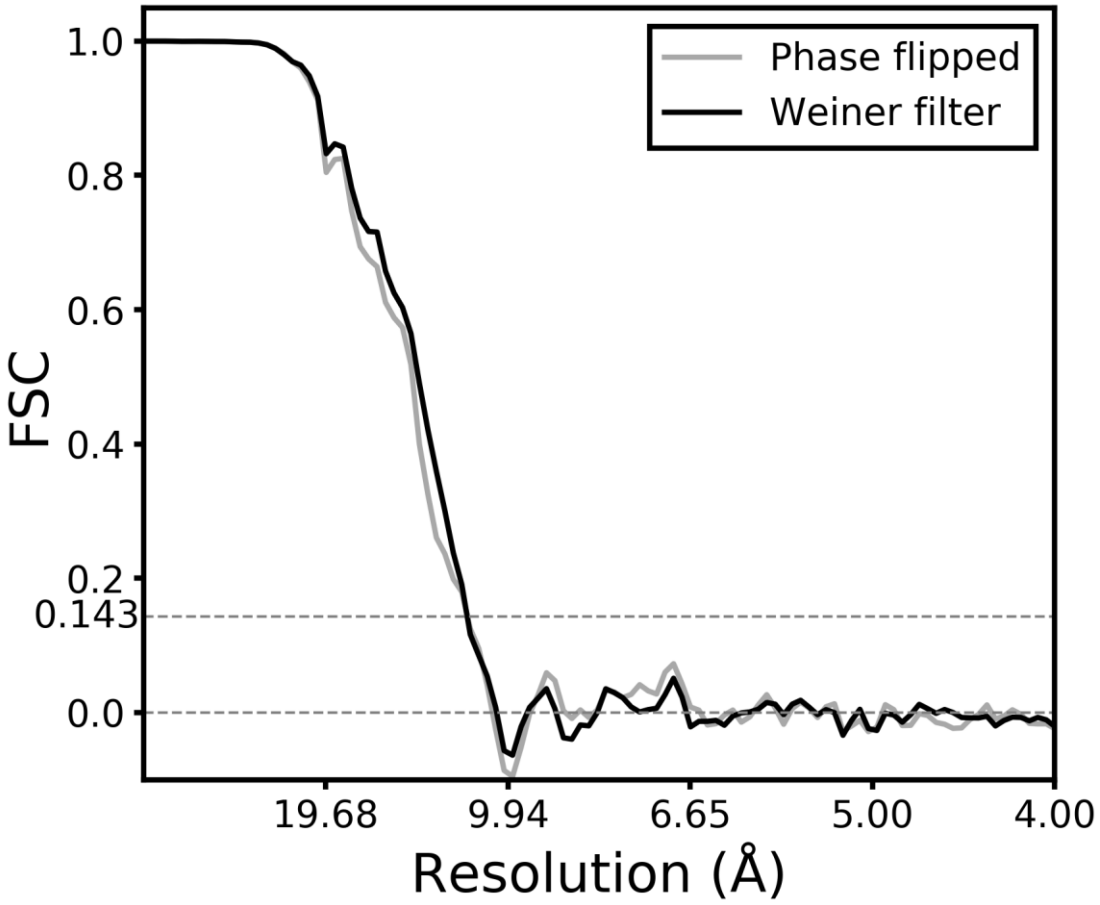


Figure 3.8 The effect of Wiener filtered CTF correction on resolution
An FSC curve displaying the effect of Wiener filtered CTF correction using Subrec on the TULV map.

However, the impact that Wiener filtering has on the appearance of the map is quite marked as displayed in Figure 3.9. Relative to the phase flipped data the density corresponding to the viral envelope is more heavily weighted and the protein density appears smaller. This means that when adjusting the threshold to match the protein molecular weight the reconstruction appears less featured. This difference is partly reduced by removing later tilt images as shown in Figure 3.9 D although not completely. One explanation for this may be that this change more accurately represents the true nature of the TULV viral envelope with correctly weighted amplitudes. However the size of the phase flipped structure does appear to fit well to the available hantaviral crystal structures so this is perhaps unlikely (see section 4.3.3). Alternatively, this may suggest that the CTF estimation for this data may not be sufficiently accurate resulting in inappropriately corrected

amplitudes. Finally, there may simply be a fundamental problem with the use of a Wiener filter for the correction of low resolution, low dose, thick, tilted tomographic data. Issues with Wiener filtered CTF correction of low SNR cryo-EM data have been previously reported but the precise nature of the difference in the maps presented here remains an area for further investigation [120].

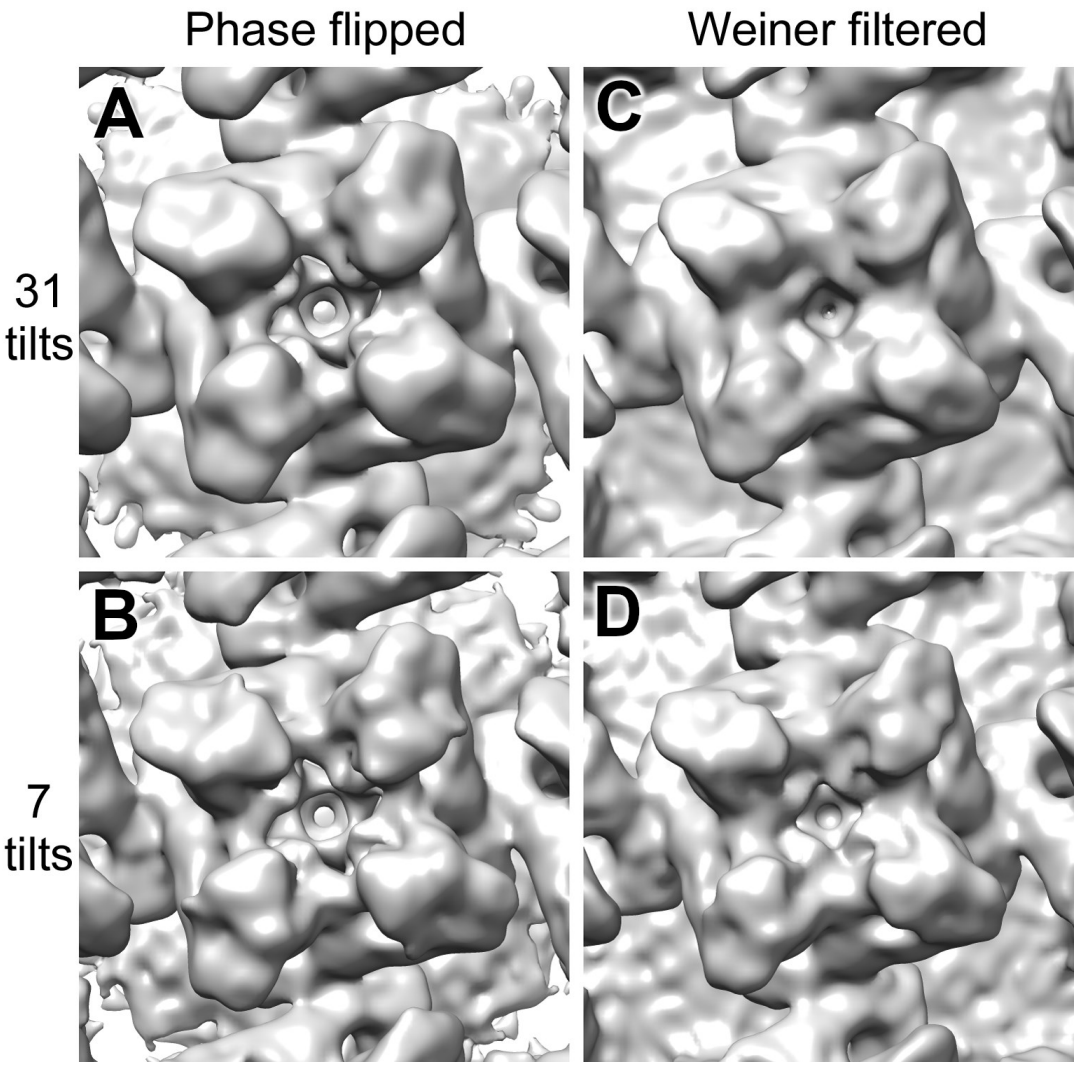


Figure 3.9 The effect of Wiener filtering on the appearance of STA data
Isosurfaces of the TULV map with phase flipping CTF correction (C-D) and Wiener Filtered correction (A-B). Maps are displayed including all the tilt images (A-C) and just the first 7 tilt images (B-D).

Also implemented in this set of tools is a script to rebuild the unaligned subtomograms using relion_reconstruct with any of the options discussed above and with the added benefit of skipping the interpolation step associated with the creation of the aligned stack. This could then be used for further alignment in any subtomogram averaging

software. As an example, the Tula virus subvolumes are reconstructed and then averaged using Dynamo. The resulting map reaches a similar, if slightly reduced resolution, as previously shown (Figure 3.10). An approach such as this could, in principle, be used to reconstruct unbinned subvolumes without the need to build a large unbinned tomogram which can be very demanding in terms of disk space requirements.

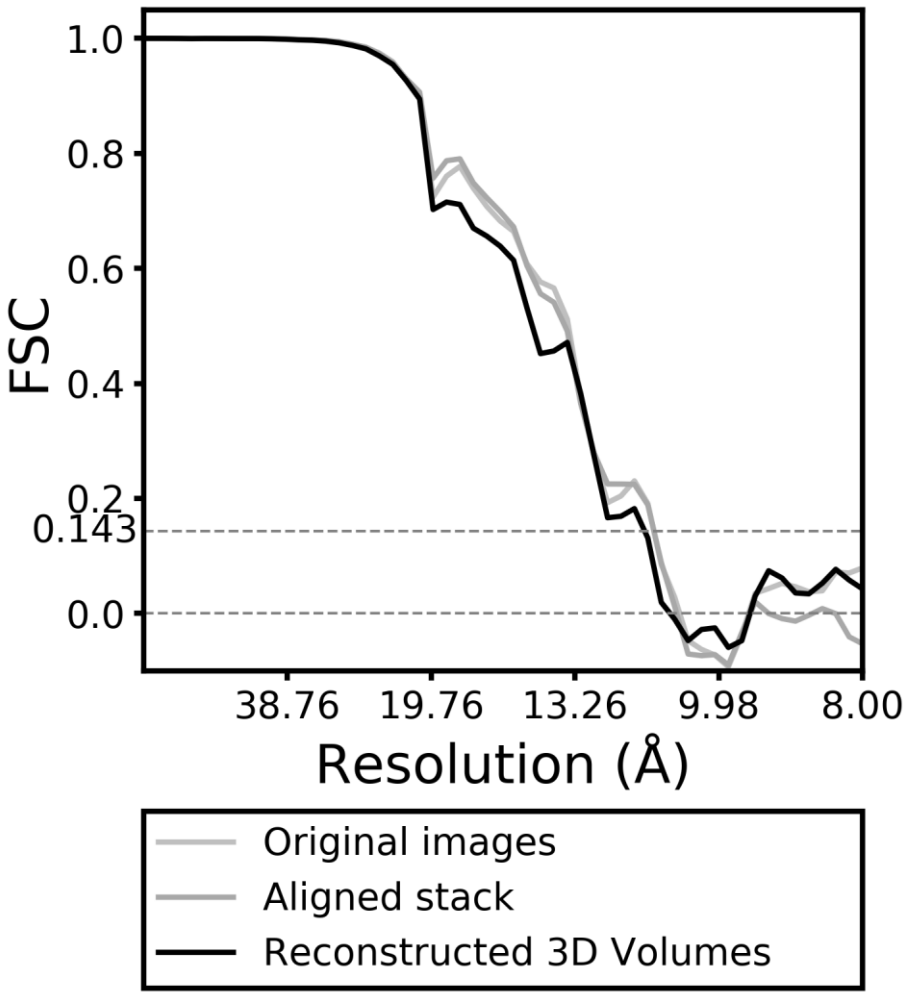


Figure 3.10 Reconstruction of unaligned subvolumes using Subrec
An FSC curve displaying the.resolution of 3 TULV maps obtained using Subrec built with a bin factor of 2. The first is reconstructed from the original 2D images.The second from the aligned stack 2D images. The third is from reconstructed unaligned subvolumes. These were then averaged in Dynamo before resolution determination with relion_postprocess.

3.7. Discussion and future directions

The tomogram preprocessing script presented here has been a useful tool throughout this project to quickly pre-process data before tilt series alignment. However, this and many of the tools associated with subtomogram averaging in general, leave a lot of room for improvement in terms of ease of use and overall simplicity. The larger and larger datasets demand a more streamlined, automated, processing approach which will hopefully continue to develop more in the future. An example of this improvement is the development of automated tilt series alignment in ETomo [107] and a more complete pipeline approach in EMClarity [109].

The data processing scheme presented here works well for the reconstruction of virus protein lattices from tomographic data and examples of its use can be found throughout the rest of this thesis. The use of the patch finder script has helped to make this process semi-automated and can, in principle, be applied to any enveloped virus displaying an ordered lattice on the surface.

The development of Subrec provides a toolkit for testing various factors associated with subtomogram averaging and its development has been a useful learning exercise. Sadly, and perhaps unsurprisingly, little benefit has been gained for the TULV reconstruction. This suggests that other factors limit the resolution on this map. These limiting factors could be any number of the issues discussed in the introduction to this chapter including inaccuracies in the alignment associated with beam induced specimen warping [111]. Perhaps the most likely explanation for the limited resolution of TULV is due to the inherent flexibility of the spike itself. The hantaviral envelope is metastable and the angle of one spike relative to its neighbours are variable which likely affects their alignment [24], [50], [121]. These problems could potentially be partially overcome in the future with a much larger dataset and more stringent 3D classification.

In the future, Subrec should be applied to different datasets that have been demonstrated to reach higher resolutions than TULV to gain a better understanding of its strengths and weaknesses. While there are a few example datasets freely available, these are frequently not released with all of the information required for Subrec to work (original images, etomo project files and refined positions) so for time-limitations and to focus more on the biological side of the project, these have not yet been attempted.

In summary, cryo-ET combined with STA is a useful technique for obtaining structures from samples where overlapping particles or other sources of contamination prevent accurate single particle refinement, such as with pleomorphic enveloped viruses. In recent years the field has rapidly developed but there are still many improvements that could be made, both in terms of attainable resolution and ease of use. As microscope technology improves it is hoped that tomographic data collection becomes simpler to set up and faster to enable the collection of larger datasets. This in turn must be met with improved strategies to deal with the vast amounts of data that will be available. As demonstrated by Subrec, there is perhaps scope for further integration between the 2D and 3D refinement paradigms of single particle refinement and STA. If acquired reasonably quickly, the alignment in some single particle projects could possibly benefit from added tilt information, as seen with the collection of tilt-pair datasets [122]. In turn, STA datasets could benefit from the more mature software tools available in single particle refinement. It is therefore possible that these two workflows could merge into a single framework in the future to facilitate easier and more robust structure determination of challenging targets like pleomorphic viruses by cryo-EM.

4. Bunyavirus ultrastructure

4.1. Introduction

The order *Bunyavirales* contains a number viruses that utilise class II fusion proteins to gain entry into the cell. Despite the structural similarity observed for class II fusion proteins, the broad architecture of these viruses is diverse. For example, Bunyamwera virus from the family *Peribunyaviridae* exhibits a tripodal spike across its surface and Rift Valley fever virus from the family P*henuiviridae* are broadly icosahedrally symmetric although its virions display a degree of flexibility (Figure 4.1 A&C) [8], [123]. Hantaviruses exhibit yet another arrangement of surface glycoproteins; a discontinuous square lattice of spikes. This lattice was first described using subtomogram averaging in [50] and later refined to a resolution of 16 Å [9] (Figure 4.1 B,D-F). There is also a low resolution reconstruction of Hantaan virus which was solved using an unconventional single particle approach that included only side views from the perimeter of the virus [124].

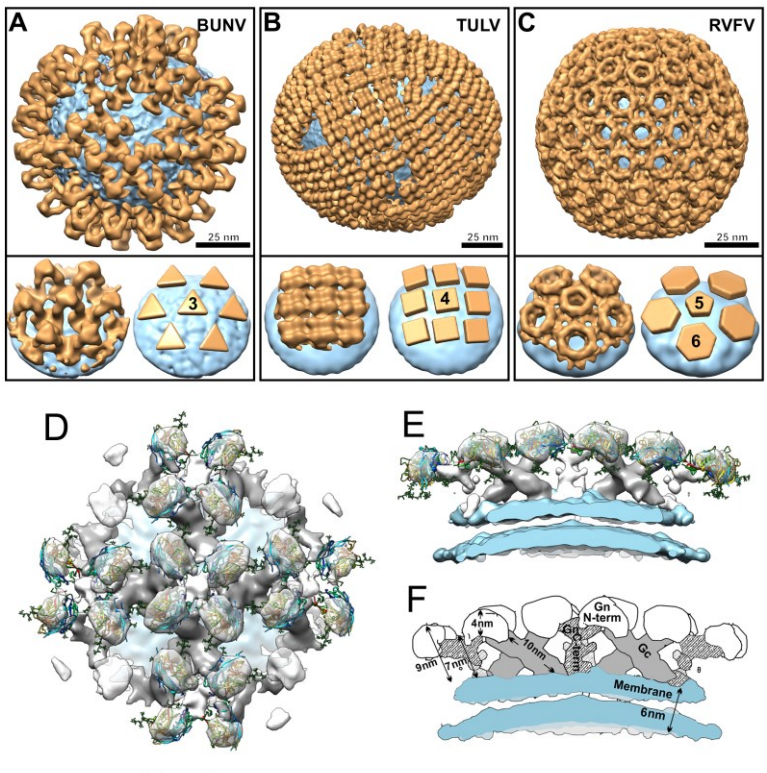


Figure 4.1 Bunyavirus glycoprotein cryo-EM

A-C) Depictions of the surface glycoprotein arrays of different bunyaviruses. Below schematics are shown of the varied symmetrical arrangements displayed. A) Bunyamwera virus; BUNV; family Peribunyaviridae; EMD-2352. B) Tula virus; TULV; family Hantaviridae. EMD-1704. C) Rift Valley fever virus; family Phenuiviridae; EMD-1550. D-E) A reconstruction of the hantavirus surface reported in [9]; EMD-3364. The membrane is displayed in blue, the glycoprotein surface in grey and a tentative fitting of the PUUV Gn crystal structure (PDB-5FXU) is shown. D) Top view. E) Side view. F) A schematic showing the probably locations of the hantaviral glycoproteins in the map. Note that an unambiguous fitting of the Gc is not available. Sections A-C adopted from [123] and sections D-E adopted from [9].

The two hantaviral surface glycoproteins have also been studied extensively in recent years by X-ray crystallography, providing higher resolution detail as summarised in

Figure 4.2. A fragment consisting of approximately 350 residues near the N-terminus of the Gn has been solved for PUUV and HNTV [9], [24]. This region forms a globular domain that is expected to occupy the four membrane distal lobes in the cryo-EM map. There is currently no high resolution data to describe the remainder of the Gn. The entire ectodomain of the HNTV Gc has been resolved in both the “pre-fusion” and “post-fusion” conformation [121]. The low pH post-fusion conformation consists of a classic post-fusion trimer that has been observed for other class II fusion protein displaying viruses [125]. The pre-fusion conformation has two main differences from the post-fusion conformation; a 26 degree bend at the boundary between domains I and II and a large change in the location of domain III which is flexibly linked to the other domains. Interestingly, a Gc:Gc homodimer interface, covering an area of ~545 Å² per subunit, can be seen in the crystal packing of this lattice. Although, Gc ectodomains appear to be monomers in solution [16] this dimer interface has been proposed to be physiologically relevant in the context of the virion [121]. Destabilising mutations in this interface impaired the formation of virus-like particles and stabilising mutations at this interface were able to increase the temperature dependent stability of the spikes. This study went on to propose a model by which hantavirus spikes exist in an equilibrium between “closed and “open” states around physiological temperatures, wherein the fusion loops of the Gc are exposed to embed in the target membrane. This suggests that the hantaviral surface is metastable and capable of undergoing significant conformational changes [24], [121]

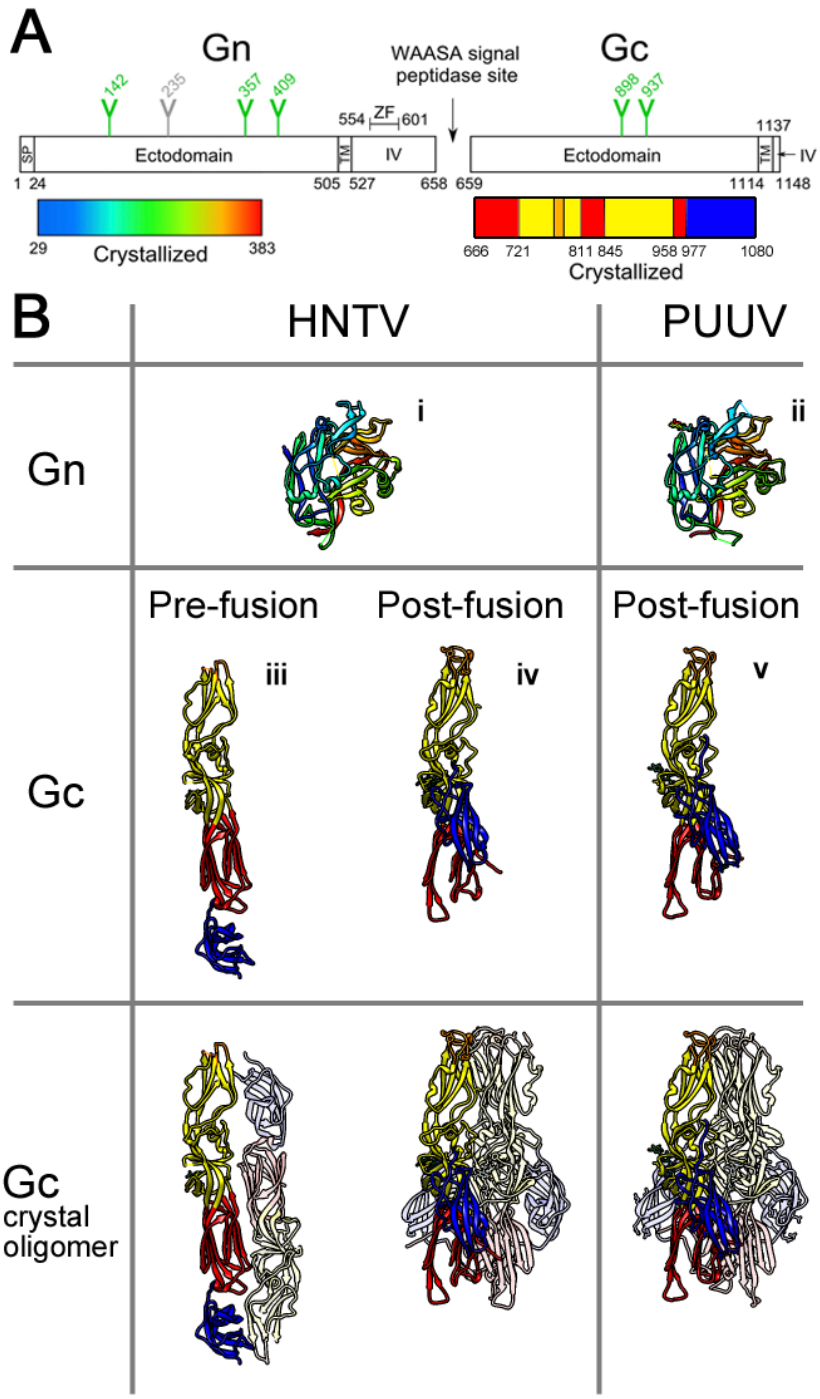


Figure 4.2 Hantavirus glycoprotein ectodomains studied by X-ray crystallography
A) A domain schematic of the Puumala virus glycoprotein precursor including the signal peptide (SP), ectodomain, transmembrane domain (TM), intravirion domain (IV), zinc finger (ZF), and WAASA signal peptidase cleavage site. N-linked glycosylation sites are displayed in green (and an additional site present in Hantaan virus but not Puumala virus is displayed in grey). Below, the portions of Gn and Gc included in the Puumala virus crystal structures are displayed including the colour scheme used in B. Note that the boundaries used for the Hantaan virus structures are similar though not identical to the Puumala virus boundaries. Figure adapted from [9]. B) Ribbon depictions of the currently available hantaviral crystal structures from Hantaan virus (HNTV) and Puumala virus (PUUV). Below each Gc ectodomain structure, the oligomeric form in which the protein crystallised is displayed with muted colours. i) HNTV Gn; PDB-5OPG. ii) PUUV Gn; PDB-5FXU. iii) HNTV Gc in the pre-fusion conformation. This was crystallised as a dimer; PDB-5LJY. iv) HNTV Gc in the post-fusion conformation. This was crystallised as a classic post fusion trimer; PDB-5LJZ. v) PUUV Gc in the post-fusion conformation. This was also crystallised as a post fusion trimer; PDB-5J81.

Despite the large amount of structural data available, the precise location and orientation of the Gn and Gc glycoproteins remains unknown. A fitting for the Gn in the existing cryo-EM map has been reported previously, in the four membrane distal lobes, but there are two equally likely orientations that cannot be distinguished at the current resolution [9] It is also not clear whether Gc glycoproteins are contained completely within a spike or bridge between adjacent spikes. One of my aims in this project is to improve the cryo-EM map in order to unambiguously determine how the crystal structures should be docked into the map.

In contrast to *Hantaviridae,* there is much less structural information about the surface glycoproteins of the nairovirus family. There are no crystal structures available of nairoviral glycoprotein ectodomains and at the beginning of this project, no microscopy data beyond basic negative staining confirming their pleiomorphic nature [82]. Sequence comparisons do suggest that the Gc glycoprotein will be similar to other class II fusion proteins but nothing else is known [126].

4.2. The surface glycoprotein array of *Nairoviridae*

4.2.1. Foreword

At the beginning of this project a sample of Hazara virus, a member of *Nairoviridae*, was provided by Prof Hewson, (Public Health England) and then purification, by ultracentrifugation through sucrose cushion, and grid preparation was performed by Dr Li, (Strubi). This sample was used to collect a small dataset that was used to generate the initial reconstruction presented here (4.2.2). The concentration of virus on the grid required improvement for the collection of a larger dataset so the remainder of this section details the attempts made to improve the quality of the sample (4.2.3).

4.2.2. Cryo-ET reconstruction of the Hazara virus surface glycoprotein array

Subtomogram averaging projects typically require the presence of an initial reference in order to start the refinement, but in the absence of any structural information about the Nairovirus surface, an initial reference was made by manually picking topviews of the virus. This was then used for subsequent template matching of the whole virion. The small size and lack of symmetry of these glycoprotein spikes makes this a particularly challenging target for subtomogram averaging. The final reconstruction required the development of a script to identify regular patches of spikes on the virion surface as a control to eliminate false positives (see section 3.5).

The arrangement of glycoproteins on the nairoviral surface was solved to a resolution of 21.6 Å based on ~4500 subtomograms (Figure 4.3). This resolution is far too low to build atomic models and there are currently no crystal structures for nairovirus glycoprotein ectodomains. However given that the class II fusion fold is highly conserved across multiple viral families, the shape of the Gc can be confidently predicted [126]. The resolution of the map is still too low to unambiguously fit a Gc crystal structure, but a tentative fitting is displayed in Figure 4.3 of a Phleboviral Gc [127].

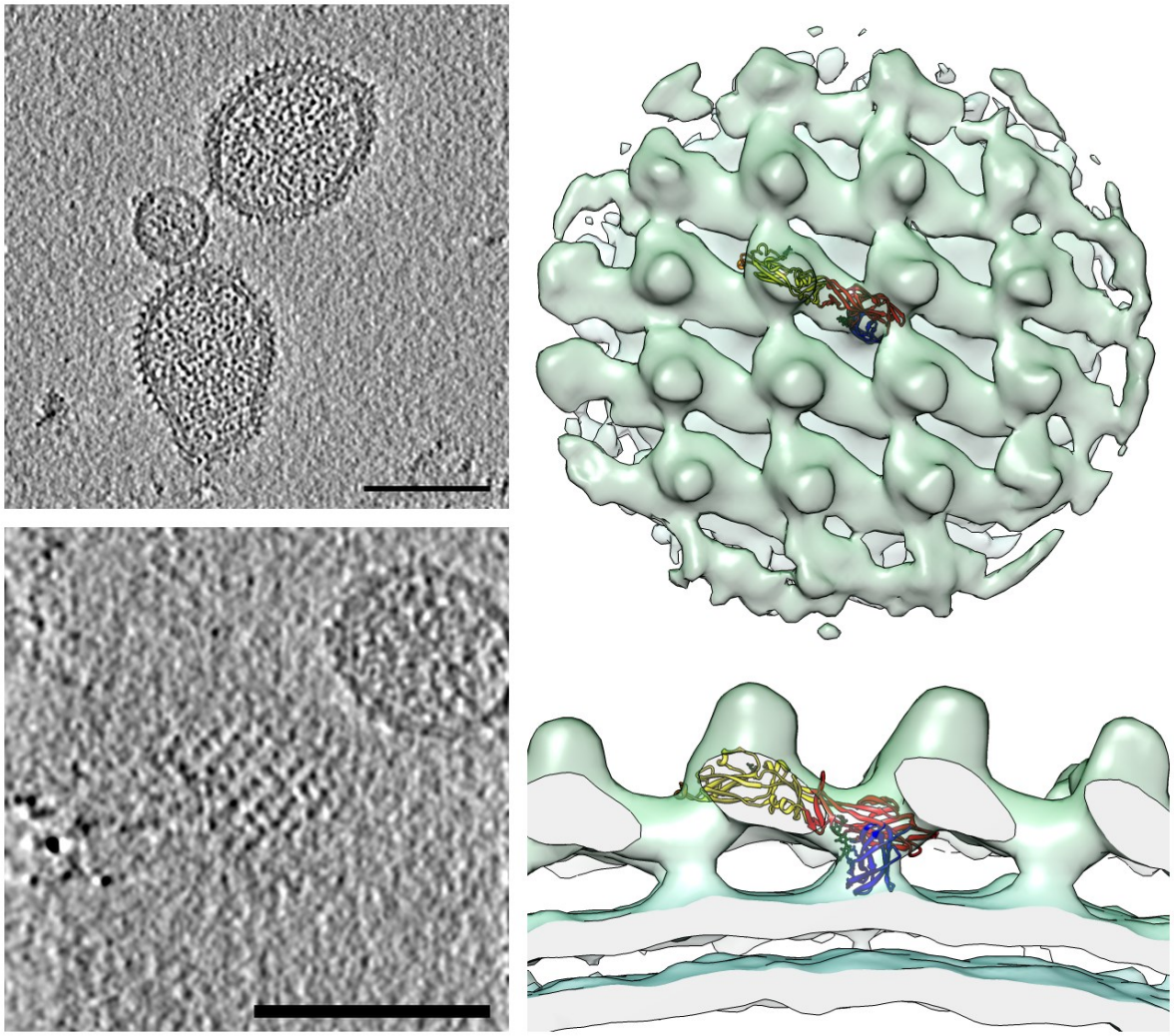


Figure 4.3 The glycoprotein array of Hazara virus
A) A slice through a tomogram of Hazara virus from the Nairovirus genus. Scale bar = 100nm. B) A magnified view of the array on the top of the virus. Scale bar = 100nm. C) A Top-view of the Hazara glycoprotein array reconstruction filtered to 21.6 Å. D) A zoomed side cut-away view of the structure in C to reveal the membrane contact points. A tentative manual fitting of a phleboviral Gc structure (from Rift Valley fever virus, PDB-4HJ1) shows a possible location of the nairoviral Gc coloured light green. The membrane is coloured blue and the possible location of Gn density is coloured dark green.

4.2.3. Hazara virus sample optimisation

The initial sample for Hazara virus collection was very low in concentration and not ideal for the collection of a large tomography dataset. In order to improve the resolution of the reconstruction and collect more data, new samples must be prepared which requires growing the virus in our own lab.

Initial attempts to grow the virus using SW-13 cells (both a kind gift by Roger Hewson, Public Health England) were successful. However, the viral titres, as measured by

plaque assay, were low. Figure 4.4 shows a test in which infections with a range of MOIs were tested and supernatant harvested at both 2 days post infection and 3 days post infection. This suggests that an MOI of 0.001 to 0.005 is ideal but this still only gives a maximum titre of ~5.5 x10⁵ PFU/ml which would be insufficient for cryo-EM grid preparation. The SW-13 cells used to grow the virus also gave inconsistent results between different viral preps so I then sought to find an alternative cell type to boost titers. Figure 4.4 displays the results of a small scale test in which a range of different cell types were seeded into a 6 well plate and infected with Hazara virus at 0.005 MOI. The viral titre of each was measured by plaque assay, revealing a large increase from the BSRT7 cell line so this was chosen for further Hazara virus production.

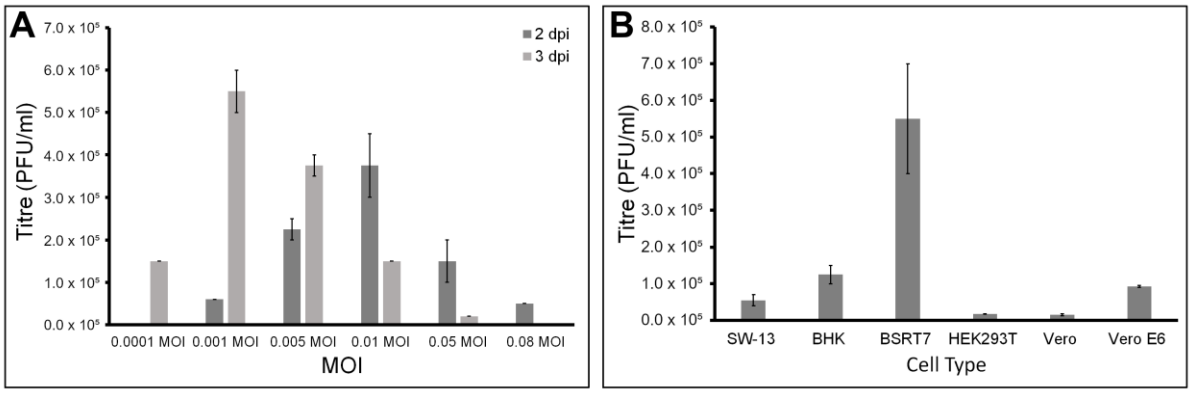


Figure 4.4 Hazara virus titre optimisation
Plaque assays performed to assess the production of Hazara virus. Experiments were performed in duplicate which are displayed as error bars. A) An experiment in which SW-13 cells grown in T175cm2 flasks were infected with HAZV at a range of MOI. The media was collected and replaced at 2dpi and collected again at 3dpi from the same flask. B) An experiment in which various cell types in a 6 well plate were infected with HAZV at an MOI of 0.005.

Throughout these tests, attempts were made to concentrate and purify small batches of viral supernatant which all failed. In one such test, 20ml of viral supernatant which had a measured titre of ~6 x 10⁵ PFU/ml was pelleted through a cushion of 20% optiprep in PBS buffer. The resulting pellet was re-suspended in 75µl which one would expect to result in a large (267x) increase in titre but instead the titre was significantly reduced to ~3 x10³ PFU/ml. In addition to this, any attempts made to band the virus on sucrose/optiprep density gradients resulted in no visible bands to harvest.

Around this time, a new publication was released entitled “Potassium is a trigger for conformational change in the fusion spike of an enveloped RNA virus” [128]. This study included low resolution cryo-electron tomography reconstructions of the Hazara virus glycoproteins as well as evidence that high salt concentrations are sufficient to trigger dramatic conformational rearrangement of the spikes. Their tomography data shows the existence of tubular projections from the viral surface that extend out towards co-purified vesicles after exposure to high salt. While this study focuses on K^+ ions as the fusion trigger, it is notable that Na^+ ion concentration is also kept low in these experiments at 20mM.

The discovery that Hazara virus is sensitive to high salt conditions may provide an explanation for the difficulty I had experienced in my purification attempts using a PBS buffer. Indeed, it was then possible to concentrate the virus using 100kDa cut-off filters and overlay onto sucrose gradients using a buffer containing 10mM Tris and 20mM NaCl. A visible band could be harvested and pelleted to maximise the virus concentration. Figure 4.5 A displays a micrograph showing the presence of virions. However the virions on this grid were rare and appeared clumped together. A subsequent purification was performed with a larger starting volume of viral supernatant and a modified protocol using techniques developed in my work with Hantaviral VLPs (see section 2.3.6 for details). The salt concentration used in this prep was also slightly increased towards a more physiologically relevant concentration of 70mM NaCl in an attempt to reduce the clumping observed on the grids. SDS-PAGE analysis confirms that the sample is relatively pure with bands that approximately correspond to the expected MW of the viral proteins (Figure 4.5 D). Many Hazara virions can be seen on the grid but aggregation of the virus is still a problem (Figure 4.5 B-C).

Tomographic slices confirm the presence of some Hazara virus spikes but there appears to be significant structural heterogeneity at the virus surface (Figure 4.5 C). Indeed, attempts to perform subtomogram averaging failed to improve the resolution over the existing map. A qualitative assessment of the refined coordinates confirmed the presence of

very little ordered lattice on the virus surface. This highlights the need for further optimisation of the purification strategy, buffer conditions and grid preparation methods to obtain virions with a regularly ordered lattice for structural determination. This has not yet been possible due to time limitations and the limited availability of microscopes capable of screening BSL-2 viruses.

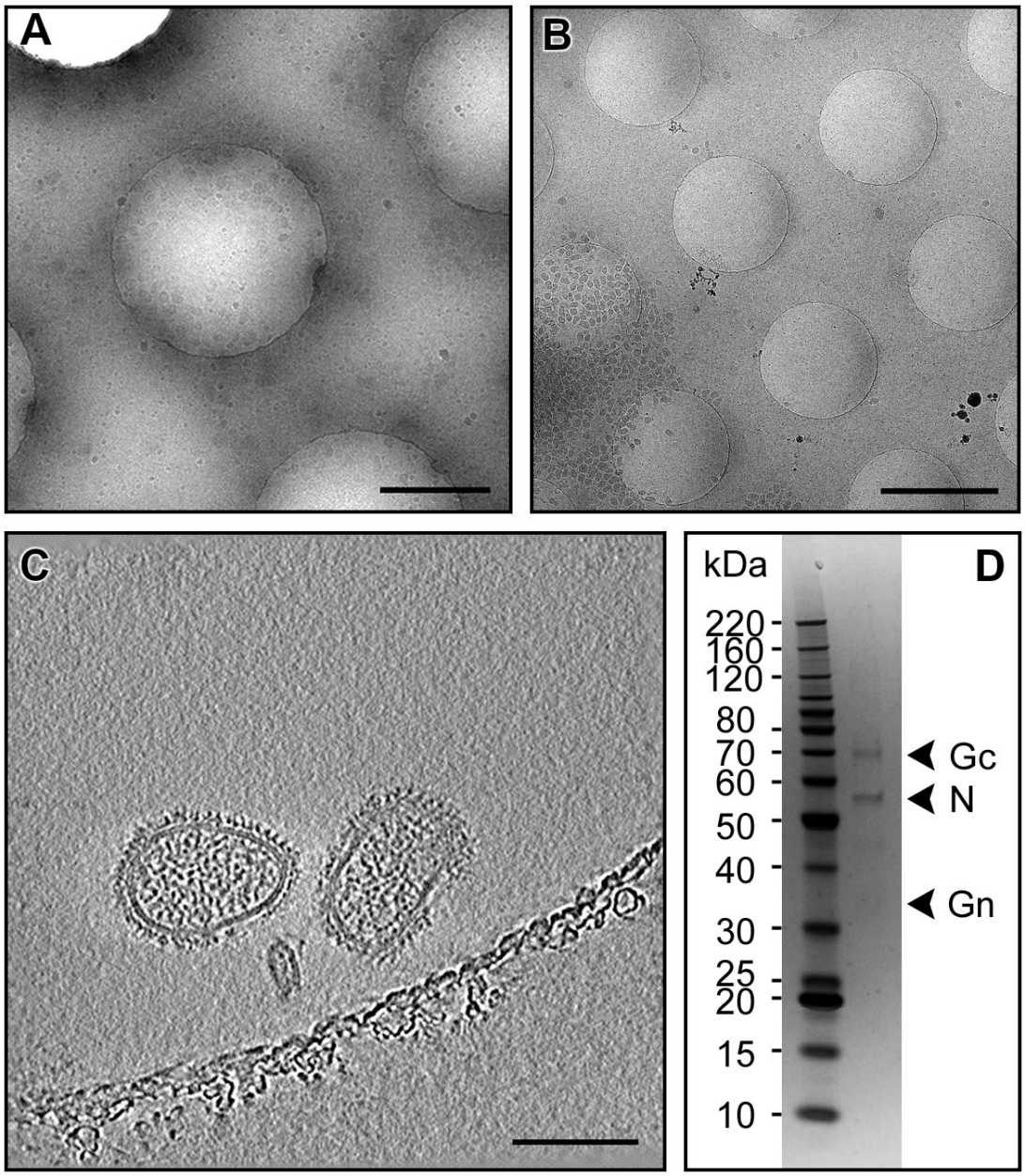


Figure 4.5 Cryo-EM sample optimisation of Hazara virus
A and B) Two representative views of cryo-EM grids from different preparations of HAZV which display aggregation of the virus. Scale bar = 1µm. C) A tomographic slice from the grid displayed in B to demonstrate the disordered array of glycoproteins at the surface. Scale bar = 100nm. D) An SDS-PAGE from 5µl of the sample shown in B and C. The viral proteins possibly associated with each of the bands from the viral proteins are indicated. Gc ≃71kDa. Gn ≃34kDa. N protein ≃ 54kDa.

4.2.4. Discussion and future directions

The reconstruction presented here gives a first look at the array of glycoproteins on the Hazara virus surface. While the current resolution is too low to unambiguously model the precise locations of the viral glycoproteins, it is possible to speculate that this lattice represents an array of heterodimers in which the lateral tubular density represents the Gc glycoprotein. The much smaller Gn (~34kDa) may then make up the globular density sitting on top of the Gc, which may serve to protect the fusion loops of the Gc as has been observed in other viruses [8], [52]. It must be stressed that such a description is highly speculative until a much higher resolution map is achieved.

A detailed comparison of the map presented here and the map presented in [128] is not possible as those reconstructions have not been deposited in the Electron Microscopy Data Bank. The broad shape of the spike appears visually similar to that presented here but one key difference is the decision by Punch *et al*. to impose C4 symmetry during the refinement. While the array of glycoproteins does form a square shaped lattice, the data presented here suggests that the individual spikes do not exhibit any symmetry.

Ultimately, a more detailed description of the nairovirus surface will have to wait for an improved resolution reconstruction. Attempts to grow Hazara virus were successful but it is clear that this virus is fairly fragile and susceptible to the harsh purification strategy or grid preparation method employed. Even the original sample prepared by Dr Li that was used for the initial reconstruction appeared to contain regions of disorder on its surface although significant regions of intact lattice were present. In the future, more work is required to optimise the ideal conditions for sample purification and grid preparation. Alternatively, it may be possible to find conditions that dramatically improve viral titre, and therefore require a more gentle purification strategy. Despite many attempts, the viral titres I achieved were typically on the order of ~1 x10^6 PFU/ml so simple pelleting of ~20ml of supernatant would not yield a sufficiently concentrated cryo-EM grid. Ultimately, it would be beneficial to find a solution to perform cryo-EM with low titre virus samples without any purification as it is likely

that harsh purification methods impact virus morphology [129]. This may be possible through the use of antibody based affinity cryo-EM grids to capture virus directly from cell culture medium [130].

4.3. The surface glycoprotein array of *Hantaviridae*

4.3.1. Sample preparation

The hantavirus work began with Tula virus, with an initial goal of obtaining a higher resolution map that could facilitate unambiguous fitting of the Gn and Gc ectodomain crystal structures into the cryo-EM density. Previous Tula virus preps in the lab had been obtained from collaborators and were not an ideal concentration for large scale data collection. Therefore, Tula virus growth and purification was optimised in our lab and new grids prepared with a high concentration of virus. This enabled the collection of a number of tomographic tilt series each with multiple virions per tomogram as shown in Figure 4.6. These tilt series were aligned and subtomogram averaging performed in dynamo to produce an updated reconstruction of the hantavirus surface glycoproteins.

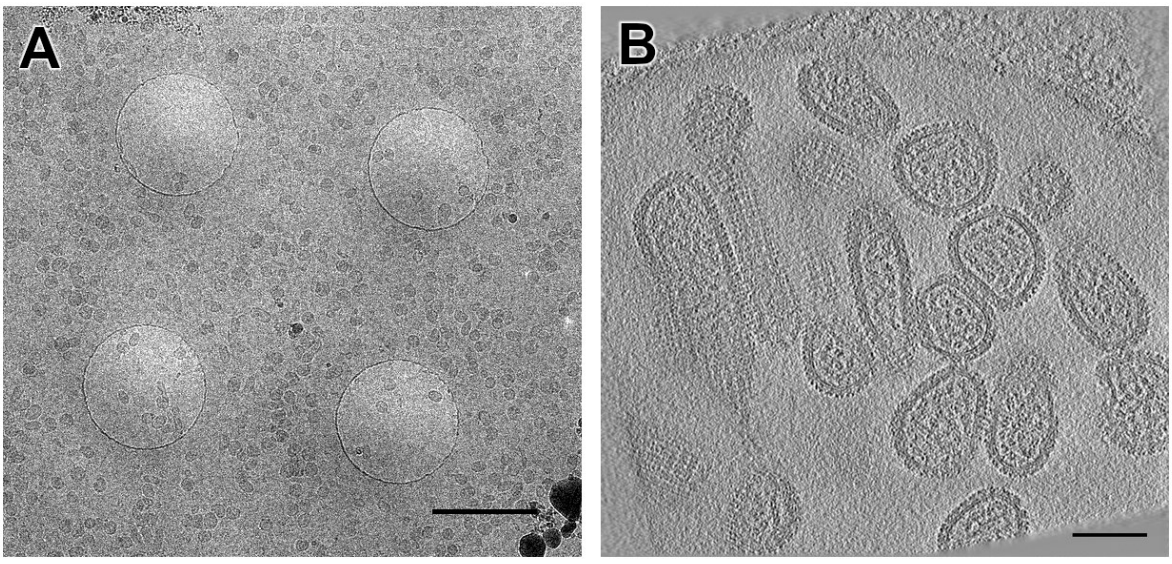


Figure 4.6 Optimised TULV preparation for cryo-EM
A) A representative image of a highly concentrated sample of TULV virions for efficient data collection. (Scale bar = 1µm) B) A slice through a tomogram containing a large number of TULV virions. This demonstrates the diverse morphologies adopted by hantaviruses. Scale bar = 100nm)

4.3.2. Handedness of the map

The new map appeared similar to the existing map (EMD-3364) but with reversed handedness (Figure 4.7).

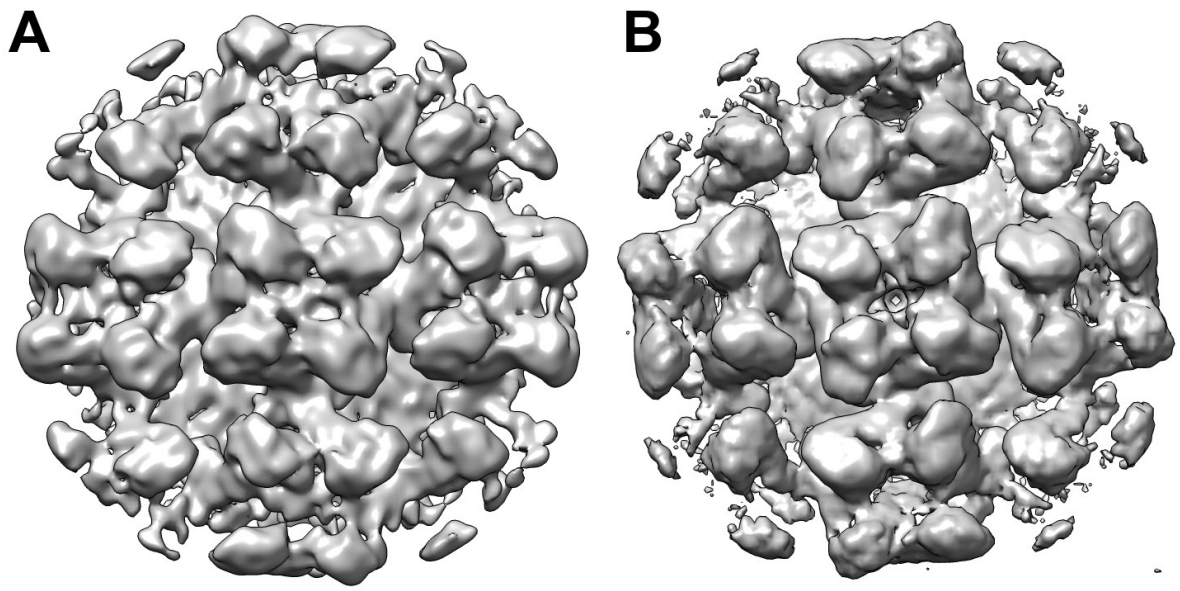


Figure 4.7 The handedness of the TULV map differs from that reported in EMD-3364
A) An isosurface depiction of EMD-3364. B) A new isosurface depiction of TULV filtered to 11.4 Å.

An obvious handedness error could not be found in this dataset. A new sample was prepared and data collected on a new sample with Tula virus mixed with human papilloma virus (HPV) pseudoviruses (grids prepared and data collected by Sai Li and the HPV pseudoviruses were a kind gift from Prof. Kay Grunewald). Refinements of HPV resulted in a map with a T=7dextro arrangement of capsomers when using both a reference of the correct hand and featureless sphere as a reference (Figure 4.8). This confirms the handedness of the data is correct. Refinements were then carried out of Tula virus from the same sample. Tula virus refinements produced a handedness flipped relative to that reported in EMD-3364 regardless of the handedness of the initial reference used. This confirms that the handedness of the new map is correct (Figure 4.8). As a result, the EMD-3364 entry was labelled obsolete and replaced with its hand-mirrored version (EMD-4867).

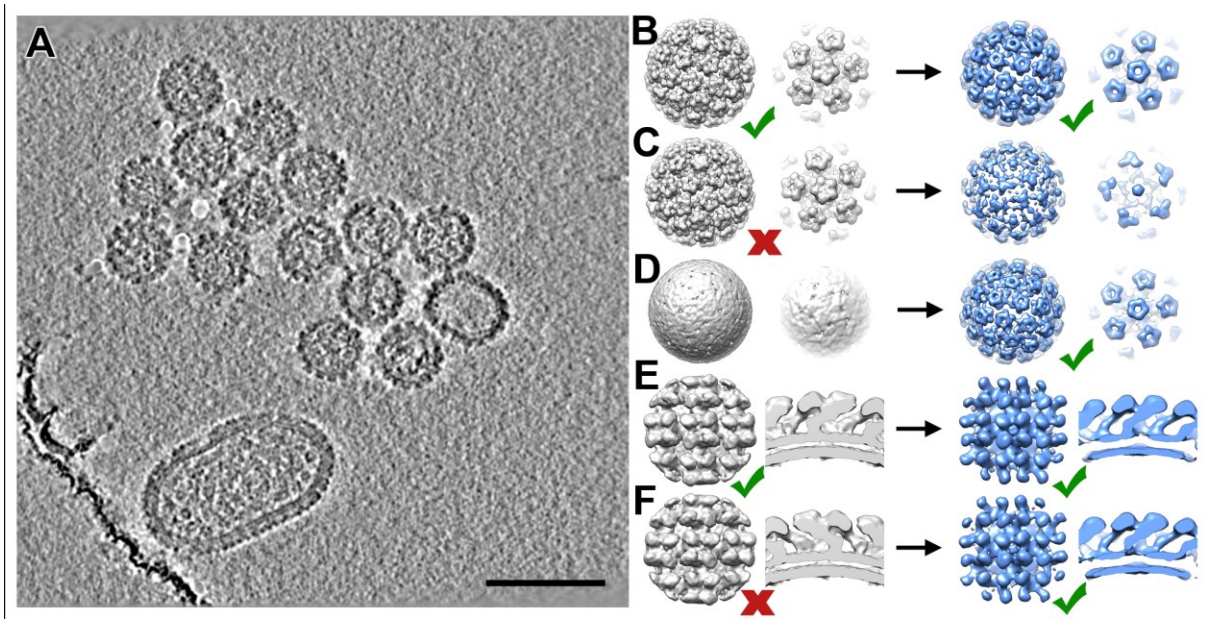


Figure 4.8 Confirming the handedness of the Tula virus map
A) A tomogram slice displaying TULV and HPV in the same sample. Scale bar = 100nm. B-F) Subtomogram averaging performed using a range of initial references displayed on the left in grey to produce the resulting reconstructions, in blue on the right. B-D) HPV refinements; Isosurfaces are displayed as a top view (left) and a magnified view of the pentons (right). E-F) TULV refinements; Isosurfaces are displayed as a top view (left) and a magnified cut-away side view (right). B) Reference = EMD-5991. C) Reference = reversed hand EMD-5991. D) Reference = a spherical reference. E) Reference = reversed hand EMD-3364. F) Reference = EMD-3364. Maps of the wrong hand are indicated with a cross and correctly handed maps indicated with ticks.

4.3.3. A pseudo-atomic model of the Hantavirus glycoprotein spike

With the new higher-resolution map of the Tula virus surface with correct hand available, it is now possible to unambiguously fit the Gn and Gc ectodomain crystal structures. At the same time an ANDV Gn-Gc complex crystal structure became available through collaboration (unpublished, Dr Guardado-Calvo and Prof. Rey). This crystal structure fits well, as a rigid body, into the density to confirm the overall architecture of the hantaviral envelope (Figure 4.9 A-D).

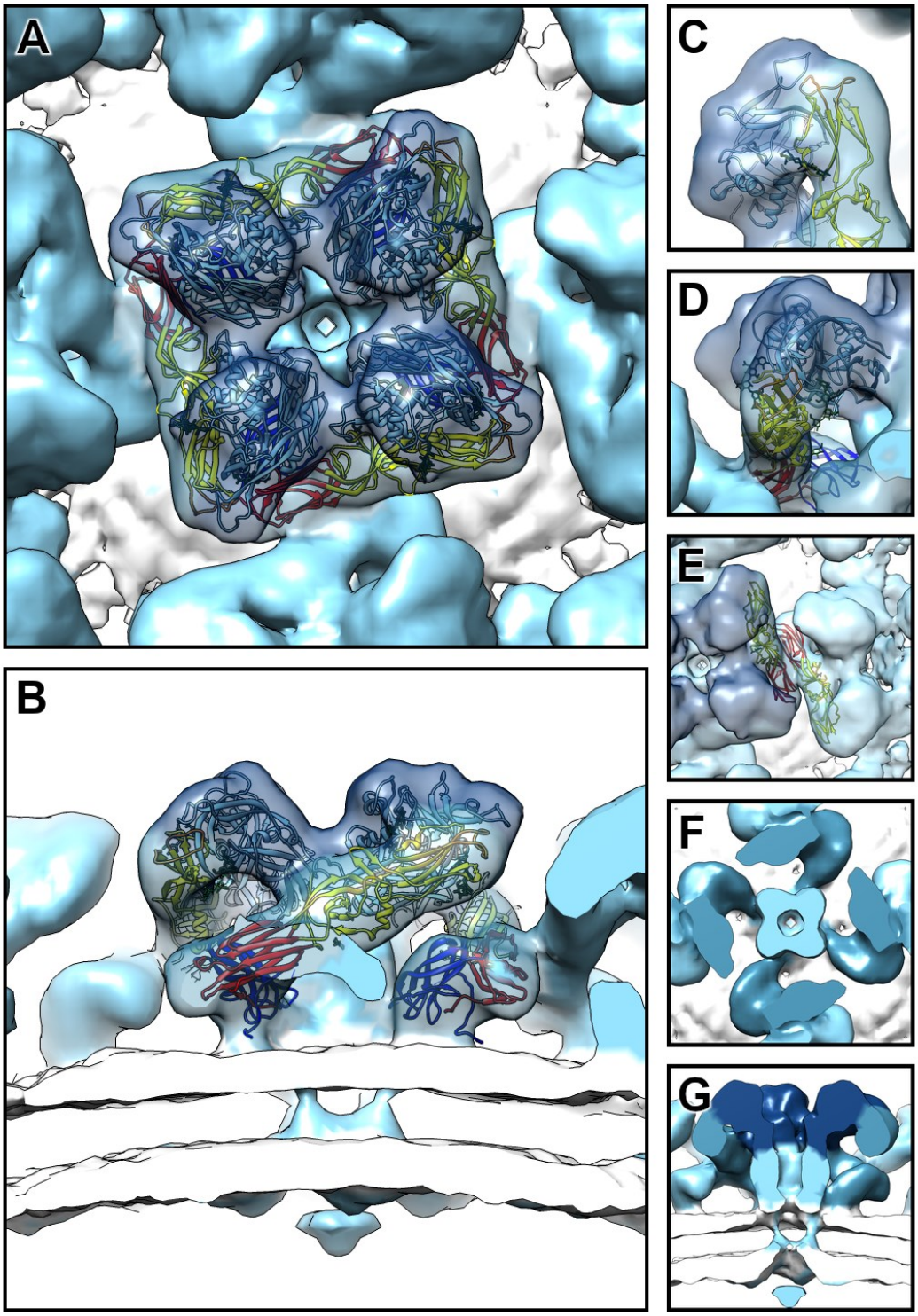


Figure 4.9 A pseudo-atomic model of the hantaviral surface glycoprotein array
A reconstruction of the TULV glycoprotein array filtered to 11.4 Å. Fitted in the map is an ANDV Gn-Gc heterodimer crystal structure (kindly supplied by Dr Guardado-Calvo and Professor Rey, Institut Pasteur). Fitting was carried out with the fitmap function in chimera assuming C4 symmetry. The Gn is coloured light blue and the Gc coloured using the standard class II fusion protein colouring scheme (red, yellow and blue for domains I, II and III respectively) A) A topview. B) A cut-away side view. C) A view demonstrating the partial shielding of the Gc fusion loops (depicted in orange). D) A view demonstrating the 3 surface exposed N-linked glycans (2 on Gn and 1 on Gc). E) A fitting of two copies of domains I and II of the pre-fusion HNTV Gc (PDB-5LJY). The contacts between these molecules observed in the crystal packing fit reasonably well into the map. F-G) Views of the stalk region that is absent from the available crystal structures. F) A cut-away topview. G) A cut-away side view.

One notable feature of this fitting is the Gc homodimer interface. In the available crystal structure for HNTV Gc, the crystal packing of the Gc revealed a dimer interface (Figure 4.2 B and Figure 4.9 E) [16]. It has further been suggested by Bignon *et al.* that this dimer interface is present in the native virus and is regulated by temperature [121]. The dimer interface can be fitted into the map reasonably well with only a slight discrepancy in the angle of the two proteins. It is conceivable that this slight discrepancy suggests that the Gc is held under tension in the native hantaviral lattice, and prevented from adopting the energy-minimal contact observed in the crystal packing. Hantaviruses lack a matrix protein, so the force required for membrane curvature while budding from cellular membranes is presumably mediated entirely by the surface glycoproteins. This fitting suggests that the Gc homodimer represents a key interface in this process.

Previous fittings of the Gn suggested it occupied the whole globular lobe at the top of the spike [9]. With the new fitting it is now clear that the Gn is further tilted inwards to the centre of the spike with the tip of the Gc occupying most of the distal portion of the lobe. In many other class II fusion protein displaying viruses, one of the purposes of the Gn protein (or similar) is to shield the hydrophobic fusion loops at the tip of the Gc until the virus is ready and triggered to fuse with the target membrane [8]. It is clear from this fitting and the heterodimer crystal structure that the tip of the Gc is partially shielded by the Gn (Figure 4.9 C). There is an unresolved, presumably flexible, loop in the PUUV ectodomain crystal structure which appears to be stabilised in the heterodimer structure by the fusion loops of the Gc [9]. It must however be noted that this shielding appears to be only partial. This suggests a new model by which the Gn stabilises the Gc by forming a tetramer that binds to the Gc and acts as a scaffold to sterically block fusion related movements until it is triggered.

This reconstruction of the Tula virus glycoproteins reveals some features that are absent from the available ectodomain crystal structures. The Gn ectodomain crystal structures lack approximately 122 residues from their C-terminus before the transmembrane region [9]. These residues can be seen in the reconstruction descending down in four pillars

of density toward the membrane (Figure 4.9 F&G). There does appear to be a region of this density part way down these pillars where the Gn density joins together, this is possibly part of the interface responsible for tetramerisation of the Gns. As the density gets close to the membrane it appears to bend outwards to meet the C-terminus of the Gc and both can be seen passing through the lipid bilayer in four stalks of density. Higher resolution is needed to confirm the precise conformation of the Gn C-terminal residues and the transmembrane regions of the two glycoproteins.

This fitting suggests that each hantaviral spike is a self-contained unit comprised of four Gn glycoproteins and four Gc glycoproteins. These units are then arranged in a lattice that is maintained by Gc heterodimers (Figure 4.10 A). However, it is clear that a square pattern cannot fully cover the surface of a sphere so break points must exist in the lattice. These break points can be seen when plotting the 3D positions of the subvolumes found by template matching. By manually picking out, and aligning these subvolumes an average was obtained of a hantaviral spike without a neighbour (Figure 4.10 B). This acts as a further validation of the fitting, confirming that a single glycoprotein does not span across two spikes.

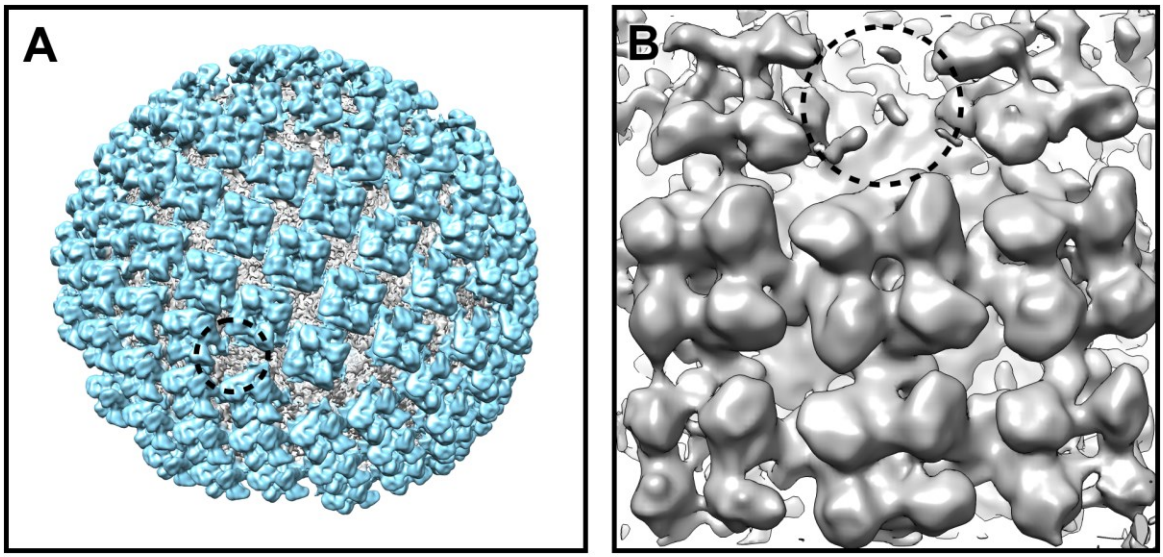


Figure 4.10 The hantaviral lattice contains regions of exposed membrane
A) A depiction of a simulated TULV virion. This is generated by plotting back the TULV spike reconstruction using the refined coordinates of one particular virion in the TULV dataset. The square lattice is discontinuous with obvious break points as indicated. B) A reconstruction generated from the TULV dataset demonstrating a spike next to exposed membrane. This was generated from a dataset refined with C4 symmetry. Obvious break points in the lattice were manually picked out and aligned, followed by reconstruction with C1 symmetry.

4.4. Tula virus conformational changes induced by exposure to low pH

The hantaviral class II fusion protein, Gc, is expected to undergo a conformational change in the low pH environment of the endocytic pathway. To gain insights into the structural changes that occur under these conditions the virus was incubated at pH 5 for 1 minute prior to plunge freezing. Figure 4.11 demonstrates the large conformational change resulting from exposure to low pH. The regular lattice of glycoproteins on the surface is completely disrupted making detailed structural determination by subtomogram averaging challenging. This can be demonstrated by 2D classification which is able to discern regular structure in the pH 7 projection images but fails to produce any ordered structure for the pH 5 sample (Figure 4.11 E&F). An averaged profile of the density distribution perpendicular to the membrane is shown in Figure 4.11 G which suggests a slight average rise in density away from the membrane.

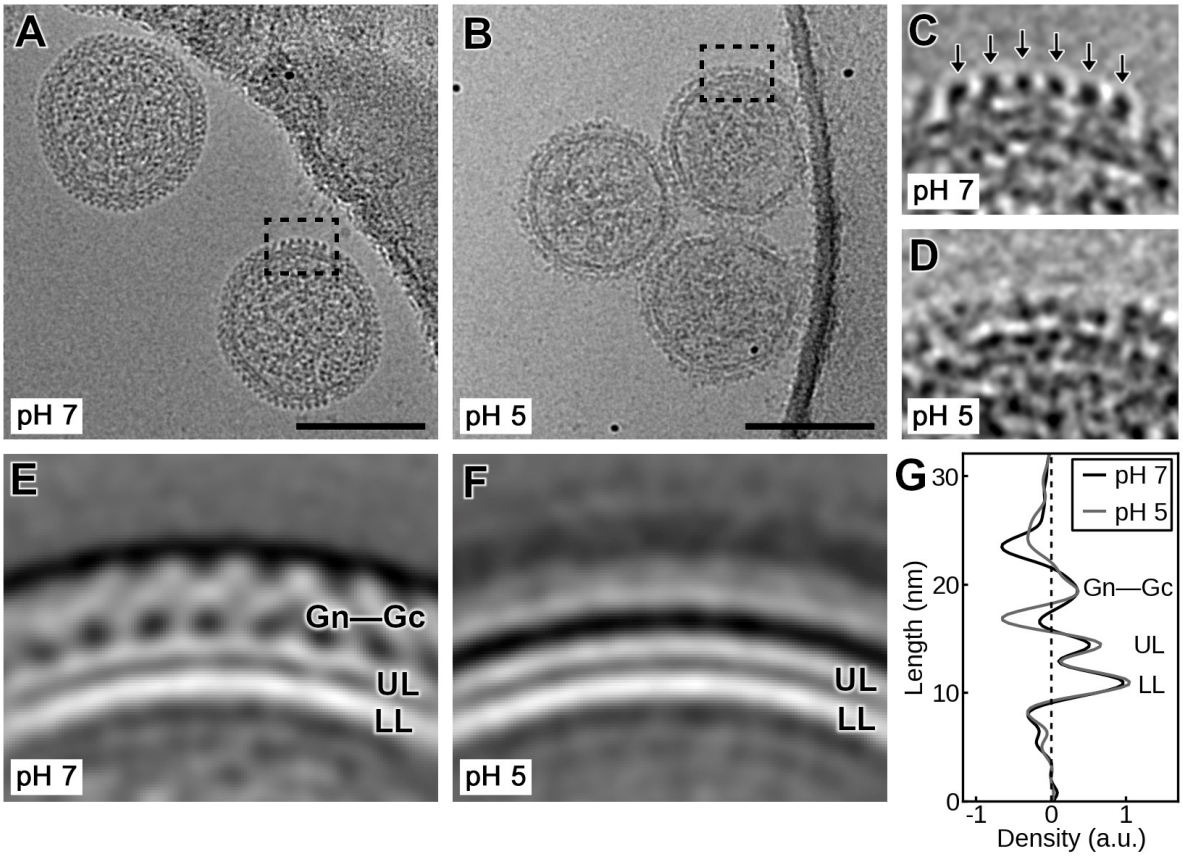


Figure 4.11 Metastability of the Tula virus envelope after exposure to low pH
A) and B) Representative cryo-electron micrographs of Tula virions (TULV) after incubation at pH 7 and pH 5, respectively. Scale bar corresponds to 100 nm. C) and D) Close-up views of the areas indicated in panels A and B. Regular lattice indicated with arrows. Virions incubated at pH 5 do not display an ordered lattice. E) and F) Representative 2D class averages of the virion surface at pH 7 and pH 5. The image contrast has been inverted so that white corresponds to the virion. G) An averaged density profile of the images in panels E and F indicating the average distribution of density perpendicular to the membrane at pH 7 and pH 5. The location of the lower membrane leaflet (LL), the upper membrane leaflet (UL) and the Gn−Gc spikes are indicated. Figure adapted from [24].

4.5. Limitations of Tula virus as a surrogate

Tula virus is considered non-pathogenic in humans so can be easily handled in a BSL-2 facility. It is an excellent choice for the study of hantavirus ultrastructure as, at the resolutions attainable by subtomogram averaging, the reconstruction is likely similar for all members of *Hantaviridae*. This is evident from the large degree of similarity between the PUUV and HNTV Gn crystal structures [9], [24]. However, as significant differences exist in receptor and antibody-binding between different hantaviruses, TULV alone was not sufficient for conducting experiments on all of the planned receptor- and antibody-complex structures.

Puumala virus has a number of reported neutralising antibodies including two bank vole antibodies 4G2 and 5A2 which have an unknown mechanism of neutralisation [131], [132]. The Tula M segment, which encodes the surface glycoproteins, has 78.9% sequence identity and 91.0% sequence similarity to Puumala virus so it was hoped these antibodies would bind to Tula virus also. Fab fragments cleaved from a sample of 4G2 and 5A2 antibodies (a kind gift of Dr Hepojoki and Dr Rissanen) were mixed with Tula virus and cryo-EM grids prepared. Unfortunately, no additional density could be clearly seen for either 4G2 or 5A2. One possible explanation for this may be that too little fab fragment was incubated with each sample. To measure the affinity of the interaction an ELISA was performed with immobilised Tula virus. The results in Figure 4.12 B demonstrate that while the antibodies do bind, the affinity is relatively low as full saturation is not reached within the concentration range tested. A low affinity interaction is unsuitable for cryo-EM as there is a limit to how much protein can be added to the sample without obscuring the sample too much. One likely explanation for the apparent low affinity is that these antibodies are raised against Puumala virus, not Tula virus, so while similar, they are perhaps not quite similar enough. (Note that hantavirus antibody interactions are revisited in Chapter 6).

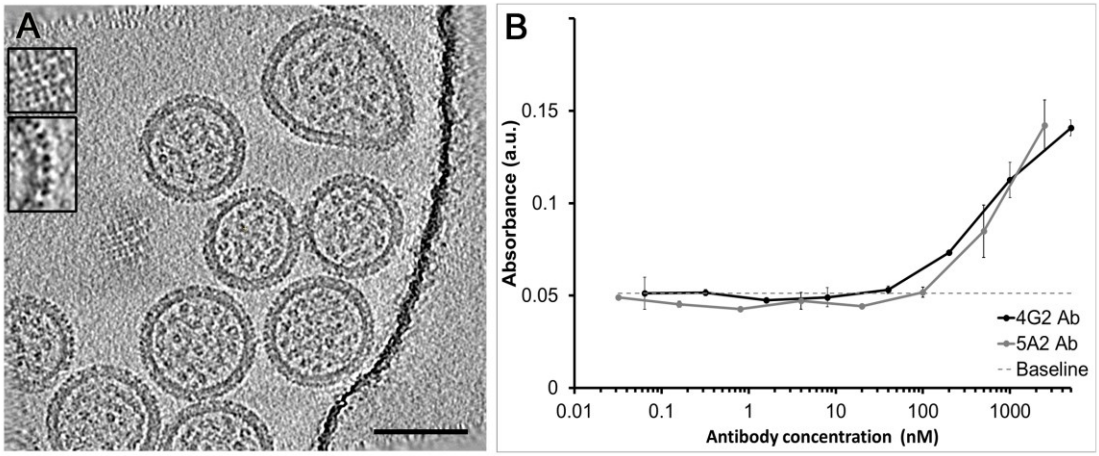


Figure 4.12 The limitation of Tula virus as a surrogate for antibody experiments
A) A slice through a tomogram of TULV incubated with the fab fragments of the anti-PUUV bank vole neutralising antibody 4G2 at a concentration of 0.5mg/ml. No extra density can be seen decorating the virus. Scale bar = 100nm. Magnified views of viral surface inset; upper = top-view, lower = side-view. B) An ELISA in which immobilised TULV was probed with 4G2 (black) and 5A2 (grey) antibodies. An analyte-free negative control well was included giving a background reading of 0.051 as indicated with a grey dashed line. A positive control was ran alongside using anti-PUUV Gn polyclonal sera (which also detects TULV) to confirm the presence of TULV (Supplementary Figure 8.1) Duplicate assay results are displayed as error bars.

In addition to this, there are some reports of differential receptor interactions between pathogenic and non-pathogenic hantaviruses. For example αVβ3 integrin is utilised for Puumala but not Tula infection . Because of these differences, it was decided that in order to gain insights into receptor binding and neutralisation of the pathogenic hantaviruses, it is important to find a way to study these viruses directly without a surrogate non-pathogenic virus like Tula [133], [134]. Some pathogenic hantaviruses have biosafety requirements too high for study in our lab such as Andes virus. However, Puumala virus, while human pathogenic, is possible to work with in a BSL-2 facility so this was an option. An alternative option is develop hantaviral virus-like particles VLPs. There are previous reports of successful hantaviral VLPs which have the added benefit of having no biosafety concerns [135]. This study suggests that simply transfecting a hantaviral glycoprotein precursor protein in HEK cells is sufficient to generate VLPs.

4.6. Puumala live virus production

To assess the feasibility of Puumala virus production a small scale test was performed to measure viral expression. These viruses do not form plaques in vero E6 cells so a simple plaque assay is not possible. It is however possible to detect vial titre with an immunofluorescence based focus forming unit (FFU) assay. The initial batches of Puumala virus were supplied by Dr Hepojoki (University of Finland) along with some advice on how to grow Puumala. For example, this virus in this cell line is very slow growing and takes up to 2 weeks to produce virus. The slow growing nature of Puumala virus is clear from a focus forming unit assay of Tula and Puumala virus performed in parallel (Figure 4.13). After 6 days each focus forming unit for Tula virus can be seen as a group of cells due to the infection spreading between cells from an initial infection (Figure 4.13 B). However, the Puumala virus sample after the same time, can only be detected in a single cell which has not yet spread to its neighbours (Figure 4.13 A). Figure 4.13 C shows a puumala plate that has been left for 11 days, by which time the infection has been able to spread to its neighbours. A number of tests were carried out with Puumala virus, including one in which

the virus was infected at a range of MOIs and incubated for 14 days (with media exchanges at 4,7,10 and 12dpi) (See supplementary Table 8.2). However, the best titres achieved for Puumala virus were approximately 5 x10^3 PFU/ml. As a result of this it was clear that, without some further optimisation, Puumala titres were not viable for cryo-EM structural studies.

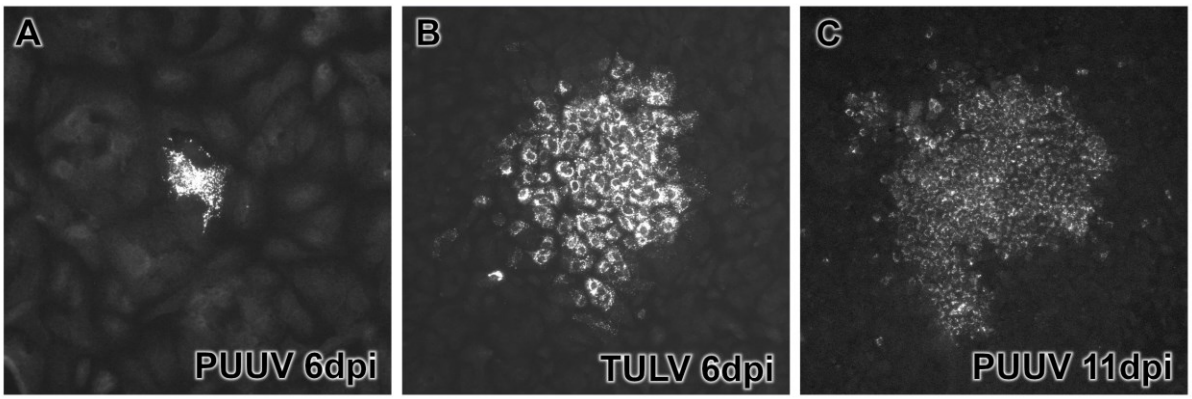


Figure 4.13 The slow growth of PUUV in VeroE6 cells
Representative images from a focus-forming unit assay with TULV and PUUV. A) PUUV at 6dpi. B) TULV at 6 dpi. C) PUUV at 11 dpi.

4.7. Hantaviral virus-like particle development

The development of a hantaviral virus-like particle system began with Puumala virus construct 1 (see 8.1.1). HEK 293T cells were transfected with the M segment of Puumala virus and the cell supernatant collected a few days later. The expression appeared to be poor but after concentration of the sample, by ultracentrifugation, VLPs could be detected and VLPs were visible by negative stain EM. Curiously, these VLPs were shown to be displaying a tubular morphology, frequently with globular heads Figure 4.14 A-B. Tubes of Tula virus had been reported previously and seen in amongst our own data but these were rare and clearly displayed hantaviral spikes (Figure 4.14 B) [50]. A helical form of the virus would be ideal for high resolution single particle electron microscopy so this was a promising target.

Subsequent work was then done to produce a more concentrated sample for cryo-electron microscopy. This was done by banding the VLP between layers of sucrose before

pelleting by ultracentrifugation from 50ml of media. Cryo-electron microscopy grids were then prepared using this sample and it was imaged by tomographic tilt series collection and single particle collection. The concentration of the sample was still too low for automated data acquisition but a small dataset could be collected manually. Inspection of the tomograms, revealed that the globular heads of the VLP did display tetrameric lattice as had been seen for Tula virus. However the tubular regions appeared disordered and segmented Figure 4.14 D-E. Any attempt at subtomogram averaging of these tubes failed as did 2D classification from the single particle images taken suggesting significant structural heterogeneity. Curiously, one tube was found with a very ordered pattern on its surface without classic hantaviral spikes, but this appeared to be extremely rare Figure 4.15 A-C. A subtomogram average of this tube could be obtained (Figure 4.15 D). However at such low resolution, and a sample size of one, it is not clear if this reconstruction is made up of hantaviral proteins or how they are arranged.

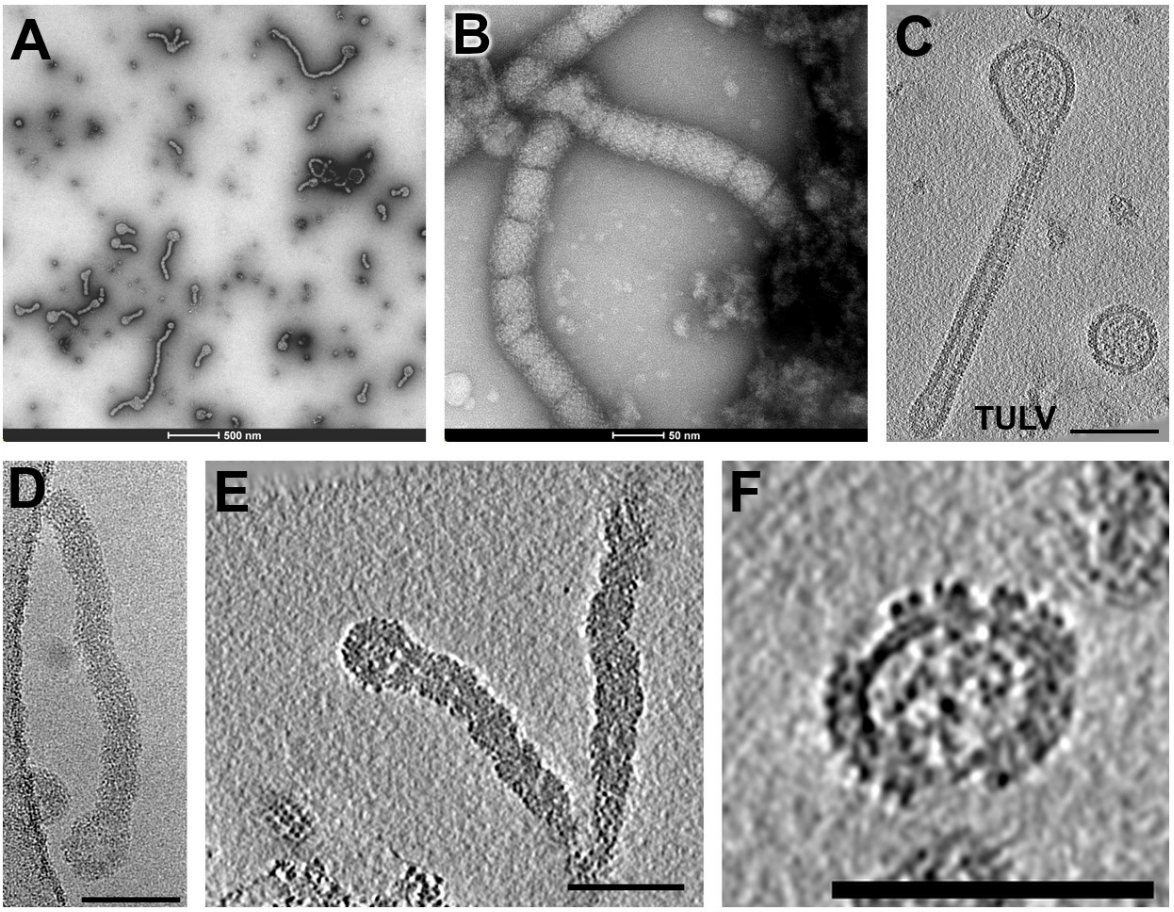


Figure 4.14 A disordered-tubular form of PUUV observed with a VLP system
A-B) Representative negative stain images of PUUV VLP exhibiting a tubular morphology. C) A tomographic slice of an ordered tube observed in our TULV dataset. D) A 2D projection image of a tubular PUUV VLP. E) A tomographic slice of a tubular PUUV VLP. A top view of a spherical VLP can be seen with ordered lattice. F) A tomographic slice of spherical PUUV from the same sample as E with clear evidence of classic hantaviral spikes. Scale bars = 100nm unless otherwise stated.

After it had become clear that the hantaviral tubes were not amenable to structure determination it became a priority to find a way to prevent the tube formation. One possibility was that it was induced by the purification. A sample of raw supernatant was imaged by negative stain EM and while extremely rare, a tubular hantaviral VLP could be observed ruling out this possibility (Figure 4.15 E). Another theory considered at the time was that the tubular VLPs were caused by incomplete cleavage between the Gn and Gc glycoproteins at the WAASA cleavage site, however both components could be observed separately in a western blot using poly-clonal anti-puuv Gn and Gc antibodies (a kind gift from Dr Hepojoki) (Figure 4.15 F). This leaves the possibility that the formation of tubes is an inherent quality of this particular Puumala virus construct.

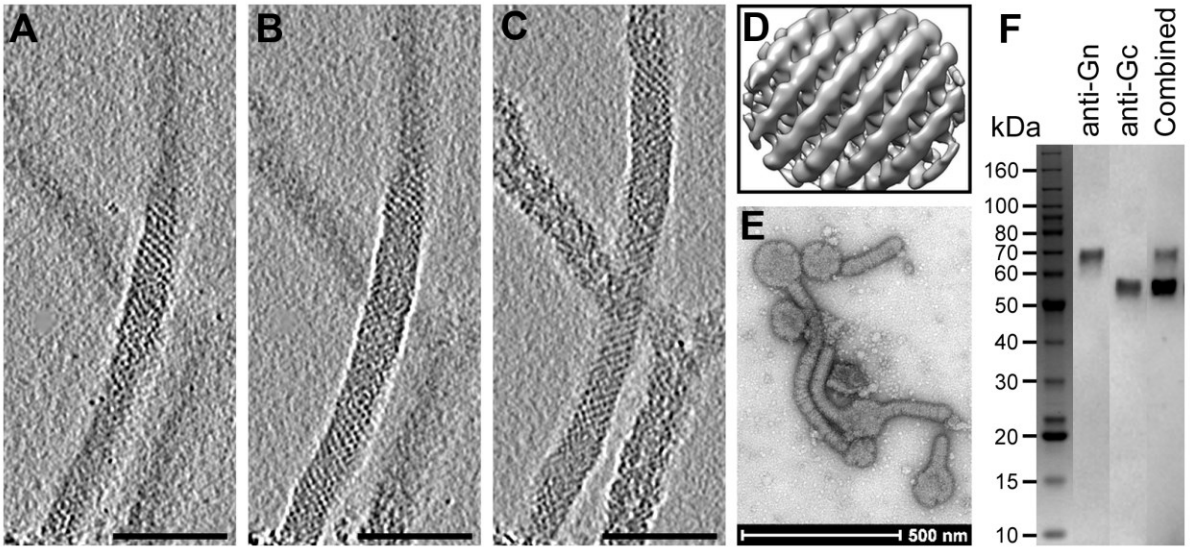


Figure 4.15 A curious ordered tube from a sample of PUUV VLPs
A-C) Tomographic slices through a single example of an ordered tube that does not appear to display classic hantaviral spikes. Scale bar = 100nm D) A very low resolution reconstruction obtained by STA of the tube in A-C only. E) A negative stain image of a collection of tubular PUUV VLPs observed in a cell supernatant sample before any purification attempted. F) A combined image of 3 western blots, run in parallel, of a sample of PUUV VLPs. The viral proteins were detected using rabbit-derived polyclonal anti-PUUV sera (a kind gift of Dr Hepojoki). The first lane using anti-Gn sera, the second using anti-Gc sera and the third with a combination of the two to demonstrate the presence of two separate bands.

New constructs were obtained for Hantaan virus and Andes virus M segments. These constructs included an internal mTurqoise tag to aid detection of the VLPs (Figure 4.16 A). Expression and pelleting of these samples were carried out as with Puumala and no tubes observed by negative stain EM (Figure 4.16 B). As with PUUV previously, a prep was prepared from 30ml of cell supernatant and by banding through sucrose and pelleting. Imaging by cryo-EM revealed that the VLP did appear to display the classic hantaviral spikes (Figure 4.16 C). However, for efficient data collection the concentration of VLP on the grid was far too low. Attempts were made to scale up the preparations to a larger volume of cell supernatant, from roller bottles, but these attempts resulted in an impure sample due to too much co-purified cell debris and FBS proteins.

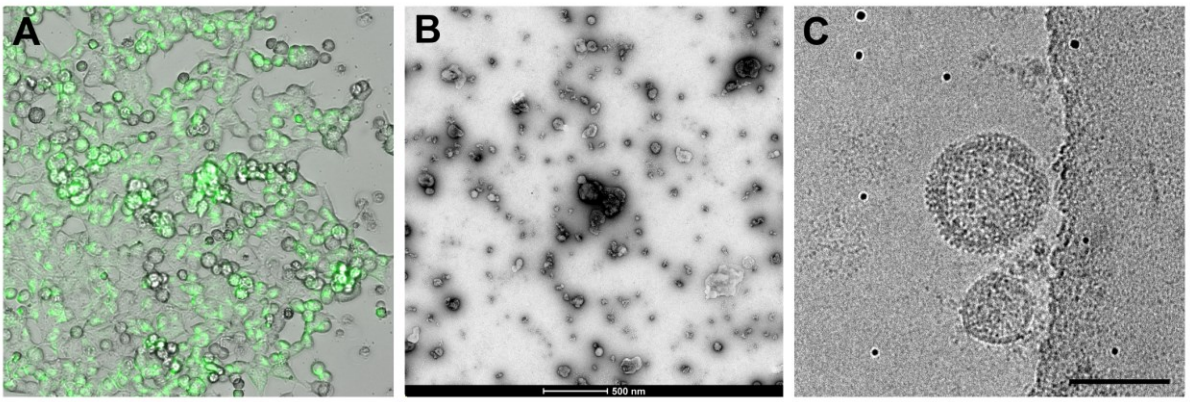


Figure 4.16 Continued development of ANDV and HNTV VLPs
A) A live cell fluorescent image of HEK293T cells transfected with the M segment of ANDV fused to a C-terminal fluorescent mTurq tag to demonstrate successful expression. B) A negative stain EM image of HNTV VLPs prepared similarly to those seen in Figure 4.14 A. A cryo-EM image of a HNTV VLP displaying classic hantaviral spikes. Scale bar = 100nm.

Purification of the VLPs proved difficult because they are of variable size and shape so do not form a tight band in a sucrose gradient (Figure 4.17 A). This means that large sections of the gradients have to be harvested so considerable amounts of cell debris and proteins from the fetal bovine serum get co-purified, making the prep too contaminated for cryo-EM. An ideal solution to this would be an affinity based purification. The C-terminus of the Gc is internal so the Histidine tag present there cannot be utilised. However the N-terminus of the Gn is external and may be relatively surface exposed. A series of HNTV VLP constructs were produced with N-terminal His-tags of varying lengths and with varying length linker peptides between the tag and the native VLP sequence. These new constructs expressed successfully as detected with an anti-HNTV Ab G2 (A kind gift from Dr Doores, KCL). However all attempts to bind the VLP to magnetic nickel beads for purification failed (Figure 4.17 B).

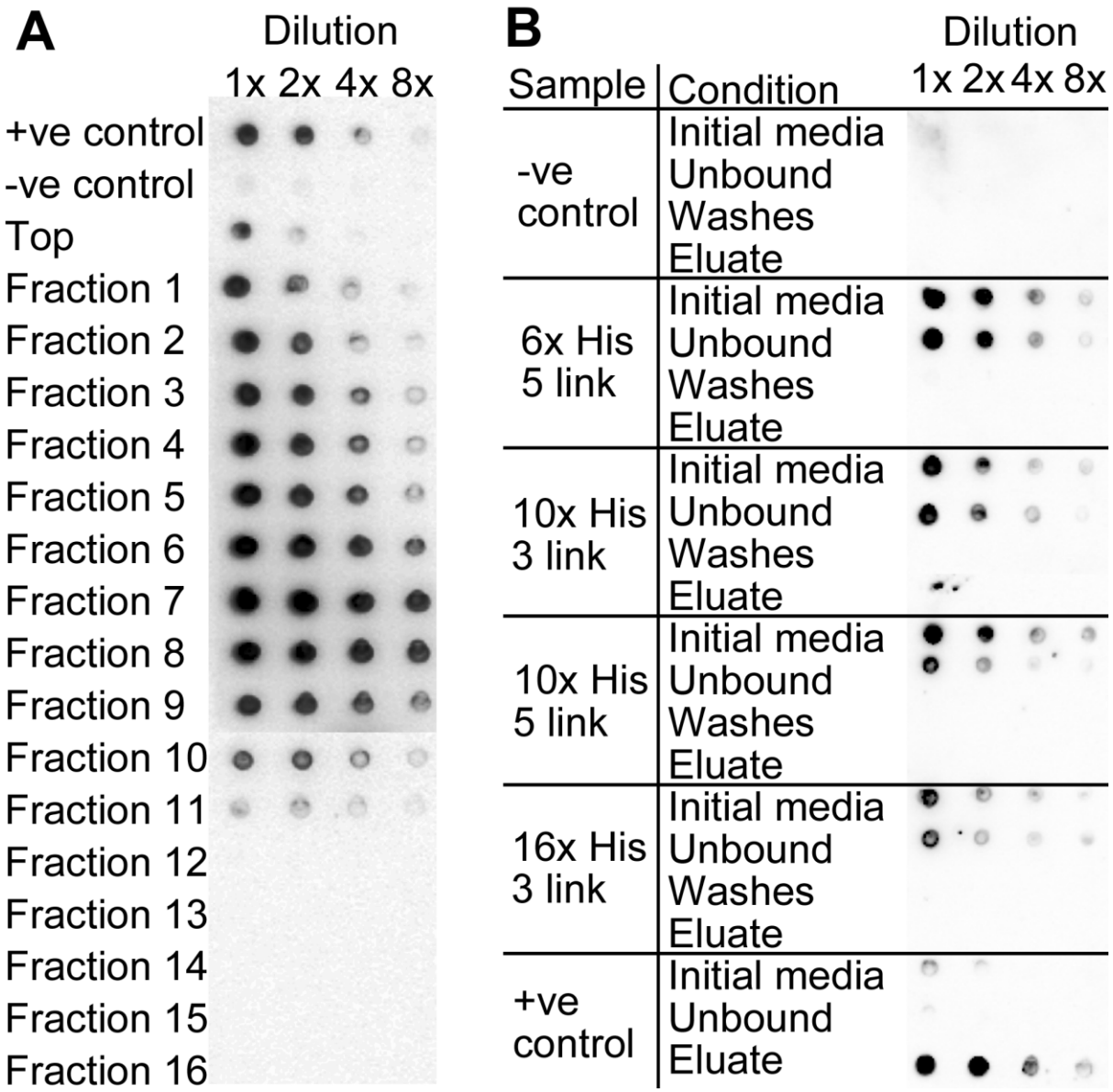


Figure 4.17 Testing purification strategies with HNTV VLPs
A) A dot blot demonstrating the diffuse distribution of HNTV particles across a 20-60% w/v sucrose density gradient in a SW32 Beckman centrifuge at 25,000rpm for 4 hours at 4°C. The positive control is a HNTV containing cell supernatant sample and the negative control is a HNTV free mock transfected cell supernatant. The gradient was harvested in approximately equal fractions from top to bottom. The dots are displayed as a 2x dilution series from left to right. B) A dot blot demonstrating the attempt to purify HNTV VLP by the use of an N-terminal poly-Histidine tag. The constructs used have different numbers of histidine residues and a different number of linker residues between the tag and the HNTV sequence and are detailed in supplementary Table 8.1 in the appendix. In order, the samples used are; Mock transfected cell supernatant, Construct B, Construct C, Construct D, Construct E, purified His-tagged and rabbit-derived G2 antibody. Samples were obtained by small scale transfection in HEK cells. The transfected cell supernatant (2.5ml) was first dialysed into standard buffer [10mM Tris pH 8.0 150mM NaCl] and the volume adjusted to 3ml (of which 1.5ml was used in the experiment). The positive control used is 25µg of purified His-tagged rabbit-derived antibody diluted in 1.5ml of standard buffer which was chosen as it will both bind to nickel beads and be detected by the anti-rabbit secondary antibody. 40µl of Magnetic Nickel beads (His Mag Sepharose Ni, GE Life Sciences) were incubated with the sample with gentle rocking for 1 hour at 4°C, followed by 4x washes with 1.5ml of standard buffer + 25mM imidazole. Elution was carried out using 100µl of standard buffer + 250mM imidazole. The dots are displayed as a 2x dilution series from left to right with the exception of the washes which are shown as 1 undiluted dot per wash.

In the absence of a successful affinity based method this left only sucrose density based methods available which suffer from co-purified vesicles and proteins from the cells and FBS. The method eventually settled on involved using 6 multi-layer TC flasks (equivalent to 30x T175 flasks). The stationary flasks, as opposed to roller bottles, allowed better control over the cell confluency at time of transfection and seemed to reduce cell death. The resulting 750ml of cell supernatant could then be concentrated down to approximately 30ml using a 100kDa cut-off “Vivaflow” concentrator. The resulting media was hyper-concentrated with FBS proteins that prevented further concentration and would contaminate the resulting sample. The majority of these proteins could be removed by dialysing the media in 1MDa cut-off dialysis membrane which retained only the large protein aggregates and vesicles including the hantaviral VLP. The resulting sample could be concentrated further in a centrifugal 100kDa cut-off filter to a volume of 3ml that could be loaded onto a sucrose density gradient. After this a faint band could be observed on the gradient that corresponds to the VLP. This band was then harvested and further pelleted through a sucrose cushion to concentrate and further clean the sample. The resulting preps were free from protein contamination but still contained unavoidable vesicle contaminants that contain small proteins that could be seen on an SDS-PAGE gel (Figure 4.18 A). However, these vesicles did not impact the data collection. This method worked well enough to produce VLP samples for efficient data acquisition.

Cryo-EM analysis revealed that the HNTV VLPs did display the classic hantaviral spikes and were not tubular. Once this expression and purification protocol had been optimised, the constructs used were switched to a form without the internal fluorescent tag. This was to ensure the VLP was as native as possible and to avoid a large internal density being visible in the resulting maps. As well as this, a new PUUV construct was obtained (kindly gifted by Dr Hepojoki and Dr Doores). This new construct was derived from a different Puumala sequence and appeared to produce a higher proportion of round VLPs. However the tubular forms of the VLP were still present suggesting that this tubular form of

the virus is a Puumala specific phenomena. The biological importance of this is unclear and may simply be an artefact from VLP production.

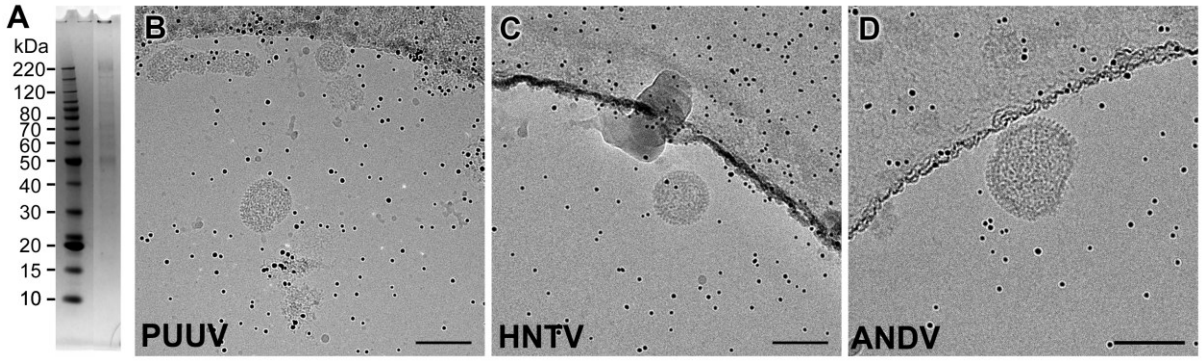


Figure 4.18 Successful production of hantaviral VLP for investigation by cryo-EM
A) A representative SDS-PAGE of a purified VLP sample. This is obtained from 5µl of the HNTV VLP sample that was used for data collection displayed in Figure 4.19. B-C) Representative cryo-EM projection images of PUUV, HNTV and ANDV VLP respectively. Gold beads are present for the alignment of tilt series. Scale bars = 100nm.

4.8. Hantaan VLP ultrastructure

As validation for the VLP production, it was important to obtain a cryo-EM reconstruction to check that the architecture of the lattice present on the VLPs is broadly equivalent to that found on the surface of live Tula virus. Figure 4.19 shows this reconstruction of HNTV virus. Fitted in the map are the available HNTV Gn and Gc crystal structures. These fittings were achieved by connecting the Gn to the Gc, based on the new hantaviral heterodimer model, and then fitting to the map. This shows that the broad architecture of the HNTV VLPs, at least at the current resolution, is indistinguishable from the TULV map.

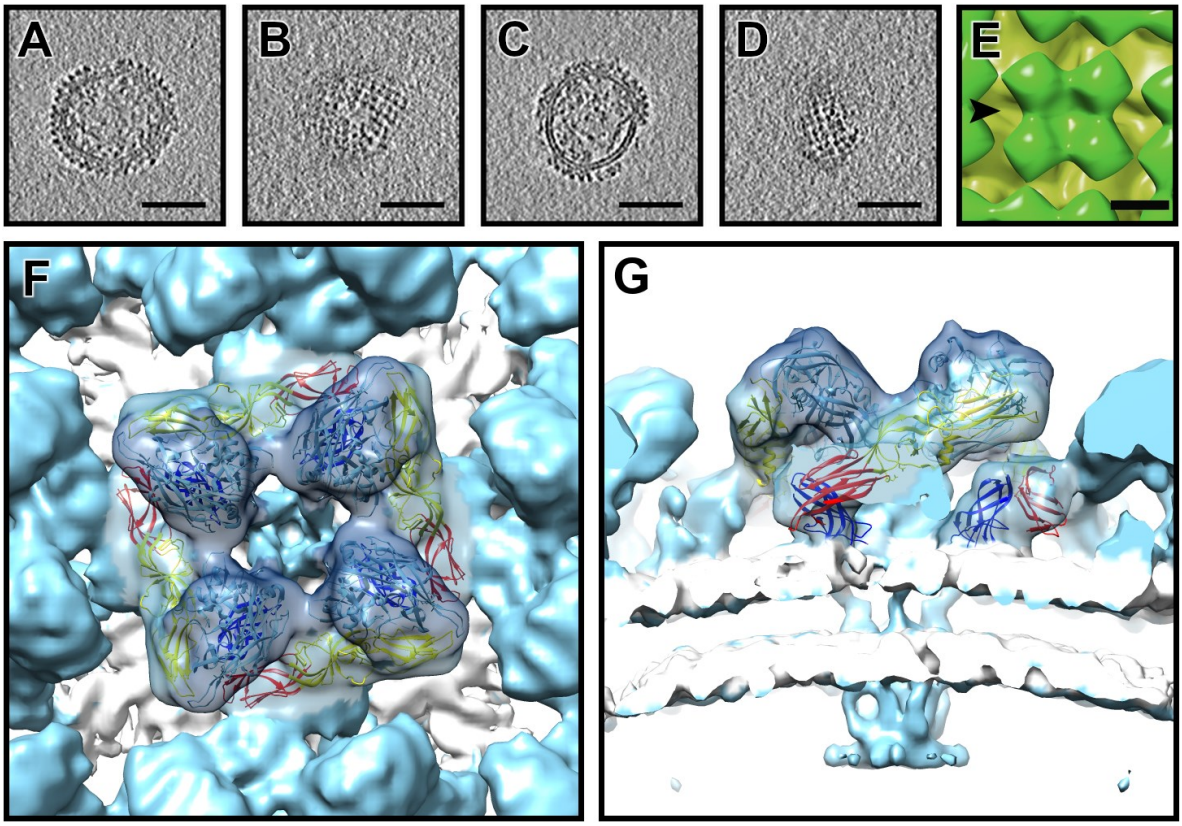


Figure 4.19 Pseudo-atomic model of HNTV obtained from a VLP system
A-D) Tomographic slices of two VLPs. Scale bar = 50nm. A&C) central slices. B&D) top views. E) A reconstruction of Hantaan virus previously published. The location of absent "peripheral stalk" density is indicated. [124]. F-G) A reconstruction of HNTV VLPs filtered to 12.6 Å. Fitted in the map are a model of the HNTV Gn (PDB-5OPG) and HNTV Gc (PDB-5LJY). These models were first aligned to each other based on the ANDV Gn-Gc heterodimer model supplied by Dr Guardado-Calvo, Institut Pasteur. The two models were then fitted together with C4 symmetry applied using the fitmap function in chimera F) A topview. G) A side cutaway view.

4.9. Discussion and future directions

Previous structural studies of Hantaan virus include crystal structures of the Gn and Gc ectodomains, which broadly appear similar to other Hantavirus species and a single electron microscopy study of live Hantaan virus displayed in Figure 4.19 E [16], [124]. This study utilised an unconventional single particle approach that included only side views from the perimeter of the virus. Single particle approaches are generally challenging for pleomorphic viruses such as Hantaan virus due to the presence of overlapping density from neighbouring spikes, but this study was able to resolve the hantaan virus lattice to 25 Å resolution. In this study, it was suggested that there were a number of differences between Hantaan virus and the Tula virus reconstruction available at the time [50]. One of these differences include the

absence of “peripheral stalks” that form the interface between the tetrameric spikes (Figure 4.19 E). Based on the current Tula virus fitting, it is now clear that these stalks are predominantly Gc density so it is unexpected to see this density absent from the reconstruction. The Hantaan VLP map presented here appears to be broadly similar to Tula virus. There is clear density present that corresponds to the “peripheral stalks” of Gc density. The differences in the two maps presented here likely stem from the difference in the techniques used to collect and process the data. It highlights the importance of obtaining three dimensional tomographic data when resolving overlapping protein density by cryo-EM.

The HNTV VLP reconstruction demonstrates that hantaviral VLPs are a suitable system for structural studies of pathogenic hantaviruses which avoid the complications associated with handling dangerous micro-organisms. However the VLP preparation method is far from ideal and could benefit from further optimisation to produce more highly concentrated and pure samples. With the current samples, the data acquisition setup is slow due to the time taken to search for ideal targets and on average only a single VLP is imaged per tomogram. Towards the end of this project, a new publication was released which included details of another hantaviral VLP system [136]. Here a streptavidin tag was fused to the VLP N-terminus which could successfully be used to purify the VLP. It is currently unclear why this approach was successful but the N-terminal His tag purification attempts presented here failed. This discovery came too late into the project to be attempted but certainly seems to be a promising avenue for future investigation. One downside to this approach is that the presence of the tag could conceivably interfere with receptor and antibody binding experiments but this could potentially be reduced by engineering a cleavable tag.

The data presented here in combination with the available x-ray crystallography data provide a good description of the native hantaviral lattice. A target now for the future would be to obtain a clearer picture of how this lattice adapts to changing environments and throughout the endocytic pathway. The investigation into Tula virus after exposure to low pH

represents the first step towards this understanding. It is unfortunate that the loss of the lattice and the high degree structural heterogeneity renders subtomogram averaging impossible. It is notable that a fellow member of the *Bunyavirales* order, Bunyamwera virus, has also been reported to undergo a similarly disordered structural rearrangement under low pH conditions [123]. In the future it will be interesting to investigate intermediate pH values between neutral and pH 5 to see if there is a more ordered transitional state between these two extremes. Another future experiment, would be to mix virus/VLP with liposomes to investigate whether the added hydrophobic surface provides a region of local order. This may make it possible to resolve fusion intermediates in which the fusion loops are seen embedded into the target membrane as have been reported recently for Rift Valley fever virus [8]. Beyond this, it may be possible to visualise post fusion forms of the viral glycoproteins.

Multiple crystal structures have now been obtained of class II fusion protein ectodomains in both pre and post fusion form [16], [125], [127]. These structures show that the class II fusion protein fold is highly conserved among distantly related virus families, highlighting the importance of this fold for the fusion process. However, despite this similarity, the way in which these basic units are arranged on the surface of the virus is highly varied. It is therefore very difficult to predict the ultrastructure of the virus from ectodomains alone, highlighting the importance of cryo-EM and cryo-ET techniques to obtain a complete description of the whole virion. Figure 4.20 displays a schematic representation of the ultrastructure of a selection of class II fusion protein displaying viruses including the new fitting confirmed here for members of *Hantaviridae*. Each of these viruses contain a class II fusion protein and an additional accessory protein, named Gn for Bunyaviruses, E2 for alphaviruses and M for flaviviruses (shown in transparent grey) [8], [137], [138]. The flavivirus M protein is small and shielded completely by the Class II E protein so is not visible in this depiction. One possible function for these accessory proteins appears to be to protect the hydrophobic fusion loops of the fusion protein. One obvious exception to this is the

flavivirus E protein which protects its own fusion loops by forming a head-to-tail homodimer. It is also notable that the protection offered by the Gn in members of *Hantaviridae* comes from just a single loop and therefore may not be complete. The accessory "Gn" glycoproteins appear to act as an "architectural mortar" that hold the class II fusion proteins in a consistent orientation that define the varied morphologies displayed by these viruses.

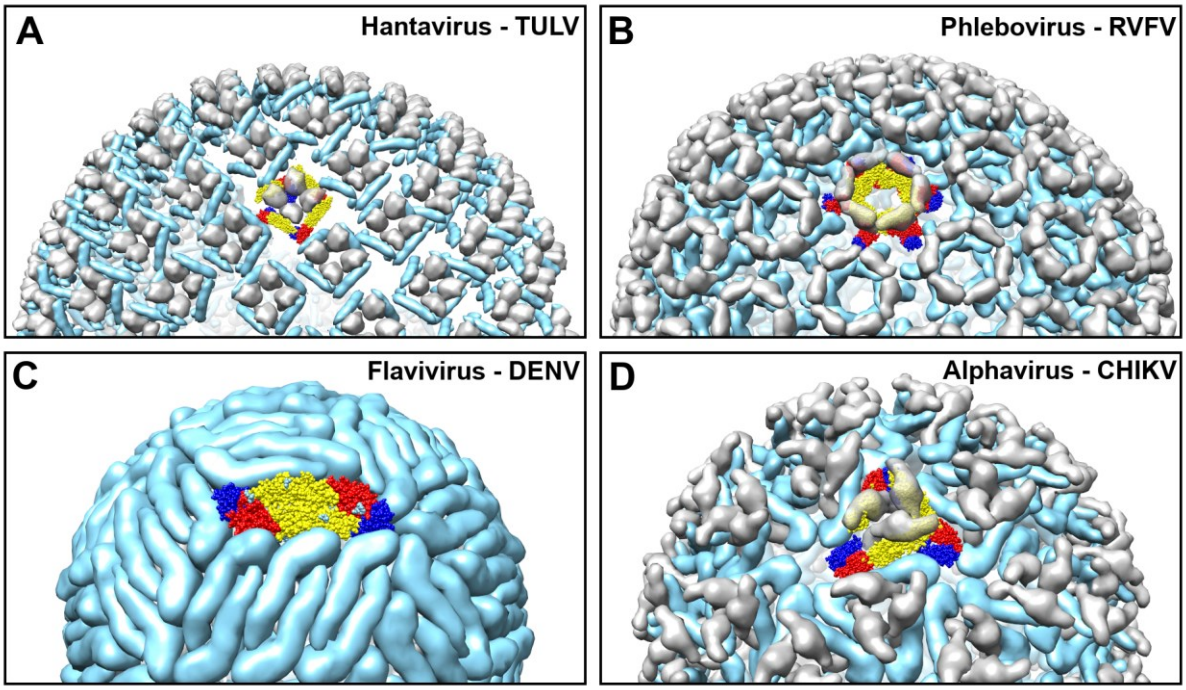


Figure 4.20 Diverse morphology of class II fusion displaying viruses
Schematic depictions of the arrangement of a range of viruses displaying class II fusion proteins. The Class II fusion proteins are displayed in blue except for the central examples which are depicted as surface models coloured in the standard red, yellow, blue colours for domains I, II and III respectively. Other accessory proteins such as the hantaviral Gn are displayed in grey. A) The pseudo-atomic hantaviral model presented here is plotted back into a whole virion using the refined coordinates from a specific virus in the dataset. Note that the stalk region is absent from this model. B) A depiction of Rift Valley fever virus made using PDB-6F9B. C) A depiction of Dengue virus made using PDB-3J27. Note that the grey accessory protein here is small and obscured entirely by the class II fusion protein. D) A depiction of Chikungunya virus made using PDB-2XFB.

5. Hantavirus interactions with cell surface receptors

5.1. Introduction

In order to gain access to the cell, viruses typically bind to a cell surface receptor protein that allows the virus to be taken into the cell via a number of endocytic pathways. Hantaviruses have been observed to utilise multiple endocytic pathways for cellular uptake including clathrin dependent endocytosis [37]. In addition to this, there are a wealth of studies reporting a number of different receptor proteins for various hantavirus species (summarised in Table 5.1). The first receptor proteins to be identified were αVβ3 integrins which have been implicated in cell entry for a range of pathogenic hantaviruses including HNTV, ANDV, SNV and PUUV and many others [133], [139]–[141]. Interestingly, two hantaviruses that are not pathogenic to humans, Prospect Hill virus and Tula virus, were shown to utilise a different integrin, α5β1, which suggests that αVβ3 integrin usage may be required for human disease [133], [134]. These studies first demonstrated the importance of integrins by showing that anti-integrin antibodies and integrin binding ligands such as vibronectin inhibit hantavirus infection of Vero E6 cells [134]. Since then, further investigations have demonstrated that hantavirus infection can be blocked by pre mixing with recombinant αvβ3 integrin [133]. This study then went further to show that a short polypeptide, belonging to the PSI domain of the integrin can alone inhibit infection and that hantavirus infection specifically requires the bent conformation of the integrin [133], [141].

In addition to integrins, the GPI-anchored decay accelerating factor (DAF) has been reported as an essential receptor for HNTV and PUUV to enter the apical surface of epithelial cells [142]. This study demonstrated that antibodies against DAF and premixing with soluble DAF was sufficient to block hantavirus infection. Another candidate hantaviral receptor is gC1qR (globular heads of complement C1q receptor, also known as C1QBP, HABP1, SF2-associated p32, p32/TAP, gC1qR-p32, gC1qR/p32 and just p32,) [140]. This was discovered by running Vero E6 cell homogenates on a gel, blotting to nitrocellulose and then probing for binding to HNTV virions before identification by mass-spectrometry. This led

to discovery of gC1qR which was shown to be important for HNTV infection by an RNAi knockout experiment. The study also demonstrated that non-permissive CHO cells could be made susceptible to HNTV infection by transfection of gC1qR [140].

More recently, another potential hantaviral receptor, Protocadherin-1 (PCDH1), has been implicated in the infection of New World hantaviruses specifically [136]. PCDH1 is a cell surface protein that contains a transmembrane domain followed by 7 extracellular cadherin repeat domains named EC1-7. PCDH1 was initially identified in a loss of function genetic screen in human cells. Following this, a number of experiments were performed which confirm its role in hantavirus infection. These include studies with VSV-pseudovirus systems that confirm that knocking out either the whole PCDH1 gene, and even just the final EC1 domain, reduced infection of the new world ANDV and SNV but not HNTV. As well as this pre-mixing the cells with anti-PCDH1 antibodies and premixing the pseudoviruses with recombinantly expressed soluble EC1-2 were both able to prevent infection. This study also confirmed direct binding of hantaviral VLPs to EC1-2 in both ELISA and by biolayer interferometry. This wealth of evidence suggests a clear role for PCDH1 in hantaviral infection but the authors note that the rate of infection in these studies did not reach zero suggesting the existence of PCDH1 independent pathways.

Receptor candidate	Hantavirus species	References
αVβ3 integrin	HNTV, ANDV, SNV and PUUV	[133], [134], [139], [141]
α5β1 integrin	TULV and Prospect Hill virus	[133], [134]
DAF	HNTV and PUUV	[142]
gC1qR	HNTV	[140]
Protocadherin-1	ANDV, SNV, Maporal virus (not HNTV)	[136]

Table 5.1 Hantavirus receptor candidates

Despite the large number of studies on candidate hantaviral receptors, there are currently no detailed molecular descriptions of how any of these proteins interact with hantaviral cell surface glycoproteins, either by x-ray crystallography or cryoEM.

5.2. Andes virus and Protocadherin-1

A recent discovery confirmed Protocadherin-1 (PCDH-1) as candidate receptor for new world Hantaviruses [136]. This study shows that a soluble fragment, comprised of the two membrane distal domains named ec1 and ec2, are able to neutralise Andes virus infection and includes evidence of high affinity direct binding to Maporal virus VLPs (another new world hantavirus). I recombinantly expressed a His-tagged PCDH-1 EC1-EC2 and purified it by IMAC and SEC. Representative SEC traces are shown in Figure 5.1 A and an SDS PAGE gel in Figure 5.1 B confirming the production of the protein.

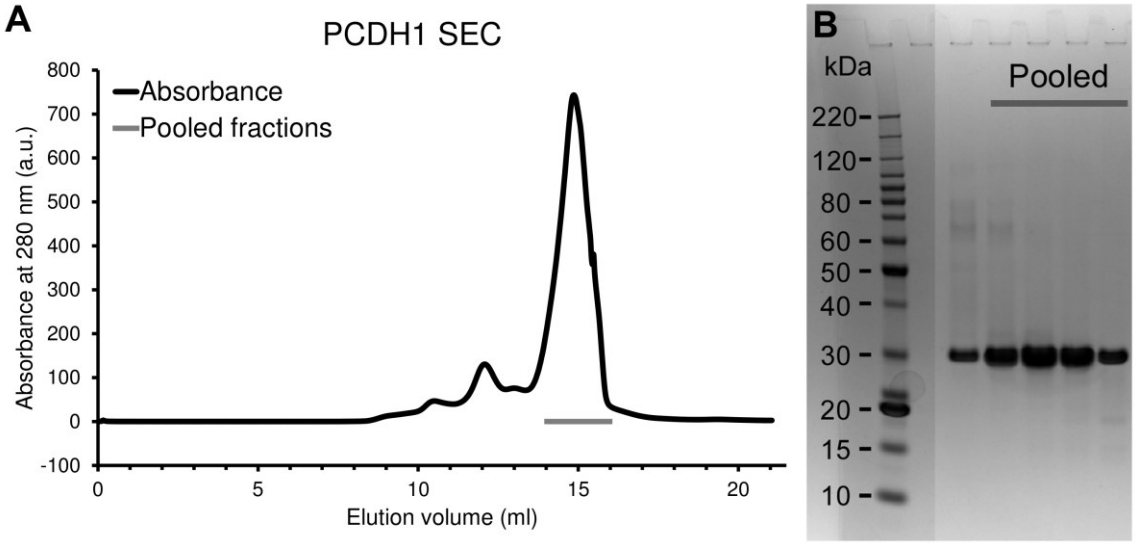


Figure 5.1 Purification of recombinantly expressed PCDH1 EC1-EC2
A) A SEC trace of the PCDH1 protein. B) An SDS-PAGE of the 0.5ml fractions from the SEC run with pooled fractions displayed.Expected MW = 26kDa.

Next, I wanted to confirm binding to ANDV VLPs by ELISA. This involved immobilising purified ANDV VLP on the ELISA plate and binding a dilution series of PCDH-1 from 0.2 to 10,000 nM. Unfortunately, no signal was detected above background (See supplementary figure Figure 8.2). In the absence of a good positive control (for example an ANDV binding antibody) it is impossible to rule out the possibility that too little ANDV VLP was immobilised on the plate. Due to the impurity of the VLP preps (mostly co-purified vesicles and large protein complexes) it is difficult to get a reliable estimate for the concentration of VLP in the sample. However, additional wells were included in the ELISA with 5 x as much VLP and these also did not produce binding signal. Furthermore, a PUUV VLP sample, prepared similarly, was immobilised on the same ELISA plate and was able to

detect binding to the fab fragment 4G2 (See Figure 6.1). A positive control was included in which PCDH1 alone was immobilised on the plate (0.001 µg/ml) which did confirm that the detection of PCDH1 was successful (of 0.45 and 0.38). An alternative explanation for the lack of signal may be that the His tag on the PCDH-1 is inaccessible when bound to ANDV VLP.

In order to confirm that the PCDH-1 protein produced in our lab is functional, a sample was sent to Dr Doores (KCL) who has a VSV-based hantaviral pseudovirus neutralisation assay similar to those reported in [136]. This confirmed that the protein I produced was able to neutralise ANDV pseudovirus infection as shown in Figure 5.2, giving some confidence that this protein is functional.

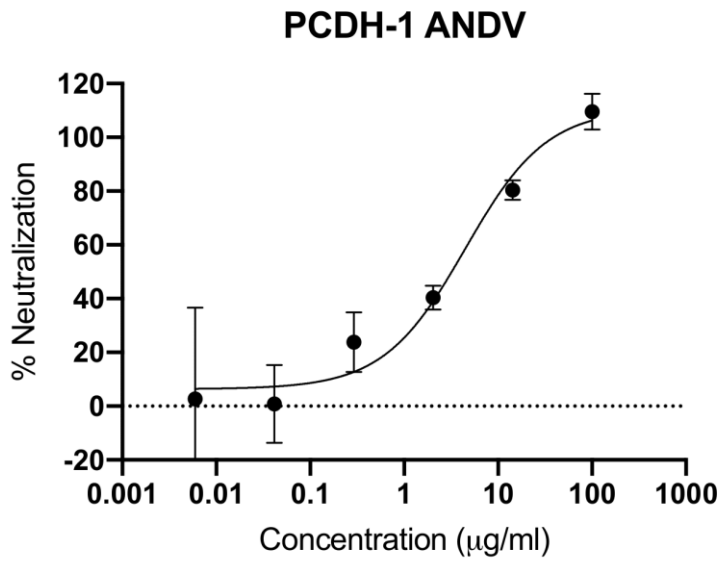


Figure 5.2 ANDV pseudovirus neutralisation by PCDH1 EC1-EC2
An ANDV VSV-pseudovirus system is used to demonstrate the neutralising effect of the PCDH-1 EC1-EC2 protein I produced. Note: This experiment was performed by Dr Katie Doores, King's College London who also supplied me with this figure.

Cryo-EM grids were then prepared with ANDV VLP and PCDH-1 mixed at a final concentration of 2.6µM and tomographic tilt series were collected. The tilt series were aligned and the data processed using the standard hantaviral processing workflow described previously with C4 symmetry imposed. The resulting reconstruction appeared similar to other hantaviral reconstructions displaying the classic tetrameric lattice of spikes. However, in this

reconstruction there was an extra piece of density observed in the centre of the spike just above the transition from the globular Gn ectodomain to the stalk region (Figure 5.3 B). For comparison a very small data set of ANDV VLP alone was collected and processed in the same way. While much lower resolution, this reconstruction clearly does not include any density this high up in the centre of the spike (Figure 5.3 B). There appeared to be minimal contact between this density and the inside surface of the spike which is likely an artefact as a result of a symmetry mismatch between the bound PCDH-1 and the C4 symmetric spike.

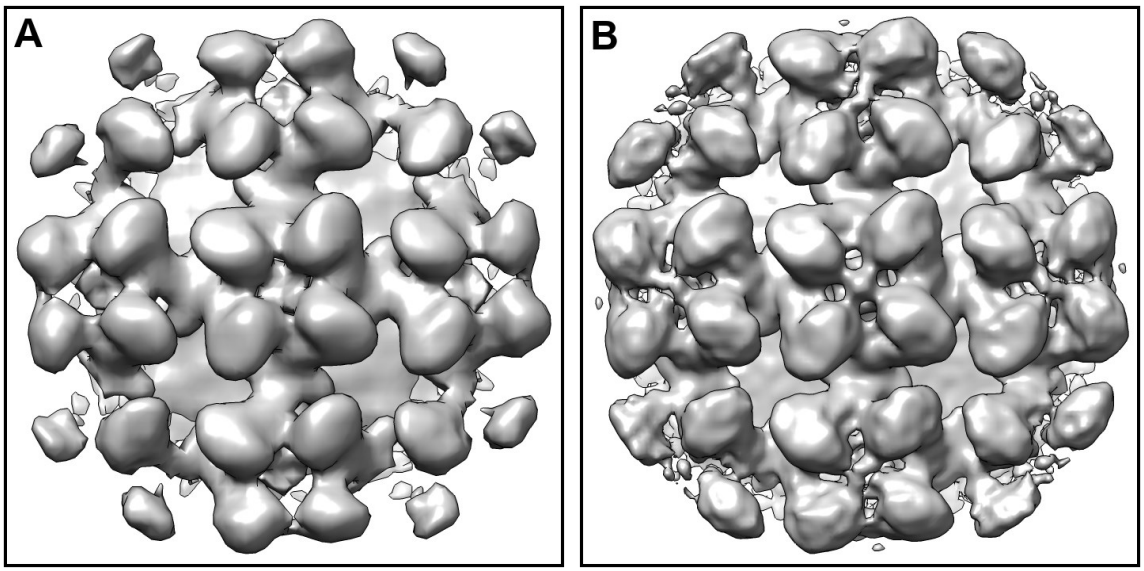


Figure 5.3 ANDV VLPs refined with C4 symmetry with and without PCDH1 EC1-EC2
A) ANDV VLP reconstruction refined with C4 symmetry imposed (339 particles at 19.84 Å). B) ANDV VLP + PCDH1 refined with C4 symmetry imposed (2823 particles at 14.31 Å)

To probe this further, 3D classification was performed in Relion using the refined spike coordinates, with C1 symmetry and without allowing image alignment. A focussed mask was used tightly around the region of interest to avoid classifying according to broader flexibility in the spike complex. Separate classes could be observed with density bridging across the lobes of the spike in two directions which could be rotated and subsequently classified further. After several rounds, this enabled the symmetry mismatch to be resolved resulting in the map displayed in Figure 5.4 in which the density can be seen projecting from a single spike.

The resulting additional density is small, and without the added benefit of C4 symmetry, the resolution is low, so a certain degree of caution is wise when interpreting this map. The density is approximately sized to accommodate a single EC domain of PCDH1. An X-ray crystal structure of EC domains 1-4 of PCDH1 has been resolved previously (PDB ID: 6BX7) so this was used to display the EC1 domain of PCDH1 in the map. It must be stressed that at such low resolution the orientation of this model cannot be unambiguously determined and is shown here for size comparison only. The EC2 domain is likely flexibly linked to the EC1 domain which may explain its apparent absence from the map. A certain degree of flexibility could be seen during the classification and the possibility that a small population of spikes containing additional PCDH1 EC1-EC2 molecules at a number of the remaining 3 sites in the tetramer cannot be excluded. However, it does appear that on average, a single EC1-2 domain binds per tetrameric spike. This may simply be partial occupancy resulting from a low affinity interaction but it is also reasonable to suggest that the EC1-EC2 protein extends far enough into the centre of the spike to sterically hinder binding to the remaining sites.

The location of the ANDV Gn can be fitted into the map using the ANDV Gn-Gc heterodimer crystal structure described previously (unpublished, Dr Guardado-Calvo, Institut Pasteur). This enables a broad patch on the surface of the Gn to be tentatively identified which may include residues responsible for the interaction with PDCH1. It should be noted that the upper stalk region of the Gn, which is excluded from the crystallised Gn models, may also be involved in the interaction. PCDH1 is reported to bind specifically to New World hantaviruses only, so an illustration of conserved residues of New World hantaviruses is displayed in Figure 5.4 D on the ANDV Gn surface which broadly displays an area of higher sequence conservation in the lower half of the patch relative to the upper half. However an obvious group of residues that are conserved among New World hantaviruses and not among the Old World hantaviruses cannot be seen. In addition to this, pairwise sequence alignments are displayed between the New World SNV, and Old World PUUV and HNTV

(Figure 5.4 E-G). The coulombic surface of EC1 is displayed in Figure 5.4 H and M which reveals a reasonably acidic surface and the coulombic surface of the Gn from various species are displayed in Figure 5.4 I-L. The putative contact patch with PCDH1 does appear to contain some positively charged residues that would be expected to enable binding to the acidic EC1.

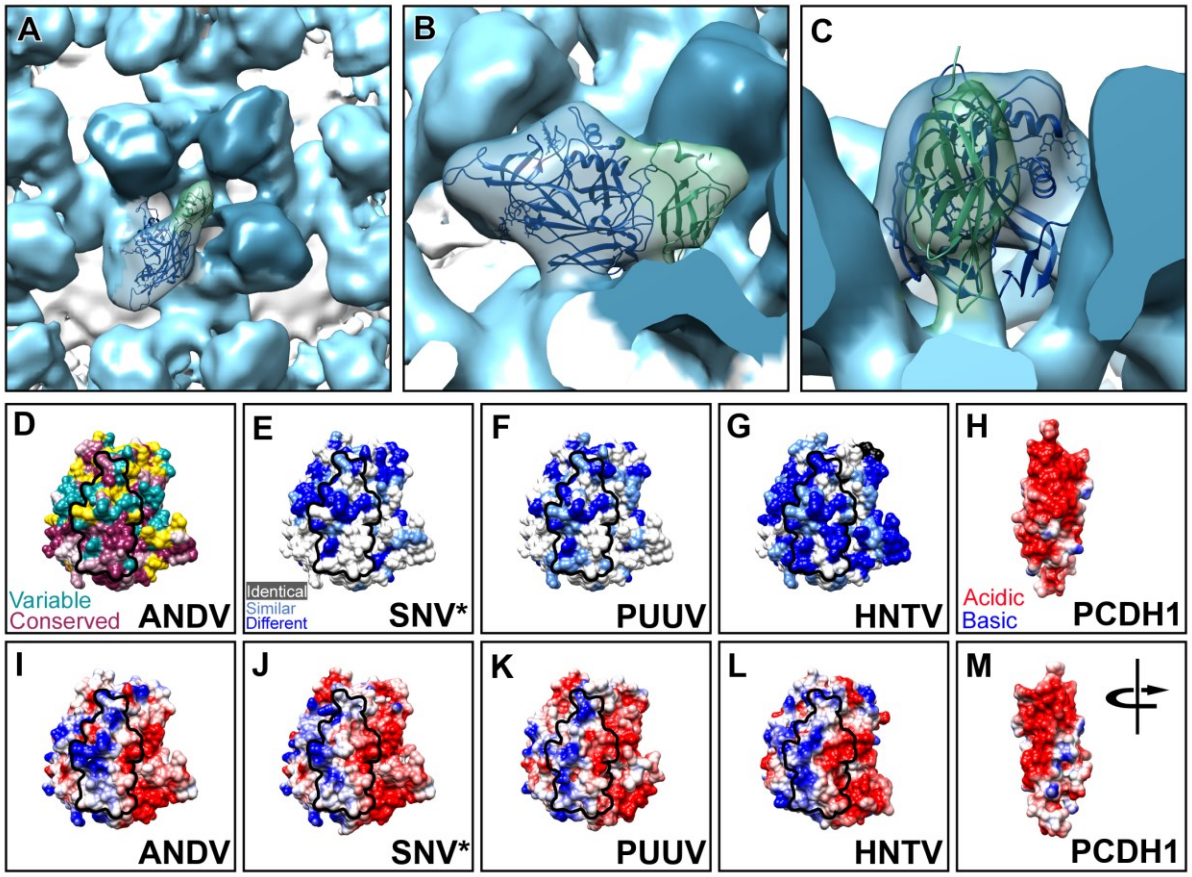


Figure 5.4 Putative ANDV VLP interactions with PCDH1 EC1-EC2
A-C) Different views of the ANDV VLP + PCDH1 reconstruction filtered to a resolution of 16.5 Å. Fitted in the map is the ANDV Gn-Gc heterodimer structure supplied by Dr Guardado-Calvo and Prof Rey, unpublished (using the fitmap function in chimera). Gc not shown. PCDH1 EC1 from PDB-6BX7 is displayed manually fitted in the map. The orientation of this cannot be unambiguously determined and is displayed here for comparison only. D-M) Surface depictions of different hantaviral Gn models and PCDH1 EC1 which match the viewing angle displayed in C. (*) The SNV homology model is derived from the ANDV model using the swiss-model server [143] in the absence of a SNV Gn model. The HNTV model is PDB-5OPG and the PUUV model is PDB-5FXU. A putative approximate contact patch is indicated as a black outline. D) A sequence conservation analysis of New world rodent-borne hantaviruses displayed on the ANDV Gn obtained using the consurf server [144]. Green = variable. Purple = conserved. Yellow = insufficient data. A Multiple sequence alignment derived using Clustal Omega [145] was supplied comprised of 66 sequences of new world rodent-borne hantavirus species including 33 ANDV sequences and 6 SNV sequences. E-F) Pairwise sequence alignments of ANDV Gn versus SNV, PUUV and HNTV respectively. White = identical. Light blue = similar. Dark blue = Different (as defined by the BLOSUM62 matrix). H-M) Coulombic surface potential (-5 red – 5 blue) displayed on the surface of each model. H and M) PCDH1 EC1 from PDB-6BX7 displayed in the orientation in C and rotated 180 degrees from C respectively.

5.3. Puumala virus VLPs and αvβ3 integrin

A small sample of human αVβ3 integrin ectodomain, was recombinantly expressed as had been previously reported for αVβ6 integrin (a kind gift of Helen Duyvesteyn, Strubi) [146]. Unfortunately, the concentration of this sample was extremely limited which made functional binding experiments impossible. Despite this, cryo-EM grids were prepared with PUUV VLPs and the integrin ectodomain. Due to the constraints of the sample, this mixture contained a relatively low concentration of integrin at 0.64μM. Included in the buffer was 1.6mM $CaCl_2$ because a previous study suggests that integrins bind hantaviruses in the bent conformation which is induced by the presence of Ca^{2+} ions [133]. A small data set was collected of these VLPs on a 200kV microscope and the data processed as before (Figure 5.5). The map is low resolution but given that the integrin ectodomain is approximately 300kDa the additional density should be clear. These data demonstrates that under these conditions, the integrin cannot be seen and therefore has likely not bound. Further studies are needed to determine whether this is simply due to extreme low occupancy, as a result of the low concentration of integrin or whether it binds to PUUV VLPs at all.

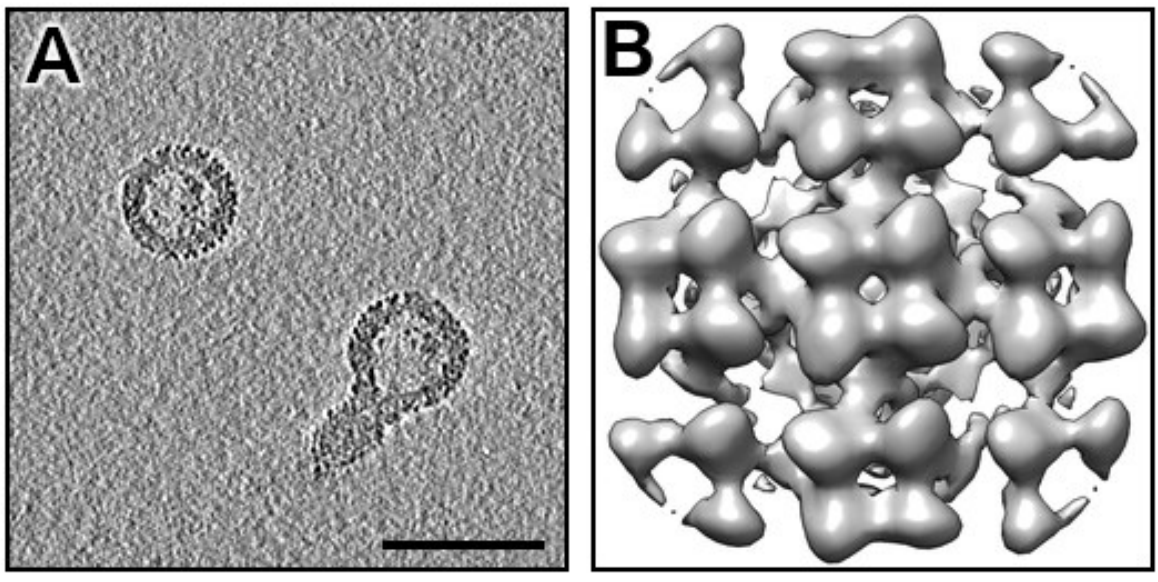


Figure 5.5 An attempt to observe an interaction with PUUV VLPs and αVβ3 integrin ectodomain
A) A representative tomogram slice of PUUV VLPs that have been mixed with a relatively low concentration of αVβ3 integrin ectodomain (kindly gifted by Helen Duyvesteyn, Strubi) B) A reconstruction of PUUV VLPs obtained from this sample filtered to 18.2 Å.

5.4. Puumala virus VLPs and gc1qr

A sample of the candidate receptor protein, gC1qR, (residues 73-282 as in the previously reported PDB ID: 1P32) was recombinantly expressed and purified (a kind gift of Dr Rissanen, Strubi). An ELISA was performed to monitor binding of this protein to PUUV VLPs (Figure 5.6) which suggests a low nM affinity interaction with the VLP.

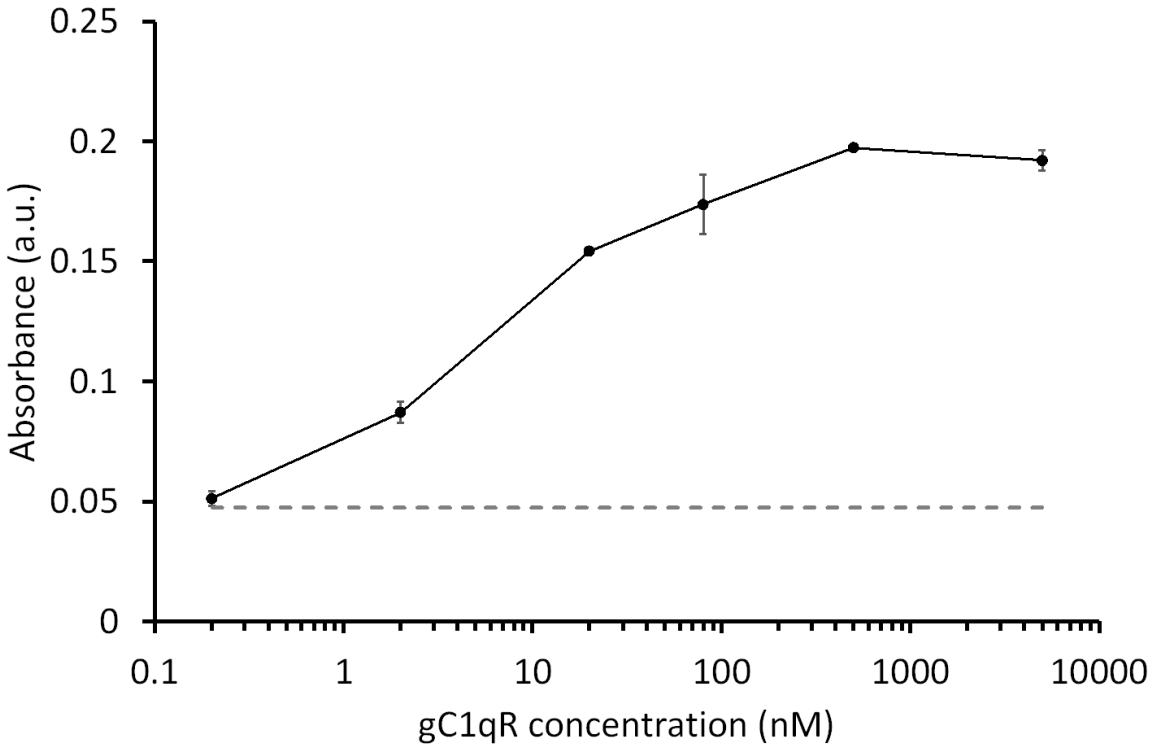


Figure 5.6 ELISA to demonstrate binding of gC1qR to PUUV VLPs
A sample purified PUUV VLPs were immobilised on a plate and probed with a dilution series of gC1qR. An analyte-free negative control well was included giving a background reading of 0.474 as indicated with a grey dashed line. The 280nm Absorbance of the original VLP sample was 3.98 a.u. which was diluted 250x before plating.

Based on this result, cryo-EM grids were prepared in which the protein was mixed with PUUV VLP at a concentration of 3.2μM. Tilt series were collected and a reconstruction obtained using the standard refinement protocol (Figure 5.7 C-F). This reconstruction appears very similar to other hantaviral reconstructions. One key difference is the presence of a small bit of density in the centre of the stalk region. For comparison, a PUUV VLP only reconstruction is displayed in Figure 5.7 A-B. In the absence of an apo-PUUV VLP reconstruction, this map was obtained from a sample mixed with fab 4G2 but does not

contain density for the 4G2 fab (see section 6.2 for details). Note that the additional density is much deeper into the stalk than was seen with ANDV and PCDH1. The density is clearly too small to be the full gC1qR protein (25.1kDa). This could be potentially due to a symmetry mismatch as was seen with PCDH1, but at the time of writing, any attempts to resolve this mismatch by classification have been unsuccessful due to the small size of the density and the crowded environment of the stalk region. Ultimately, the question of whether gC1qR can be seen in this reconstruction remains unresolved. It is generally wise to exercise extreme caution when interpreting small pieces of density, especially around symmetry axes, so further study is required to determine whether this an EM artefact or gC1qR.

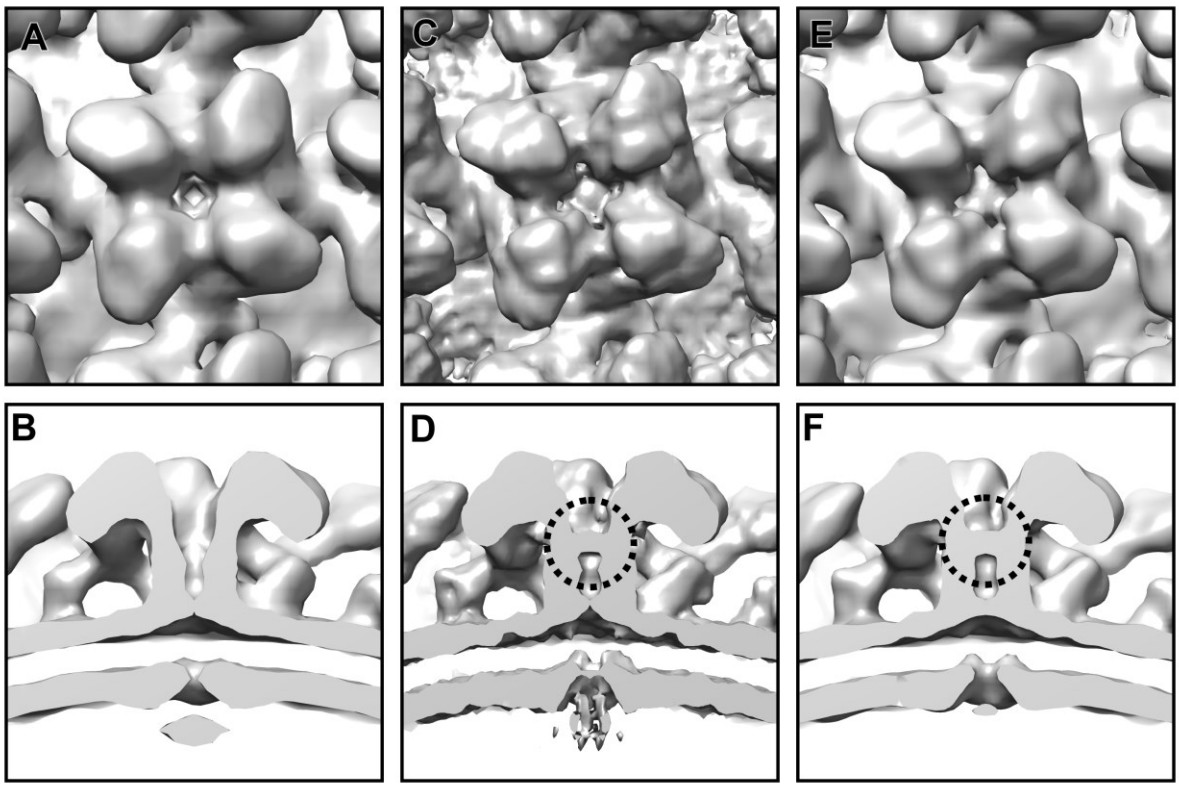


Figure 5.7 Additional density seen in PUUV VLPs in the presence of gC1qR
A-B) PUUV VLP reconstruction filtered to 16.8 Å. Note that in the absence of a PUUV VLP only dataset, this is from the regions of lattice in a dataset with fab fragment 4G2, but the fab is absent here. (See section 6.2 for details). C-D) PUUV VLP mixed with gC1qR filtered to 10.74 Å. E-F) The same reconstruction in C-D filtered to the same resolution as A-B for fair comparison.

5.5. Discussion and future directions

The investigations into receptor interactions with hantaviruses have been challenging. Despite a wealth of evidence supporting a role for αvβ3 integrins in hantaviral disease, I was

unable to demonstrate binding in the limited experiment I was able to carry out. When more integrin sample becomes available, an obvious next step would be to carry out a binding experiment such as an ELISA. This should be done in the presence and absence of Ca^{2+} ions and Mn^{2+} ions which each induce the formation of the bent and extended conformation of integrins, respectively [133], [146]. An approximate assessment of the affinity of the interaction would also provide an insight into whether the concentration of integrin used in this experiment was high enough. It is also possible that the integrin does not bind directly to native hantaviral particles and may only bind to an intermediate conformation that exists further down the endocytic pathway, but the data presented here is certainly not enough to rule out binding to the native lattice.

The investigation into gC1qR and PUUV VLP requires further study to determine whether binding has occurred. The small bit of density in the map may simply be an artefact from the refinement although it is notable that a similar density is not seen in either the TULV or HNTV VLP reconstruction, both of which were resolved to a similar resolution. Care was taken to avoid common problems with subtomogram averaging data (such as the missing wedge, preferential orientation and overfitting) but these are difficult to rule out completely without comparison to atomic models.

The ELISA data appears to support the conclusion that gC1qR does bind to PUUV VLPs. There is however an important caveat to make with the ELISAs performed with these VLP samples which comes from the fact that these VLP samples are relatively impure and therefore a meaningful estimation of the concentration of the VLP cannot be accurately determined. This makes it difficult to properly control the amount of VLP immobilised on the ELISA plate and therefore a perfect negative control sample cannot be included. This leaves open the possibility that the analyte protein could bind to something else in the VLP prep that is not the VLP itself. In parallel to the PUUV:gC1qR ELISA a sample of ANDV VLP was also tested as a negative control which was prepared in the same way and diluted by the same amount (See supplementary Figure 8.3). However, the amount of VLP protein and

contaminating protein between these two preps is likely different. The PUUV VLP sample had an A280 absorbance reading of 3.98 (a.u.) versus just 0.57 (a.u.) for the ANDV VLP sample, as measured with a nanodrop spectrophotometer. However given that the samples may have different relative amounts of VLP to contaminating proteins, which cannot be easily quantified, the A280 absorbance reading does not give a good indication of the relative levels of VLP in each sample. As a result of this, it is not meaningful to compare the absolute ELISA signal between the two preps. One thing that can be compared is the general shape of the curve which appears to demonstrate a higher affinity interaction for PUUV compared to ANDV, although the ANDV signal was low and fairly close to the background reading. These problems indicate that the use of ELISA for monitoring receptor interactions with these VLP samples is limited.

As previously mentioned in section 4.9, an alternative method for preparing VLPs may be possible with the use of an N-terminal streptavidin tag [136]. This would allow for the preparation of a purer sample of VLP which could enable better control of the amount of VLP used in functional experiments. The surface accessible strep tag would also provide a consistent way detect the relative level of VLP across different species which would make it possible to perform an ELISA with immobilised receptor protein and VLP as the analyte.

The limitations of the use of ELISA discussed here may also explain the failure of the ELISA with ANDV and PCDH1 although this requires further investigation. The neutralisation data provided in collaboration with Katie Doores does provide good evidence that the PCDH1 EC1-EC2 I produced is able to neutralise ANDV glycoprotein decorated VSV-pseudoviruses. Note that the ANDV M segment construct used for that experiment is identical to that used to generate VLP here. Moreover, the buffer conditions used in the preparation of the cryo-EM sample (PBS) are identical to those used in the binding experiments performed in [136].

The additional density observed in the ANDV:PCDH1 EC1-EC2 cryo-EM reconstruction is reasonably convincing but further validation is required given its small size

and the data processing complications associated with mismatched symmetry and flexibility. One potential solution to this would be to repeat the cryo-EM reconstruction with PCDH1 EC1 domain only. Some evidence is provided in [136] that the EC1 domain alone is capable of neutralising ANDV which was unfortunately not available at the time the PCDH1 EC1-EC2 construct was designed. This would hopefully avoid the problems with a flexibly linked EC2 domain and the smaller size may even allow four EC1 proteins to bind to the spike to eliminate the symmetry mismatch problem. Alternatively a much larger reporter protein or gold tag could be fused to the PCDH1 EC1-EC2 which would make it obvious to observe binding to the VLP from visual inspection of a tomogram.

Beyond this, it would then be important to validate the putative binding site on the surface of the VLP with a mutagenesis experiment. For example, a selection of residues could be identified at the site which, when mutated, prevent binding of PCDH1 or PCDH1 mediated neutralisation of an ANDV pseudovirus. Panels D-M in Figure 5.4 represent an attempt made to identify likely residues at this interface. The sequence conservation among New World rodent-borne hantaviruses (Figure 5.4 D) does display some residues that appear to be reasonably well conserved which would be expected at a receptor binding site. However it is difficult to discern any residues which are present in the New World hantaviruses but absent from both the Old World hantaviruses. One notable difference between the available models of New World and Old World hantaviruses in the region of interest is the loop between residues 116 and 127. This is α-helical in both the HNTV and PUUV models but not in the ANDV model. It may therefore be interesting to graft this loop from Old World to New World viruses (and vice-versa) to investigate its impact on PCDH1 binding.

Hantaviral interactions with cell surface receptors remain a key area for further investigation. While many candidate receptor proteins have been reported there appears to be little consensus on how and when each receptor is used during the viral life cycle and in what tissues of the body. A differential ability to bind human and animal receptors may also

provide an explanation as to why only particular species of hantaviruses undergo zoonotic transmission into humans and why this manifests in disease with varying severity. It is hoped that future study, including insights from structural biology, may help to uncover these unknowns to help inform novel strategies to prevent hantaviral disease.

6. The host immune response to hantaviruses

6.1. Introduction

There are limited options available for the prevention and treatment of hantavirus caused diseases HPS and HFRS [4]. Immunotherapy is one possible treatment option that has proved successful in the treatment of other haemorrhagic fever causing viruses [147]. However there are currently no FDA approved immunotherapeutic treatments available for any hantavirus disease [4].

Despite the lack of approved immunotherapeutic treatments for hantaviruses, the neutralisation of hantaviruses by a range of antibodies has been well studied. In one early example, a range of antibodies were isolated from hybridoma cell lines obtained from the PUUV host species, the Bank Vole [33]. The antibodies were tested to monitor both binding to PUUV virions by ELISA and for their ability to neutralise PUUV infection *in vitro*. This process identified two anti-PUUV neutralising antibodies named 5A2 and 4G2 which appear to target Gn and Gc respectively. A similar study also identified a human antibody named 1C9 [148]. Later attempts were then made to categorise the neutralisation epitopes of these antibodies using phage display libraries of short peptides that could then be matched to the hantaviral glycoproteins Gn and Gc [131]. Similar studies have also been carried out using Hantaan virus to identify a number of anti-Hantaan antibodies [149]. Among these, is the anti-HNTV Ab named HCO2 which was later used in an *in vivo* protection study in hamsters [32]. This study demonstrates that passive transfer of recombinantly expressed antibodies can be used to prevent HNTV infection *in vivo*. More recently, two monoclonal antibodies were derived from a patient who had been infected with ANDV named JL16 and MIB22 [31]. This study demonstrated that passive transfer of these antibodies yielded complete protection from a lethal ANDV infection in hamsters. Studies such as this highlight the potential that antibody based therapeutics have for the treatment of hantaviral disease in humans.

Despite a large number of studies of anti-hantaviral antibodies, there is a lack of cross-reactive monoclonal antibodies that can neutralise a wide range of different hantavirus species [4]. There are also no detailed molecular descriptions of how these antibodies interact with hantaviral glycoproteins and how they exert their neutralising effect.

6.2. Neutralisation of Puumala virus by 4G2 fab

A commonly reported neutralising antibody against PUUV was obtained from the host reservoir, a bank vole [33]. This has been shown to protect against hantavirus infection *in vitro* [33]. However, the mechanism by which this antibody neutralises infection remains unknown. A fab fragment of 4G2 was produced by Dr Rissanen (Strubi) and a crystal structure obtained of this antibody in complex with the ectodomain of Puumala Gc (unpublished). This crystal structure reveals an interaction with the Gc that spans across domains I and II and is close to the proposed Gc homodimer interface. This homodimer interface is required to hold the four fold symmetric lattice together and likely plays an important role in inducing the membrane curvature required for virus budding. However, it is not clear from the crystal structure alone how this antibody interacts with the native hantaviral lattice. It is possible to model how the antibody may interact with the surface by docking the antibody onto the current Tula virus fitting but from this it is not obvious whether the antibody has space to fully clear the neighbouring spike or whether the fab sterically hinders Gc homodimerisation.

Firstly, to check if the 4G2 fab fragment is able to bind to PUUV VLP an ELISA was performed with immobilised PUUV VLP and a dilution series of 4G2 fab Figure 6.1. This ELISA demonstrates a high affinity interaction between PUUV VLP and the 4G2 fab in the low nM range.

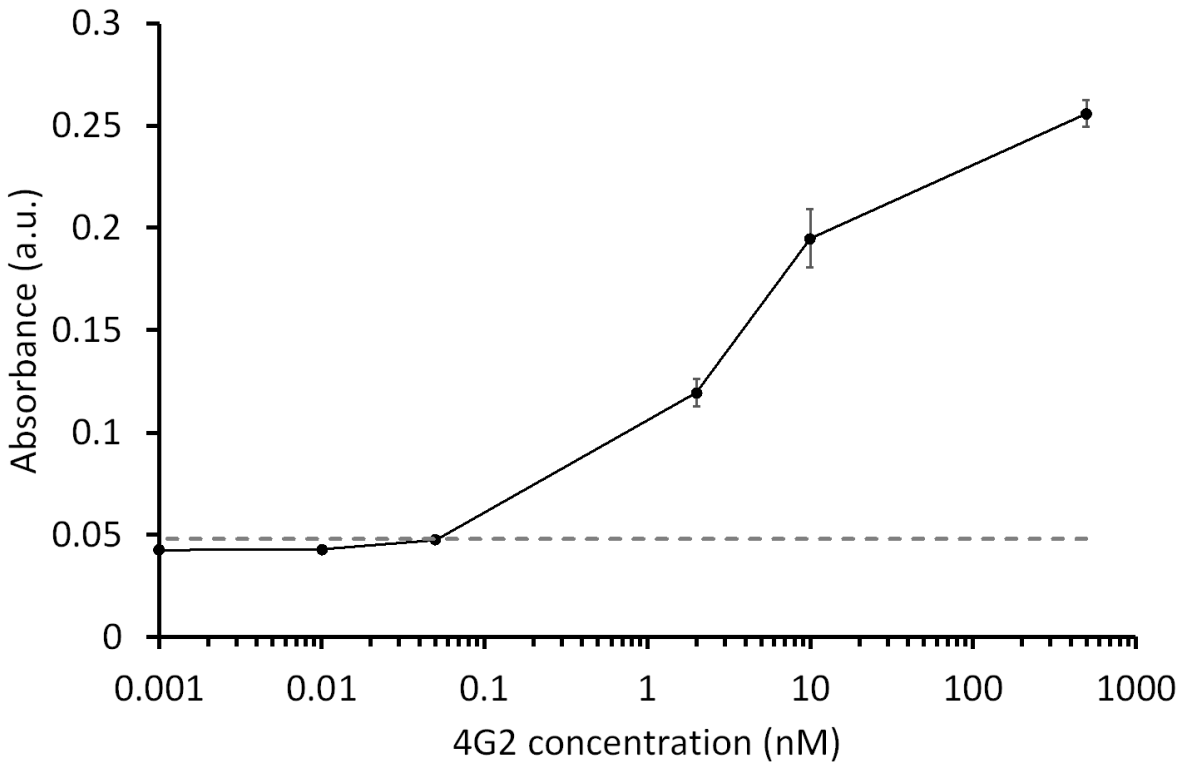


Figure 6.1 ELISA to demonstrate binding of 4G2 fab to PUUV VLPs
A sample purified PUUV VLPs were immobilised on a plate and probed with a dilution series of 4G2 fab. An analyte-free negative control well was included giving a background reading of 0.0479 as indicated with a grey dashed line. The 280nm absorbance of the original VLP sample was 0.91 a.u. which was diluted 250x before plating.

Next, cryo-EM grids were prepared with PUUV VLP and fab 4G2. Tomographic data was collected and the data processed using a standard workflow as previously described. The resulting reconstruction appears broadly similar to the TULV and HNTV reconstructions obtained previously, with no obvious extra density present corresponding to the fab (Figure 6.2 A). The first insights into why the fab is missing came from a repeat of the technique used for TULV (see Figure 4.10) in which break points in the lattice were manually selected and aligned, and reconstructed without symmetry applied. Like Tula, this revealed an empty space next to the central spike but with the addition of a small piece of density (Figure 6.2 B).

The standard hantavirus processing workflow that was used to process this data contains a step that restricts the data to include only particles that belong to a regular lattice (using the patch finder script detailed in section 3.5). A qualitative assessment of the number of particles after this restriction reveals fewer particles belonging to lattice than expected

based on experience with other hantaviral samples. This suggests that these particles are less decorated with complete hantaviral lattice than expected despite appearing fairly indistinguishable from regular VLPs when inspecting tomographic slices of the VLP. To investigate the regions of the surface that are not coated in lattice, the data was instead restricted to only include particles that are not found in the regions of lattice. The downside of this approach is that the dataset now includes many “bad” particles that contribute only noise. To mitigate this effect, only particles that meet a minimum cross-correlation threshold were included (>0.15). This reconstruction revealed a neighbour-less hantaviral spike fully decorated with fab around its perimeter. Subsequent refinements were then carried out using this a reference to maximise the number of particles picked up by the refinement and the same restrictions performed to produce the reconstruction in Figure 6.2. The PUUV Gc:4G2 crystal structure can be successfully docked into the map confirming the presence of the fab in the peripheral density.

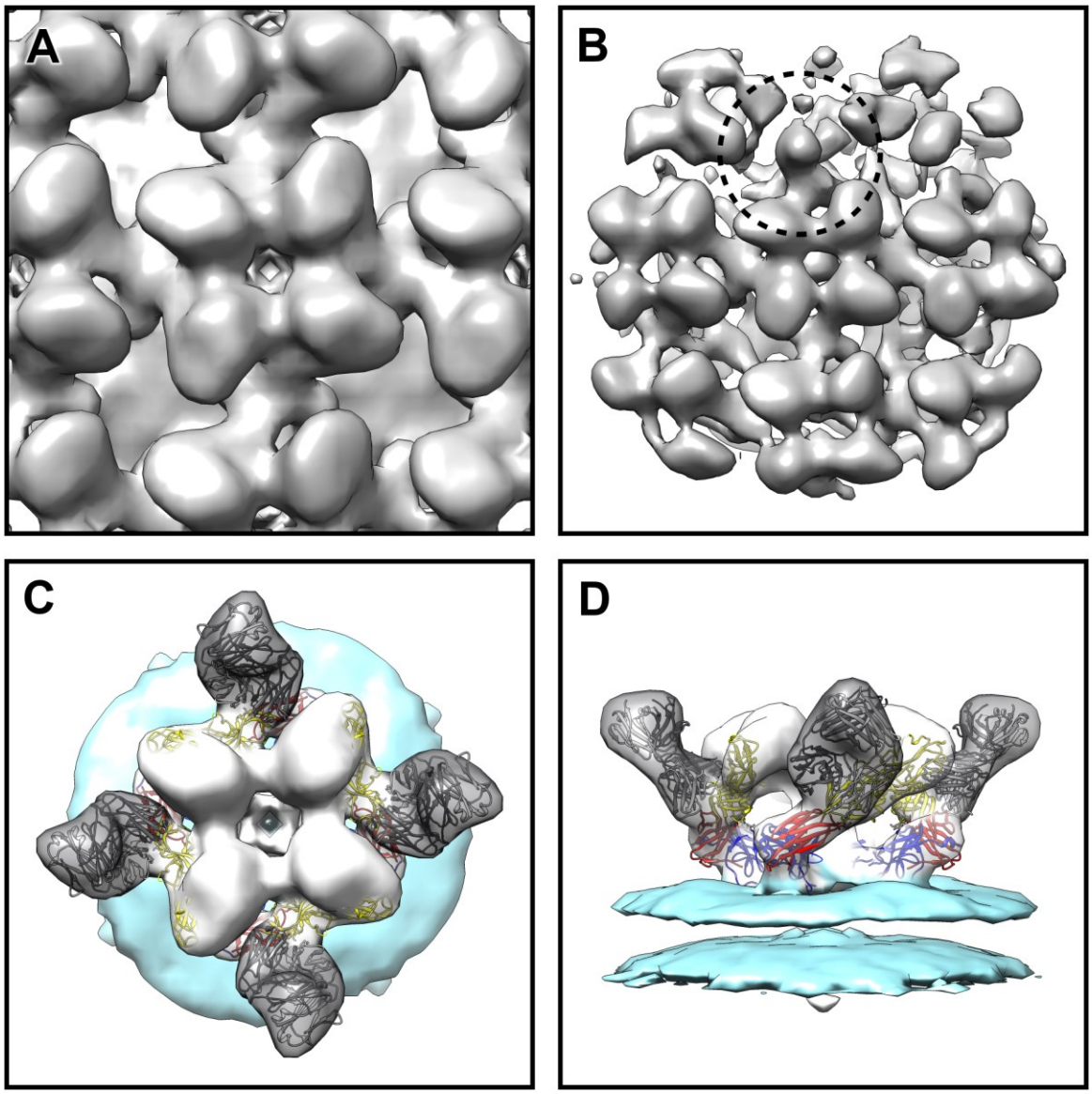


Figure 6.2 The interaction of 4G2 fab with PUUV VLPs
A) A reconstruction of PUUV VLPs mixed with fab using the standard hantavirus processing workflow including lattice restrictions. B) A reconstruction obtained from the refined positions in A where lattice break points were manually selected and aligned. Additional density can be seen in these membrane exposed break points as indicated. C-D) A reconstruction of a single PUUV VLP spike bound to fab 4G2 filtered to 15.0 Å resolution. Fitted in the map is the PUUV Gc: fab 4G2 crystal structure obtained by Dr Rissanen, Strubi, unpublished. Fitting was performed using the fitmap function in chimera using this model aligned to the PUUV Gn (PDB-5XFU, not displayed) based on the aforementioned ANDV Gn-Gc heterodimer model. The Gc model is coloured according to standard class II fusion colouring and the fab chains coloured grey. C) Topview. D) Side View.

To demonstrate the disruptive effect of the 4G2 antibody on the PUUV VLP lattice, a computational experiment was performed in which refinement of PUUV VLPs in the presence and absence of fab were carried out in parallel. Currently, a large dataset of PUUV VLPs on their own has not been acquired so the fab-free sample used in this test is the VLP sample that was mixed with candidate receptor protein gC1qR. This protein is small and

appears to have a minimal effect on the hantaviral lattice (see section 5.4) so is appropriate for use in this test. The data was also acquired with similar parameters to the 4G2 bound dataset so offers a fair comparison. A simple comparison of the whole VLP dataset is not possible. This is because these VLPs are not all completely decorated with spikes in the first place (For example see Figure 4.19 C-D) and also come in a variety of shapes. The VLPs also tend to stick to the air-water interface which sometimes, but not always, causes a loss of visible lattice on the top of the VLP. Therefore, the first step of this analysis required manual selection of eight roughly spherical, and visibly well decorated VLP from the two datasets. It should be noted that by doing so necessarily comes with a certain degree of human bias. However, the selection of only the best looking VLPs, does appear to have biased the 4G2 dataset towards VLPs with more complete lattice rather than the other way around.

Each set of eight VLPs were then refined using the standard workflow previously described, up to the point where the “patch finder” script is first used. The initial density of “seed” particles were kept constant between datasets. Following this the patch finder script was applied with standard parameters to locate particles found in lattice. The particles on the top and bottom of the virus were then removed from further analysis to prevent artefacts arising from the air-water interface. To do this the second euler angle, “tilt”, was restricted to be between 45 and 135 degrees. Finally the 3D locations of the particles before and after patch finder can be plotted to 2D via a Mercator projection as shown in Figure 6.3 and Figure 6.4 along with tomographic slices of the VLPs themselves. These plots demonstrate the completeness of the lattice present in the fab-free data set. In the 4G2 dataset there are some VLPs which appear to be completely covered in lattice like the fab-free dataset. However, there are other VLPs which, despite appearing to be well decorated, are either partially or not covered in any regions of lattice.

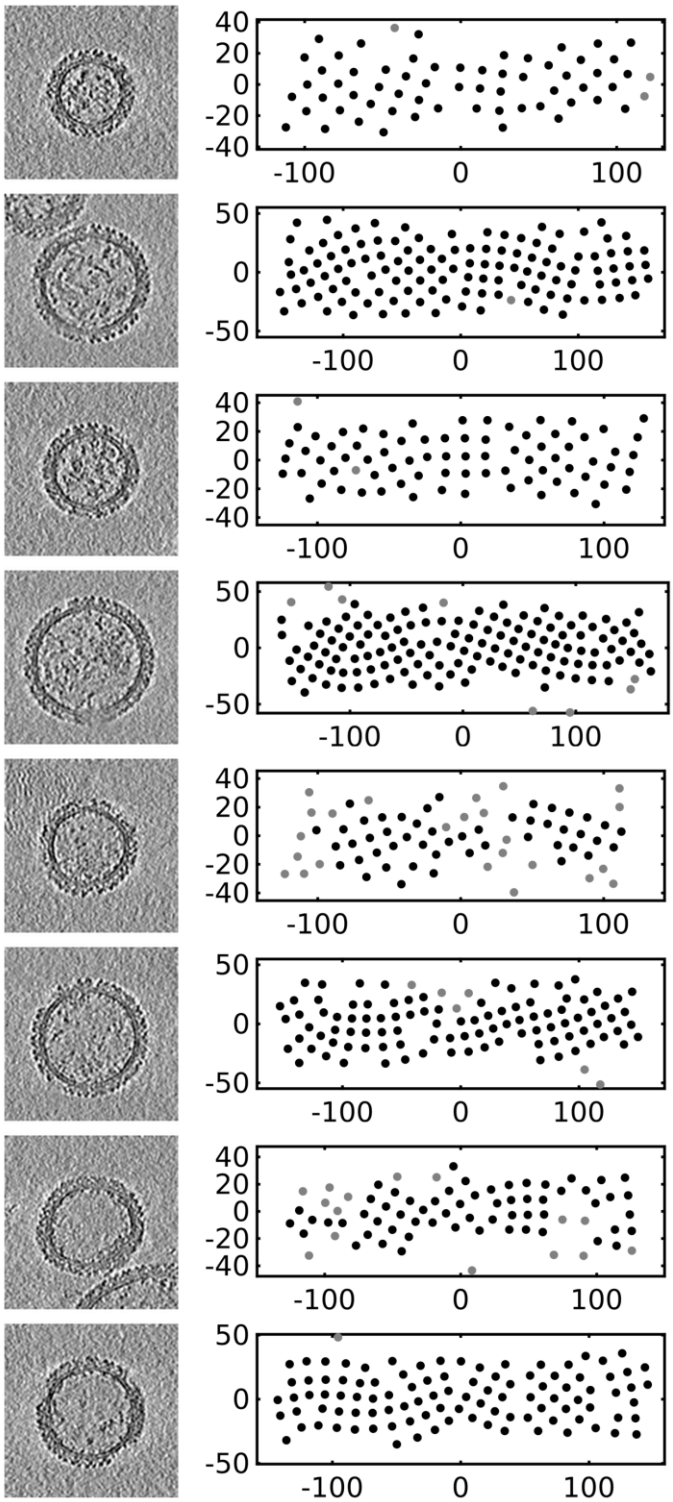


Figure 6.3 A demonstration of the lattice present on PUUV VLPs in the absence of fab 4G2
A computational experiment performed in which the eight VLPs displayed as tomographic slices on the left were refined using STA. These roughly spherical and well decorated VLPs were manually selected from a larger dataset for use in the experiment. The corresponding refined positions of the hantaviral spikes are displayed to the right as a Mercator projections (excluding particles on the top and bottom of the VLP. Units=nm). The subset of refined coordinates determined to belong to regular patches (using the patch finder script from section 3.5) are displayed in black. Note that in the absence of a large PUUV VLP dataset, the dataset used here is the PUUV VLP sample that has been mixed with candidate receptor gC1qR. This appears to have minimal impact on the lattice (See section 5.4 for details).

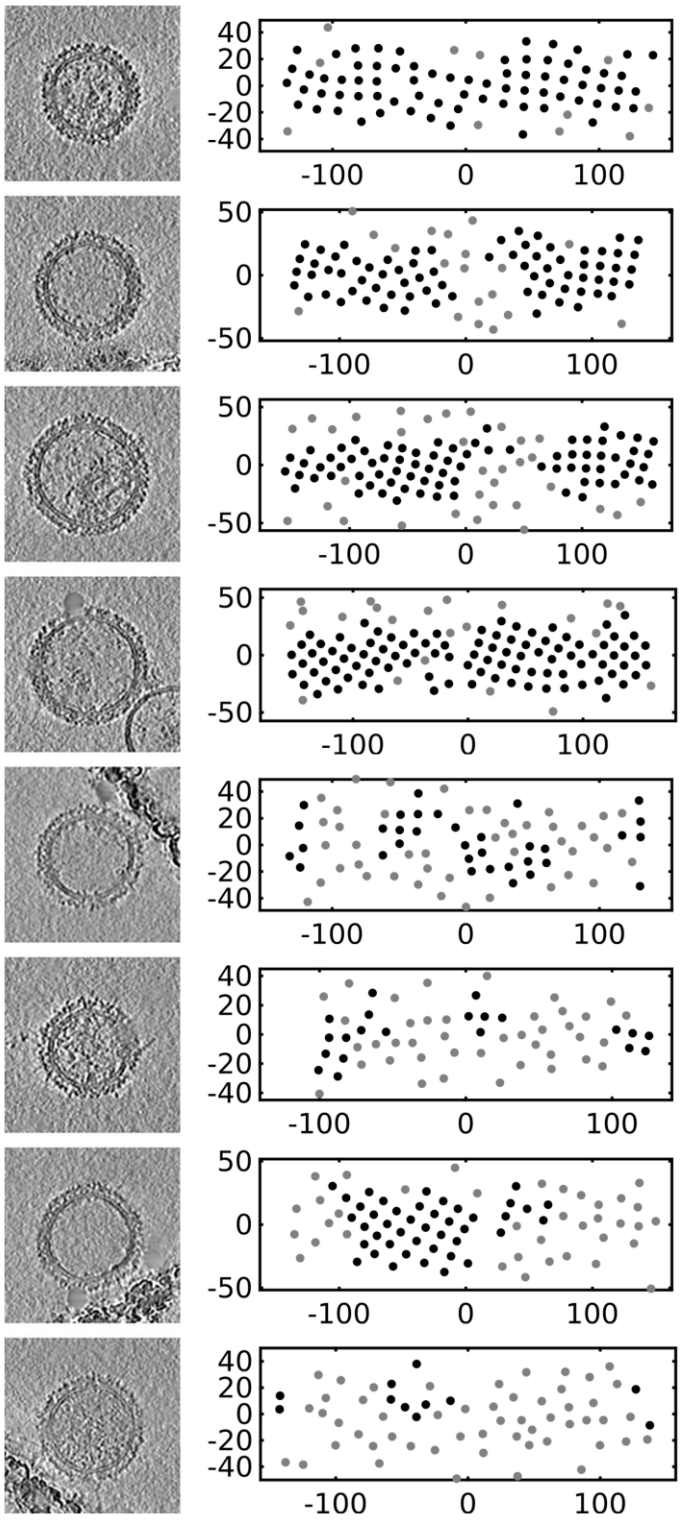


Figure 6.4 A demonstration of the lattice present on PUUV VLPs in the presence of fab 4G2
A computational experiment performed in which the eight VLPs displayed as tomographic slices on the left were refined using STA. These roughly spherical and well decorated VLPs were manually selected from a larger dataset for use in the experiment. The corresponding refined positions of the hantaviral spikes are displayed to the right as a Mercator projections (excluding particles on the top and bottom of the VLP. Units=nm). The subset of refined coordinates determined to belong to regular patches (using the patch finder script from section 3.5) are displayed in black.

The number of lattice-associated particles for each VLP is plotted in Figure 6.5 A which has been scaled to the size of the VLP. One of the parameters to the patch finder script is the minimum number of lattice-compatible neighbours that a particle must have to be classified as belonging to a patch. As standard, this value is set conservatively to 3 to maximise the number of particles found and to include, for example, a particle that is on the corner of a square lattice. However, patch finder can be re-run to output the particles found when this value is set to the full range of values from 0 up to the maximum of 8 (defined by the square of neighbours around the central particle). This data is plotted in Figure 6.5 B which demonstrates the difference in the distributions of the two data sets. The proportion of neighbour-less spikes in the 4G2 data set is higher than in the fab-free dataset which typically has a higher proportion of spikes with multiple neighbours. This demonstrates that incubation with the 4G2 fab does appear to reduce the number of spikes that are associated to neighbours via the formation of Gc homodimers.

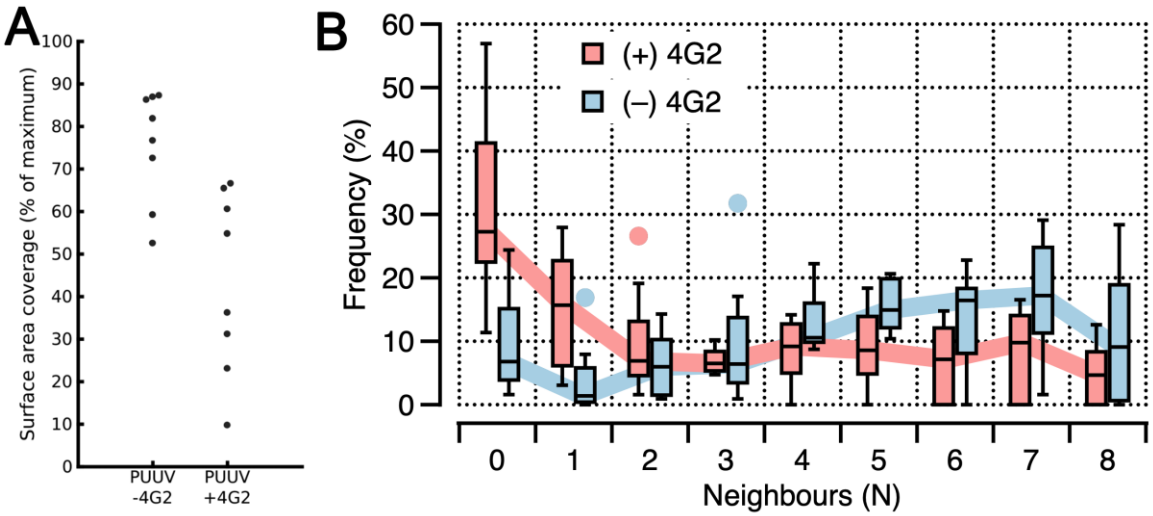


Figure 6.5 Comparison of the PUUV VLP lattice in the presence an absence of fab 4G2
The data displayed here is from a computational experiment performed in which the eight VLPs were refined as displayed in Figure 6.3 and Figure 6.4 above. A) A beeswarm plot of the number of particles determined to belong to patches using the patch finder script. Particles with at least 3 lattice-compatible neighbours were selected. The number of particles is scaled to the theoretical maximum number of particles that could fit on the surface of each VLP. The surface area of each VLP is obtained assuming spherical geometry based on the measured diameter. The theoretical maximum number of spikes is then obtained by dividing this by the approximate surface area of a single spike (196nm2). B) A box-and-whisker plot showing the number of particles that have a given number of lattice compatible neighbours from 0 to the maximum of 8, as determined using patch finder. The number of particles is displayed as a frequency scaled to the number of starting particles to account for variations in VLP size. Outliers, as determined by IQR are displayed. Note that in the absence of a large PUUV VLP dataset, the dataset used here is the PUUV VLP sample that has been mixed with candidate receptor gC1qR. This appears to have minimal impact on the lattice (See section 5.4 for details).

6.3. Neutralisation of Hantaan virus by the novel fab C4

Through collaboration with Dr Rissanen, Dr Bowden and Dr Doores (KCL), an attempt has been made to develop broadly neutralising antibodies by inoculating rabbits with the ectodomain of three different hantaviral Gns; HNTV, PUUV and ANDV. This project produced an antibody, named C4, that has been confirmed to be neutralising against live Hantaan virus by Prof. Botten and Annalis Norman at the University of Vermont. The neutralisation curve obtained by this group is displayed in Figure 6.6. Having demonstrated its neutralising ability, structural biology can be a powerful technique to elucidate the mechanism of action of this antibody.

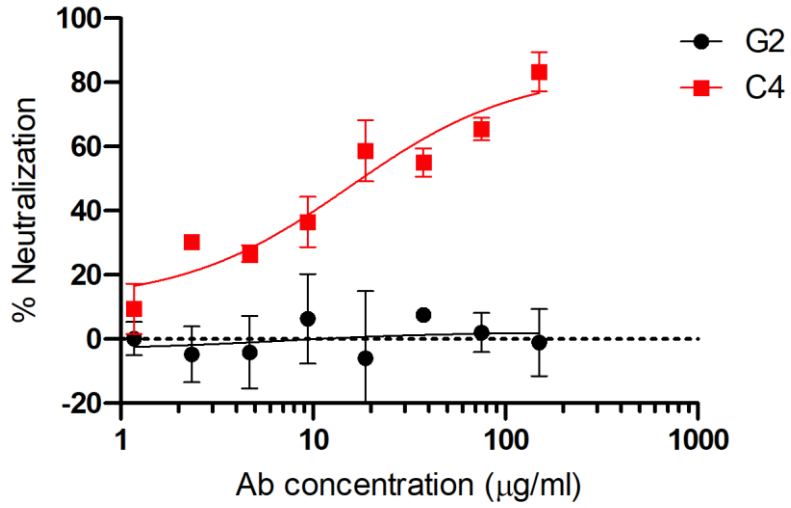


Figure 6.6 Neutralisation of live Hantaan virus by antibody C4
A live Hantaan virus neutralisation experiment performed with the neutralising Ab C4 as well as the non-neutralising Ab G2. This work was carried out in collaboration with Dr Doores, King's college London who supplied the figure and the experiment performed by Prof. Botten and Annalis Norman at the University of Vermont.

A fab fragment from the C4 antibody was produced by Dr Rissanen who then was able to obtain a crystal structure of the fab in complex with the HNTV Gn. This shows that the fab binds to a region of the Gn containing a loop comprised of residues 84-96. Using the Tula virus fitting presented previously, it is possible to model how this antibody would look when bound to the Hantavirus lattice. From this model it appeared that there would be very little clearance between it and the neighbouring Gn tetramer. It is also not immediately

obvious whether two fab fragments could both be bound to neighbouring spikes at the same time. The fab also appears to be bound to a region of the Gn that is in close contact with the fusion loops of the Gc. It is therefore not obvious whether the fab can bind to a Gn that is in complex with Gc.

To investigate the effect of the fab on the native hantaviral lattice I mixed the fab with a sample of HNTV VLPs. Figure 6.7 displays a number of tomographic slices through some of the VLPs imaged. These VLPs appear different to the unbound VLPs imaged previously and even give the impression of a disordered lattice. However, when these particles are fed through the standard hantavirus spike refining workflow, there is a reasonable coverage of lattice on the surface as demonstrated in the Mercator projections shown in Figure 6.7. To investigate whether temperature affects binding of the fab a smaller dataset was also collected that had been prepared at 21°C instead of 4°C. This revealed no clear and obvious difference in the coverage of lattice (See supplementary figure Figure 8.11).

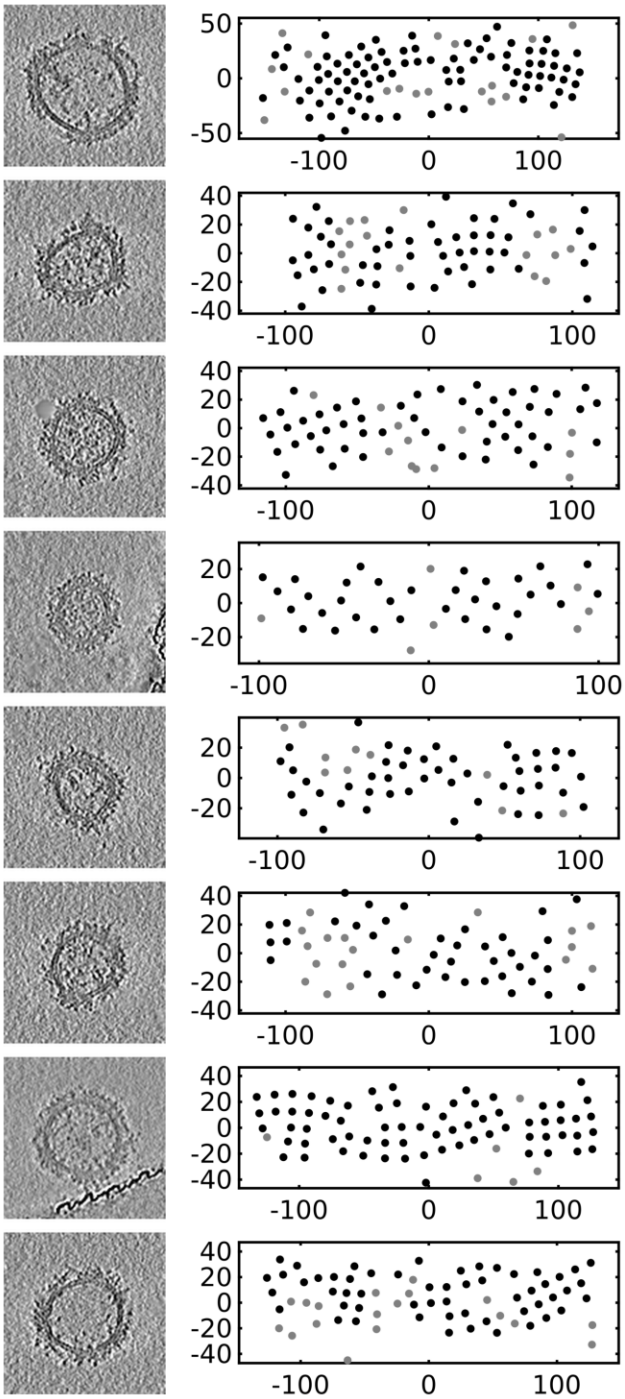


Figure 6.7 A demonstration of the lattice present on HNTV VLPs in the presence of fab C4
A computational experiment performed in which the eight VLPs displayed as tomographic slices on the left were refined using STA. These roughly spherical and well decorated VLPs were manually selected from a larger dataset for use in the experiment. The corresponding refined positions of the hantaviral spikes are displayed to the right as a Mercator projections (excluding particles on the top and bottom of the VLP. Units=nm). The subset of refined coordinates determined to belong to regular patches (using the patch finder script from section 3.5) are displayed in black. This sample was prepared at 4°C similarly to that reported in section 4.8.

The resulting reconstruction, with C4 symmetry applied, contains density in the inter-spike regions corresponding to the expected location of the fab from the crystal structure (Figure 6.8 A). However, this density was smaller and not the expected shape for a fab

fragment so it was clear that this reconstruction suffers from either partial occupancy or from incoherently averaged density resulting from a mismatch in symmetry. To resolve this the location of each spike was shifted to focus on the inter-spike area as this would leave just a single fab to classify instead of four. The orientations of these new shifted positions were randomised according to the new C2 symmetry and any overlapping particles removed. From this shifted location, I performed 3D classification of the subvolumes without symmetry imposed. The classification was able to distinguish two separate orientations of the fab fragment which made it possible to rotate one set of particles to match the other set, leaving only a single orientation of the fab present in the final reconstruction. The classification also revealed a population of completely unbound spikes consisting of 2777 particles that were removed from further processing. Further classification with the coherently aligned fab fragments reveal separate classes with fab fragments at a range of angles with respect to the spike, suggesting that there is a degree of flexibility in the hantaviral spike, at least at the site where the fab is bound (Figure 6.8 B-C). Just two of these classes (1 and 4) were selected to obtain the final reconstruction shown in Figure 6.8.

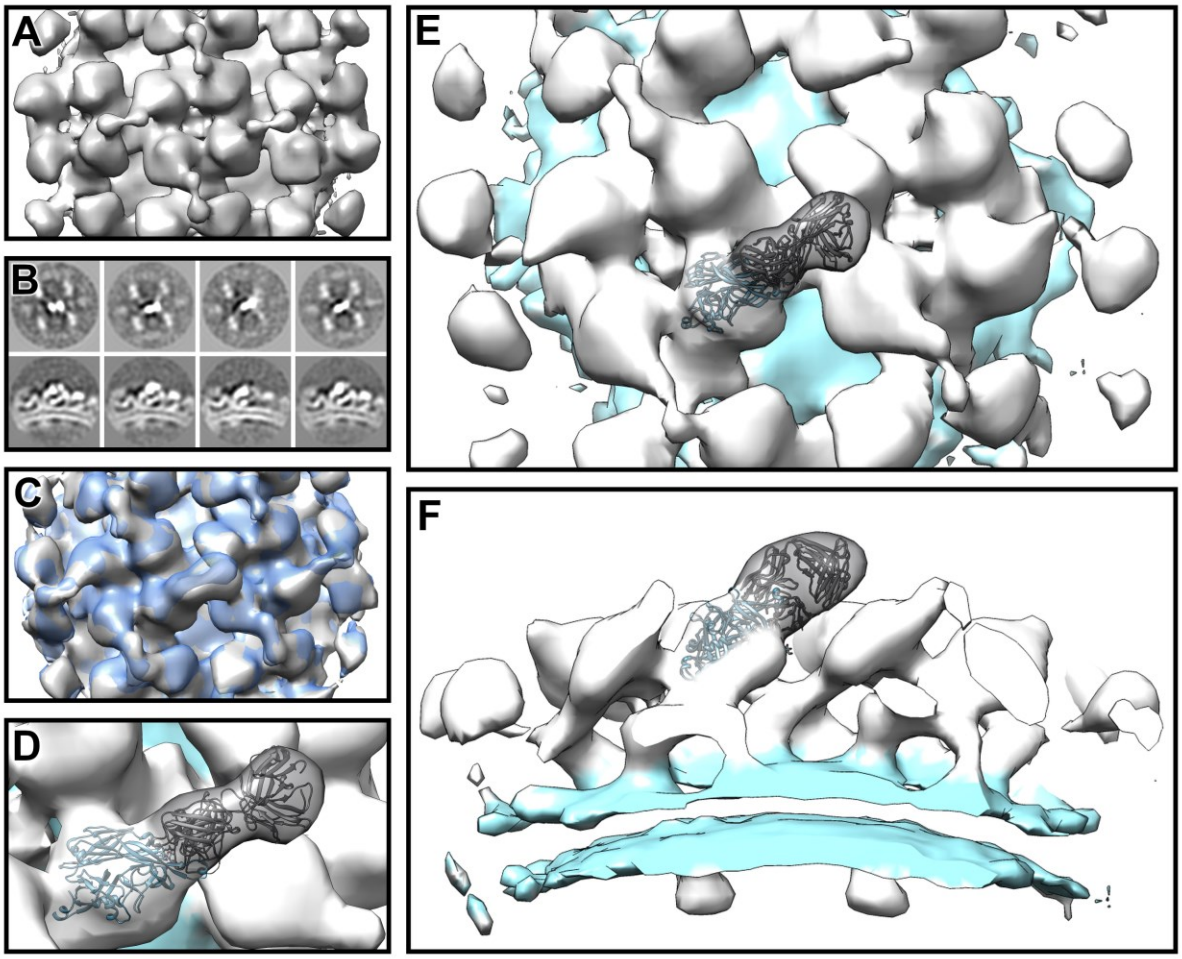


Figure 6.8 The interaction of antibody C4 with HNTV VLPs
A) The output from refinement using standard methods refined with C4 symmetry. B) 3D class averages obtained with relion to demonstrate the flexibility of the fab with respect to the lattice. The top are topviews sliced through the fab. The bottom row are central slices from side views. In order from left to right, number of particles in each class are; 893,1235,775,1487. A fifth class of just 228 particles is not shown. The final reconstruction was built using particles from the first and last classes. C) Isosurfaces from the 2nd and 4th classes in B overlaid on each other. D-F) A reconstruction of HNTV VLP bound to Ab C4 filtered to 19.0 Å. Note that the resolution estimation is based on a small masked region including the fab and half of each neighbouring spike. Fitted in the map is the HNTV Gn ectodomain:C4 fab complex crystal structure obtained by Dr Rissanen, Strubi, unpublished. This was fit into the map using the fitmap function in chimera. D) A magnified view, E) A top view. F) A cut-away side view.

Taken together, these results demonstrate that the fab is able to bind to the native lattice, without disrupting lattice contacts, and that only a single fab can bind at each inter-spike region. This mixture of orientations as well as the apparent flexibility of the contact site contribute to the disordered appearance of the VLP.

6.4. Discussion and future directions

Taking structural evidence from both the ectodomain crystal structure and the EM data presented here it is possible to speculate on the mechanism of action of the 4G2 antibody. It

appears this antibody is able to partially disrupt the lattice by disconnecting the Gc dimers that are responsible for spike-to-spike contacts. It is not clear whether 4G2 can do this actively or whether it merely stabilises the monomeric Gc that transiently forms on the hantaviral surface. It is known from [121] that the hantaviral spike is metastable, particularly around physiological temperatures and above, and that the Gc-Gc interface is weak [16]. It is conceivable that as the pleomorphic virus naturally changes shape, there are associated lattice rearrangements which due to the nature of the incomplete hantaviral lattice, exposes new monomeric Gc for 4G2 to bind to. In the future, it would be interesting to test this hypothesis by incubating the VLP with 4G2 at a range of temperatures to see if there is a greater reduction in lattice as temperature increases. The fact that 4G2 is able to disrupt Gc-Gc contacts may hint at the physiological importance of the Gc homodimers for the cell entry of hantaviruses.

Antibody 4G2 may also block virus fusion. It is expected that class II fusion proteins like the hantaviral Gc, form a “postfusion trimer” during the fusion process as the virus enters the cell [8], [125]. It is clear that the fab would sterically block the formation of this postfusion trimer which may prevent the fusion process (Figure 6.9 B). One unusual finding from the cryo-EM data is the existence of some, albeit less complete, lattice on the hantavirus surface. These regions contain fab-free Gc that is presumably still able to undergo fusion if triggered. One explanation for this may be that at physiological temperatures, the weak Gc interaction is less stable leading to a greater proportion of fab bound Gc. An alternative explanation may be that the Gc proteins buried deep in the hantaviral lattice are not fusogenic. Instead it may be only those Gc that surround break points in the lattice, where there is exposed membrane, that are able to fuse. Therefore, by specifically targeting the edges and break points in the lattice, 4G2 is able to achieve full neutralisation. A more detailed view of the fusion process would be required to confirm this hypothesis.

The region of the Gc attacked by 4G2 may be a site of vulnerability on the hantaviral surface utilised by other antibodies. A search through the literature uncovered two studies in

which escape mutations were found that prevent neutralisation of PUUV and HNTV by antibodies 1C9 and HCO2 respectively [150], [151]. 1C9 is a patient derived human antibody and HCO2 is an anti-HNTV antibody that has been demonstrated to be an effective post-exposure therapeutic in animal models [32]. The two escape mutations mapped onto the PUUV Gc are displayed in Figure 6.9 B and are in close proximity to the 4G2 site. It is also notable that attempts have been made in the past to locate the epitope of the 4G2 antibody using phage display libraries [152]. Two of these epitopes are displayed in green in Figure 6.9 A. This demonstrates the susceptibility of the phage display technique to false positives and highlights the need for validation from structural biology.

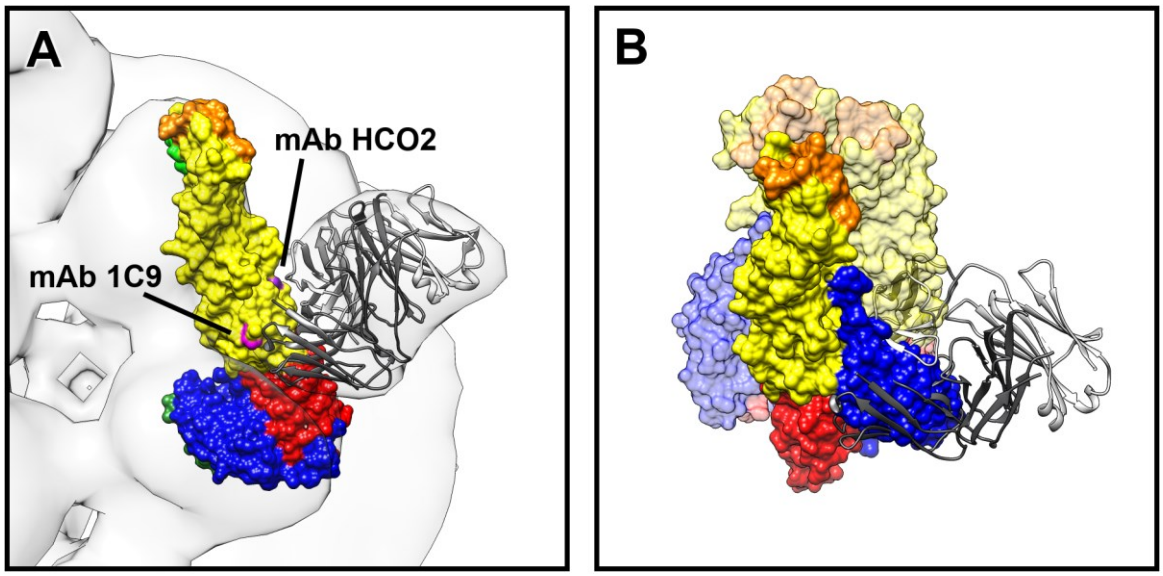


Figure 6.9 Common hantaviral immune epitopes and the 4G2 Ab mechanism of action
The PUUV Gc:4G2 fab model fitted into the EM map. The Gc is displayed as a surface model and coloured in standard class II fusion protein colours. Neutralisation evasion mutation sites from two other previously reported Abs, 1C9 and HCO2 are displayed in pink and Purple respectively [150], [151]. These residues are 944 and 805 respectively. Note that HCO2 is an anti-HNTV Ab so the site has been mapped to PUUV by sequence alignment. Two previously reported 4G2 epitopes from phage display experiments are displayed in green (residues 909-915 and 1057-1069) [152]. B) A demonstration of steric incompatibility with fab 4G2 and the post fusion trimer. The PUUV Gc : fab 4G2 model has been aligned to a PUUV Gc model in the post fusion conformation (PDB 5J81) and the original Gc removed. Clear steric clashes can be seen.

The C4 fab data presented here complements the x-ray structure obtained by Dr Rissanen and answer a few key questions about how the C4 fab binds to the native hantaviral lattice. It is clear that the fab is able to bind to the native lattice despite its close proximity to the neighbouring spike and that it is only able to bind in one orientation. The

precise mechanism by which the fab neutralises the virus remains unknown but there are a few possible candidates. One option, that cannot be ruled out is that the fab sterically hinders receptor binding. There is no structural information available for Hantaan virus receptor interactions so this option cannot be ruled out. Secondly, the fab binds to a site on the Gn that is very close to the fusion loops of the Gc. It is therefore possible that the fab sterically blocks fusion by preventing the fusion loops of the Gc from embedding into the target membrane. This however would require the Gn to remain bound to the Gc throughout the fusion process which, in the absence of a more detailed model of hantaviral fusion, is unknown.

Finally, the evidence presented here does tentatively suggest a mechanism by which the C4 fab could disrupt the native hantaviral lattice as has been observed for 4G2. It is known that the hantaviral spike is metastable, particularly around physiological temperatures, and that the Gc homodimer interface is weak [16], [121]. It is therefore possible that at higher temperatures more spikes may break free from the lattice and are able to become completely decorated in the fab fragment. Once fully decorated, the spikes would then be unable to re-form the native lattice because the fab is only able to bind to one of its neighbours. In the future, this hypothesis could be tested by incubating the VLP and the fab at higher temperatures before plunging and then assessing the lattice coverage as has been done for 4G2. The evidence from the 21°C experiment would suggest that this temperature would need to be higher than 21°C (Supplementary Figure 8.11).

The results presented here for fab 4G2 and C4 demonstrate the power of combining high resolution x-ray crystallography data with low resolution EM data for studying hantaviruses. The crystal structure can provide a detailed map of the residues involved at the contact site which are well beyond the resolutions achievable for subtomogram averaging. However, the structures of individual ectodomains cannot always predict exactly how the antibody will behave on the native virion, for example, whether 4G2 or C4 have space to bind to the native lattice. It can also suggest potential mechanisms for how the

antibody functions. Furthermore, the work with 4G2 specifically, and potentially C4 too, hint at the importance of the lattice breaks for hantavirus cell entry.

The original aim when raising the antibody C4 was to generate a broadly-neutralising antibody that can neutralise Old and New world hantaviruses. Currently, none of the antibodies produced from this study have yielded an antibody that can neutralise HNTV, PUUV and ANDV despite all three being used as antigens in the study. With thousands of candidates to choose from, it hoped that further work may yield such an antibody (Personal correspondence, Dr Doores, King's college London). An alternative, approach to achieve this goal may be to use the structural data available to make small mutations to the fab that allow it bind to more than one hantaviral species. It is hoped that attempts such as these could yield a broadly neutralising antibody that could then be tested *in vivo* and eventually form part of a much needed therapeutic to treat hantaviral disease.

7. References

[1] Y. Sun, J. Li, G. F. Gao, P. Tien, and W. Liu, “Bunyavirales ribonucleoproteins: the viral replication and transcription machinery,” *Crit. Rev. Microbiol.*, vol. 44, no. 5, pp. 522–540, Sep. 2018.

[2] T. S. Y. Guu, W. Zheng, and Y. J. Tao, “Bunyavirus: Structure and Replication,” Springer, Boston, MA, 2012, pp. 245–266.

[3] A. Vaheri *et al.*, “Uncovering the mysteries of hantavirus infections.,” *Nat. Rev. Microbiol.*, vol. 11, no. 8, pp. 539–50, Aug. 2013.

[4] R. L. Brocato and J. W. Hooper, “Progress on the Prevention and Treatment of Hantavirus Disease,” *Viruses*, vol. 11, no. 7, p. 610, Jul. 2019.

[5] M. Hilsch *et al.*, “Influenza A matrix protein M1 multimerizes upon binding to lipid membranes.,” *Biophys. J.*, vol. 107, no. 4, pp. 912–23, Aug. 2014.

[6] M. Iwasaki, M. Takeda, Y. Shirogane, Y. Nakatsu, T. Nakamura, and Y. Yanagi, “The matrix protein of measles virus regulates viral RNA synthesis and assembly by interacting with the nucleocapsid protein.,” *J. Virol.*, vol. 83, no. 20, pp. 10374–83, Oct. 2009.

[7] T. Strecker *et al.*, “Lassa Virus Z Protein Is a Matrix Protein Sufficient for the Release of Virus-Like Particles,” *J. Virol.*, vol. 77, no. 19, p. 10700, 2003.

[8] S. Halldorsson, S. Li, M. Li, K. Harlos, T. A. Bowden, and J. T. Huiskonen, “Shielding and activation of a viral membrane fusion protein.,” *Nat. Commun.*, vol. 9, no. 1, p. 349, 2018.

[9] S. Li *et al.*, “A Molecular-Level Account of the Antigenic Hantaviral Surface.,” *Cell Rep.*, vol. 15, no. 5, pp. 959–67, May 2016.

[10] D. A. Bente, N. L. Forrester, D. M. Watts, A. J. McAuley, C. A. Whitehouse, and M. Bray, “Crimean-Congo hemorrhagic fever: history, epidemiology, pathogenesis, clinical syndrome and genetic diversity.,” *Antiviral Res.*, vol. 100, no. 1, pp. 159–89, Oct. 2013.

[11] H. JP and R. F, “MRBAYES: Bayesian Inference of Phylogenetic Trees,” *Bioinformatics*, vol. 17, no. 8, 2001.

[12] K. S *et al.*, “Re-assessing the Diversity of Negative Strand RNA Viruses in Insects,” *PLoS Pathog.*, vol. 15, no. 12, 2019.

[13] J. M. White, S. E. Delos, M. Brecher, and K. Schornberg, “Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme.,” *Crit. Rev. Biochem. Mol. Biol.*, vol. 43, no. 3, pp. 189–219, Jan. .

[14] A. Grafmüller, J. Shillcock, and R. Lipowsky, “The fusion of membranes and vesicles: pathway and energy barriers from dissipative particle dynamics.,” *Biophys. J.*, vol. 96, no. 7, pp. 2658–75, Apr. 2009.

[15] Y. Modis, *Viral Entry into Host Cells*, vol. 790. New York, NY: Springer New York, 2013.

[16] P. Guardado-Calvo *et al.*, “Mechanistic Insight into Bunyavirus-Induced Membrane Fusion from Structure-Function Analyses of the Hantavirus Envelope Glycoprotein Gc.,” *PLoS Pathog.*, vol. 12, no. 10, p. e1005813, Oct. 2016.

[17] E. Baquero, A. A. Albertini, P. Vachette, J. Lepault, S. Bressanelli, and Y. Gaudin, “Intermediate conformations during viral fusion glycoprotein structural transition,” *Curr. Opin. Virol.*, vol. 3, no. 2, pp. 143–150, Apr. 2013.

[18] Y. Modis, S. Ogata, D. Clements, and S. C. Harrison, “Structure of the dengue virus envelope protein after membrane fusion,” *Nature*, vol. 427, no. 6972, pp. 313–319, Jan. 2004.

[19] O. Vapalahti, J. Mustonen, A. Lundkvist, H. Henttonen, A. Plyusnin, and A. Vaheri, “Hantavirus infections in Europe.,” *Lancet. Infect. Dis.*, vol. 3, no. 10, pp. 653–61, Oct. 2003.

[20] C. Martinez-Valdebenito *et al.*, “Person-to-person household and nosocomial transmission of andes hantavirus, Southern Chile, 2011.,” *Emerg. Infect. Dis.*, vol. 20, no. 10, pp. 1629–36, Oct. 2014.

[21] C. B. Jonsson, L. T. M. Figueiredo, and O. Vapalahti, “A global perspective on hantavirus ecology, epidemiology, and disease.,” *Clin. Microbiol. Rev.*, vol. 23, no. 2, pp. 412–41, Apr. 2010.

[22] H. W. Lee, P. W. Lee, and K. M. Johnson, “Isolation of the etiologic agent of Korean Hemorrhagic fever.,” *J. Infect. Dis.*, vol. 137, no. 3, pp. 298–308, Mar. 1978.

[23] J. Mustonen *et al.*, “The pathogenesis of nephropathia epidemica: New knowledge and unanswered questions,” *Antiviral Res.*, vol. 100, no. 3, pp. 589–604, Dec. 2013.

[24] I. Rissanen *et al.*, “Structural Transitions of the Conserved and Metastable Hantaviral Glycoprotein Envelope,” *J. Virol.*, vol. 91, no. 21, Nov. 2017.

[25] R. K. Rowe and A. Pekosz, “Bidirectional Virus Secretion and Nonciliated Cell Tropism following Andes Virus Infection of Primary Airway Epithelial Cell Cultures,” *J. Virol.*, vol. 80, no. 3, p. 1087, Feb. 2006.

[26] J. W. Golden, C. D. Hammerbeck, E. M. Mucker, and R. L. Brocato, “Animal Models for the Study of Rodent-Borne Hemorrhagic Fever Viruses: Arenaviruses and Hantaviruses,” *Biomed Res. Int.*, vol. 2015, 2015.

[27] P. Shrivastava-Ranjan, P. E. Rollin, and C. F. Spiropoulou, “Andes Virus Disrupts the Endothelial Cell Barrier by Induction of Vascular Endothelial Growth Factor and Downregulation of VE-Cadherin,” *J. Virol.*, vol. 84, no. 21, pp. 11227–11234, Nov. 2010.

[28] V. Matthys and E. R. Mackow, “Hantavirus Regulation of Type I Interferon Responses,” *Adv. Virol.*, vol. 2012, 2012.

[29] T. K. Outinen *et al.*, “The severity of Puumala hantavirus induced nephropathia epidemica can be better evaluated using plasma interleukin-6 than C-reactive protein determinations,” *BMC Infect. Dis.*, vol. 10, p. 132, 2010.

[30] G. Salazar, N. Zhang, T.-M. Fu, and Z. An, “Antibody therapies for the prevention and treatment of viral infections,” *npj Vaccines*, vol. 2, no. 1, p. 19, Dec. 2017.

[31] J. L. Garrido *et al.*, “Two recombinant human monoclonal antibodies that protect against lethal Andes hantavirus infection in vivo,” *Sci. Transl. Med.*, vol. 10, no. 468, p. eaat6420, Nov. 2018.

[32] M. LIANG, Y.-K. CHU, and C. SCHMALJOHN, “Bacterial Expression of Neutralizing Mouse Monoclonal Antibody Fab Fragments to Hantaan Virus,” *Virology*, vol. 217, no. 1, pp. 262–271, Mar. 1996.

[33] A. Lundkvist and B. Niklasson, “Bank vole monoclonal antibodies against Puumala

virus envelope glycoproteins: identification of epitopes involved in neutralization.," *Arch. Virol.*, vol. 126, no. 1–4, pp. 93–105, 1992.

[34] C.-F. Chiang, M. Flint, J.-M. S. Lin, and C. F. Spiropoulou, "Endocytic Pathways Used by Andes Virus to Enter Primary Human Lung Endothelial Cells.," *PLoS One*, vol. 11, no. 10, p. e0164768, 2016.

[35] M. Jin *et al.*, "Hantaan Virus Enters Cells by Clathrin-Dependent Receptor-Mediated Endocytosis," *Virology*, vol. 294, no. 1, pp. 60–69, Mar. 2002.

[36] N. Cifuentes-Muñoz, N. Salazar-Quiroz, and N. Tischler, "Hantavirus Gn and Gc Envelope Glycoproteins: Key Structural Units for Virus Cell Entry and Virus Assembly," *Viruses*, vol. 6, no. 4, pp. 1801–1822, Apr. 2014.

[37] C.-F. Chiang, M. Flint, J.-M. S. Lin, and C. F. Spiropoulou, "Endocytic Pathways Used by Andes Virus to Enter Primary Human Lung Endothelial Cells," *PLoS One*, vol. 11, no. 10, p. e0164768, Oct. 2016.

[38] J. Petersen *et al.*, "The major cellular sterol regulatory pathway is required for Andes virus infection.," *PLoS Pathog.*, vol. 10, no. 2, p. e1003911, Feb. 2014.

[39] J. Arikawa, I. Takashima, and N. Hashimoto, "Cell fusion by haemorrhagic fever with renal syndrome (HFRS) viruses and its application for titration of virus infectivity and neutralizing antibody," *Arch. Virol.*, vol. 86, no. 3–4, pp. 303–313, Sep. 1985.

[40] N. Cifuentes-Munoz, G. P. Barriga, P. D. T. Valenzuela, and N. D. Tischler, "Aromatic and polar residues spanning the candidate fusion peptide of the Andes virus Gc protein are essential for membrane fusion and infection," *J. Gen. Virol.*, vol. 92, no. 3, pp. 552–563, Mar. 2011.

[41] S. Rothenberger, G. Torriani, M. U. Johansson, S. Kunz, and O. Engler, "Conserved Endonuclease Function of Hantavirus L Polymerase.," *Viruses*, vol. 8, no. 5, 2016.

[42] D. Olal and O. Daumke, "Structure of the Hantavirus Nucleoprotein Provides Insights into the Mechanism of RNA Encapsidation," *Cell Rep.*, vol. 14, no. 9, pp. 2092–2099, Mar. 2016.

[43] R. Acuna *et al.*, "Hantavirus Gn and Gc Glycoproteins Self-Assemble into Virus-Like Particles," *J. Virol.*, vol. 88, no. 4, pp. 2344–2348, Feb. 2014.

[44] X. Shi and R. M. Elliott, "Analysis of N-linked glycosylation of hantaan virus glycoproteins and the role of oligosaccharide side chains in protein folding and intracellular trafficking.," *J. Virol.*, vol. 78, no. 10, pp. 5414–22, May 2004.

[45] J. Hepojoki, T. Strandin, H. Wang, O. Vapalahti, A. Vaheri, and H. Lankinen, "Cytoplasmic tails of hantavirus glycoproteins interact with the nucleocapsid protein," *J. Gen. Virol.*, vol. 91, no. 9, pp. 2341–2350, Sep. 2010.

[46] P. Gómez-Puertas, C. Albo, E. Pérez-Pastrana, A. Vivo, and A. Portela, "Influenza virus matrix protein is the major driving force in virus budding.," *J. Virol.*, vol. 74, no. 24, pp. 11538–47, Dec. 2000.

[47] E. V Ravkov, S. T. Nichol, and R. W. Compans, "Polarized entry and release in epithelial cells of Black Creek Canal virus, a New World hantavirus.," *J. Virol.*, vol. 71, no. 2, pp. 1147–54, Feb. 1997.

[48] C. S. Goldsmith, L. H. Elliott, C. J. Peters, and S. R. Zaki, "Ultrastructural characteristics of Sin Nombre virus, causative agent of hantavirus pulmonary syndrome.," *Arch. Virol.*, vol. 140, no. 12, pp. 2107–22, 1995.

[49] T. Strandin, J. Hepojoki, and A. Vaheri, "Cytoplasmic tails of bunyavirus Gn

glycoproteins—Could they act as matrix protein surrogates?," *Virology*, vol. 437, no. 2, pp. 73–80, Mar. 2013.

[50] J. T. Huiskonen *et al.*, "Electron cryotomography of Tula hantavirus suggests a unique assembly paradigm for enveloped viruses.," *J. Virol.*, vol. 84, no. 10, pp. 4889–97, May 2010.

[51] M. Kielian, "Class II virus membrane fusion proteins," *Virology*, vol. 344, no. 1, pp. 38–47, Jan. 2006.

[52] H. Song *et al.*, "Molecular Basis of Arthritogenic Alphavirus Receptor MXRA8 Binding to Chikungunya Virus Envelope Protein," *Cell*, vol. 177, no. 7, pp. 1714-1724.e12, Jun. 2019.

[53] J. Casals and G. H. Tignor, "The <i>Nairovirus </i>Genus: Serological Relationships," *Intervirology*, vol. 14, no. 3–4, pp. 144–147, 1980.

[54] P. J. Walker, S. G. Widen, T. G. Wood, H. Guzman, R. B. Tesh, and N. Vasilakis, "A Global Genomic Characterization of Nairoviruses Identifies Nine Discrete Genogroups with Distinctive Structural Characteristics and Host-Vector Associations," *Am. J. Trop. Med. Hyg.*, vol. 94, no. 5, pp. 1107–1122, May 2016.

[55] J. V. Dzimianski *et al.*, "Probing the impact of nairovirus genomic diversity on viral ovarian tumor domain protease (vOTU) structure and deubiquitinase activity," *PLOS Pathog.*, vol. 15, no. 1, p. e1007515, Jan. 2019.

[56] B. I. Marczinke and S. T. Nichol, "Nairobi Sheep Disease Virus, an Important Tick-Borne Pathogen of Sheep and Goats in Africa, Is Also Present in Asia," *Virology*, vol. 303, no. 1, pp. 146–151, Nov. 2002.

[57] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, "MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods," *Mol. Biol. Evol.*, vol. 28, no. 10, p. 2731, 2011.

[58] S. S. Al-Abri *et al.*, "Current status of Crimean-Congo haemorrhagic fever in the World Health Organization Eastern Mediterranean Region: issues, challenges, and future directions," *Int. J. Infect. Dis.*, vol. 58, pp. 82–89, May 2017.

[59] A. K. Celikbas *et al.*, "Crimean-Congo hemorrhagic fever among health care workers, Turkey.," *Emerg. Infect. Dis.*, vol. 20, no. 3, pp. 477–9, Mar. 2014.

[60] J. R. Spengler and D. A. Bente, "Crimean-Congo Hemorrhagic Fever in Spain - New Arrival or Silent Resident?," *N. Engl. J. Med.*, vol. 377, no. 2, pp. 106–108, Jul. 2017.

[61] A. Negredo *et al.*, "Autochthonous Crimean–Congo Hemorrhagic Fever in Spain," *N. Engl. J. Med.*, vol. 377, no. 2, pp. 154–161, Jul. 2017.

[62] B. S, L. G, K. H, A. S, K. J, and M. A, "Crimean-Congo Hemorrhagic Fever Virus Infection Is Lethal for Adult Type I Interferon Receptor-Knockout Mice," *J. Gen. Virol.*, vol. 91, no. Pt 6, 2010.

[63] A. R. Garrison, D. R. Smith, and J. W. Golden, "Animal Models for Crimean-Congo Hemorrhagic Fever Human Disease," *Viruses*, vol. 11, no. 7, 2019.

[64] F. Begum, C. L. Wisseman, and J. Casals, "Tick-borne viruses of West Pakistan. II. Hazara virus, a new agent isolated from Ixodes redikorzevi ticks from the Kaghan Valley, W. Pakistan.," *Am. J. Epidemiol.*, vol. 92, no. 3, pp. 192–4, Sep. 1970.

[65] S. D. Dowall *et al.*, "Hazara virus infection is lethal for adult type I interferon receptor-knockout mice and may act as a surrogate for infection with the human-pathogenic

Crimean–Congo hemorrhagic fever virus," *J. Gen. Virol.*, vol. 93, no. 3, pp. 560–564, Mar. 2012.

[66] E. Akıncı, H. Bodur, and H. Leblebicioglu, "Pathogenesis of Crimean-Congo Hemorrhagic Fever," *Vector-Borne Zoonotic Dis.*, vol. 13, no. 7, pp. 429–437, Jul. 2013.

[67] O. Ergonul, S. Tuncbilek, N. Baykam, A. Celikbas, and B. Dokuzoguz, "Evaluation of Serum Levels of Interleukin (IL)–6, IL- 10, and Tumor Necrosis Factor–α in Patients with Crimean- Congo Hemorrhagic Fever," *J. Infect. Dis.*, vol. 193, no. 7, pp. 941–944, Apr. 2006.

[68] A. Saksida, D. Duh, B. Wraber, I. Dedushaj, S. Ahmeti, and T. Avsic-Zupanc, "Interacting roles of immune mechanisms and viral load in the pathogenesis of crimean-congo hemorrhagic fever.," *Clin. Vaccine Immunol.*, vol. 17, no. 7, pp. 1086–93, Jul. 2010.

[69] X. Xiao, Y. Feng, Z. Zhu, and D. S. Dimitrov, "Identification of a putative Crimean-Congo hemorrhagic fever virus entry factor," *Biochem. Biophys. Res. Commun.*, vol. 411, no. 2, pp. 253–258, Jul. 2011.

[70] S. Bose, M. Basu, and A. K. Banerjee, "Role of Nucleolin in Human Parainfluenza Virus Type 3 Infection of Human Lung Epithelial Cells," *J. Virol.*, vol. 78, no. 15, pp. 8146–8158, Aug. 2004.

[71] S. Nisole *et al.*, "The Anti-HIV Pentameric Pseudopeptide HB-19 Binds the C-terminal End of Nucleolin and Prevents Anchorage of Virus Particles in the Plasma Membrane of Target Cells," *J. Biol. Chem.*, vol. 277, no. 23, pp. 20877–20886, Jun. 2002.

[72] U. R. de Verdugo *et al.*, "Characterization of a 100-kilodalton binding protein for the six serotypes of coxsackie B viruses.," *J. Virol.*, vol. 69, no. 11, pp. 6751–7, Nov. 1995.

[73] A. Bertolotti-Ciarlet *et al.*, "Cellular Localization and Antigenic Characterization of Crimean-Congo Hemorrhagic Fever Virus Glycoproteins," *J. Virol.*, vol. 79, no. 10, pp. 6152–6161, May 2005.

[74] A. R. Garrison *et al.*, "Crimean–Congo hemorrhagic fever virus utilizes a clathrin- and early endosome-dependent entry pathway," *Virology*, vol. 444, no. 1–2, pp. 45–54, Sep. 2013.

[75] O. Shtanko, R. A. Nikitina, C. Z. Altuntas, A. A. Chepurnov, and R. A. Davey, "Crimean-Congo Hemorrhagic Fever Virus Entry into Host Cells Occurs through the Multivesicular Body and Requires ESCRT Regulators," *PLoS Pathog.*, vol. 10, no. 9, p. e1004390, Sep. 2014.

[76] A. J. Sanchez, M. J. Vincent, B. R. Erickson, and S. T. Nichol, "Crimean-congo hemorrhagic fever virus glycoprotein precursor is cleaved by Furin-like and SKI-1 proteases to generate a novel 38-kilodalton glycoprotein.," *J. Virol.*, vol. 80, no. 1, pp. 514–25, Jan. 2006.

[77] D. F. Estrada and R. N. De Guzman, "Structural characterization of the Crimean-Congo hemorrhagic fever virus Gn tail provides insight into virus assembly.," *J. Biol. Chem.*, vol. 286, no. 24, pp. 21678–86, Jun. 2011.

[78] G. Simmons, R. J. Wool-Lewis, F. Baribaud, R. C. Netter, and P. Bates, "Ebola virus glycoproteins induce global surface protein down-modulation and loss of cell adherence.," *J. Virol.*, vol. 76, no. 5, pp. 2518–28, Mar. 2002.

[79] N. Altgärde *et al.*, "Mucin-like Region of Herpes Simplex Virus Type 1 Attachment

Protein Glycoprotein C (gC) Modulates the Virus-Glycosaminoglycan Interaction," *J. Biol. Chem.*, vol. 290, no. 35, pp. 21473–21485, Aug. 2015.

[80] H. H. Tantawi *et al.*, "Crimean-Congo haemorrhagic fever virus in Iraq: isolation, identification and electron microscopy.," *Acta Virol.*, vol. 24, no. 6, pp. 464–7, Dec. 1980.

[81] A. Jelínková, R. Benda, and M. Novák, "Electron microscopic demonstration of Crimean hemorrhagic fever virus in CV-1 cells.," *Acta Virol.*, vol. 19, no. 5, pp. 369–73, Sep. 1975.

[82] S. Shimada *et al.*, "Tofla virus: A newly identified Nairovirus of the Crimean-Congo hemorrhagic fever group isolated from ticks in Japan.," *Sci. Rep.*, vol. 6, p. 20213, Feb. 2016.

[83] M. S. Smyth and J. H. J. Martin, "x Ray crystallography," *Mol. Pathol.*, vol. 53, no. 1, p. 8, Feb. 2000.

[84] D. Marion, "An Introduction to Biological NMR Spectroscopy," *Mol. Cell. Proteomics*, vol. 12, no. 11, pp. 3006–3025, Nov. 2013.

[85] R. Henderson, "From Electron Crystallography to Single Particle CryoEM (Nobel Lecture)," *Angew. Chemie Int. Ed.*, vol. 57, no. 34, pp. 10804–10825, Aug. 2018.

[86] E. V. Orlova and H. R. Saibil, "Structural Analysis of Macromolecular Assemblies by Electron Microscopy," *Chem. Rev.*, vol. 111, no. 12, pp. 7710–7748, Dec. 2011.

[87] S. De Carlo and J. R. Harris, "Negative staining and Cryo-negative Staining of Macromolecules and Viruses for TEM," *Micron*, vol. 42, no. 2, p. 117, 2011.

[88] J. Dubochet *et al.*, "Cryo-electron microscopy of vitrified specimens.," *Q. Rev. Biophys.*, vol. 21, no. 2, pp. 129–228, May 1988.

[89] S. Asano, B. D. Engel, and W. Baumeister, "In Situ Cryo-Electron Tomography: A Post-Reductionist Approach to Structural Biology," *J. Mol. Biol.*, vol. 428, no. 2, pp. 332–343, Jan. 2016.

[90] R. H. Wade, "A brief look at imaging and contrast transfer," *Ultramicroscopy*, vol. 46, no. 1–4, pp. 145–156, Oct. 1992.

[91] A. Rohou and N. Grigorieff, "CTFFIND4: Fast and accurate defocus estimation from electron micrographs," *J. Struct. Biol.*, vol. 192, no. 2, pp. 216–221, Nov. 2015.

[92] S. Wu, J.-P. Armache, and Y. Cheng, "Single-particle cryo-EM data acquisition by using direct electron detection camera.," *Reprod. Syst. Sex. Disord.*, vol. 65, no. 1, pp. 35–41, Feb. 2016.

[93] A. R. Aricescu, W. Lu, and E. Y. Jones, "A time- and cost-efficient system for high-level protein production in mammalian cells," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 62, no. 10, pp. 1243–1250, Oct. 2006.

[94] O. Vapalahti *et al.*, "Vaccinia virus-free rescue of fluorescent replication-defective vesicular stomatitis virus and pseudotyping with Puumala virus glycoproteins for use in neutralization tests," *J. Gen. Virol.*, vol. 97, no. 5, pp. 1052–1059, May 2016.

[95] S. Zheng, E. Palovcak, J.-P. Armache, Y. Cheng, and D. Agard, "Anisotropic Correction of Beam-induced Motion for Improved Single-particle Electron Cryo-microscopy," *bioRxiv*, 2016.

[96] T. Grant and N. Grigorieff, "Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6.," *Elife*, vol. 4, p. e06980, May 2015.

[97] T. A. M. Bharat, C. J. Russo, J. Löwe, L. A. Passmore, and S. H. W. Scheres, “Advances in Single-Particle Electron Cryomicroscopy Structure Determination applied to Sub-tomogram Averaging,” *Structure*, vol. 23, no. 9, pp. 1743–1753, Sep. 2015.

[98] D. Castaño-Díez, M. Kudryashev, M. Arheit, and H. Stahlberg, “Dynamo: A flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments,” *J. Struct. Biol.*, vol. 178, no. 2, pp. 139–151, May 2012.

[99] S. H. W. Scheres, “RELION: implementation of a Bayesian approach to cryo-EM structure determination.,” *J. Struct. Biol.*, vol. 180, no. 3, pp. 519–30, Dec. 2012.

[100] J. B. Heymann and D. M. Belnap, “Bsoft: Image processing and molecular modeling for electron microscopy,” *J. Struct. Biol.*, vol. 157, no. 1, pp. 3–18, Jan. 2007.

[101] K. Murata and M. Wolf, “Cryo-electron microscopy for structural analysis of dynamic biological macromolecules,” *Biochim. Biophys. Acta - Gen. Subj.*, vol. 1862, no. 2, pp. 324–334, Feb. 2018.

[102] T. A. M. Bharat, C. J. Russo, J. Löwe, L. A. Passmore, and S. H. W. Scheres, “Advances in Single-Particle Electron Cryomicroscopy Structure Determination applied to Sub-tomogram Averaging.,” *Structure*, vol. 23, no. 9, pp. 1743–53, Sep. 2015.

[103] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, “cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination,” *Nat. Methods*, vol. 14, no. 3, pp. 290–296, Mar. 2017.

[104] M. A. Herzik, M. Wu, and G. C. Lander, “High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM,” *Nat. Commun.*, vol. 10, no. 1, p. 1032, Dec. 2019.

[105] R. Stass, S. L. Ilca, and J. T. Huiskonen, “Beyond structures of highly symmetric purified viral capsids by cryo-EM,” *Curr. Opin. Struct. Biol.*, vol. 52, pp. 25–31, Oct. 2018.

[106] J. R. Kremer, D. N. Mastronarde, and J. R. McIntosh, “Computer visualization of three-dimensional image data using IMOD.,” *J. Struct. Biol.*, vol. 116, no. 1, pp. 71–6, 1996.

[107] D. N. Mastronarde and S. R. Held, “Automated tilt series alignment and tomographic reconstruction in IMOD.,” *J. Struct. Biol.*, vol. 197, no. 2, pp. 102–113, 2017.

[108] J. M. Heumann, A. Hoenger, and D. N. Mastronarde, “Clustering and variance maps for cryo-electron tomography using wedge-masked differences,” *J. Struct. Biol.*, vol. 175, no. 3, pp. 288–299, Sep. 2011.

[109] B. A. Himes and P. Zhang, “emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging,” *Nat. Methods*, vol. 15, no. 11, pp. 955–961, Nov. 2018.

[110] R. Henderson and R. M. Glaeser, “Quantitative analysis of image contrast in electron micrographs of beam-sensitive crystals,” *Ultramicroscopy*, vol. 16, no. 2, pp. 139–150, Jan. 1985.

[111] J.-J. Fernandez, S. Li, T. A. M. Bharat, and D. A. Agard, “Cryo-tomography tilt-series alignment with consideration of the beam-induced sample motion.,” *J. Struct. Biol.*, vol. 202, no. 3, pp. 200–209, 2018.

[112] S. Chen *et al.*, “High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy.,” *Ultramicroscopy*, vol. 135, pp. 24–35, Dec. 2013.

[113] C. J. Russo and L. A. Passmore, “Ultrastable gold substrates: Properties of a support for high-resolution electron cryomicroscopy of biological specimens.,” *J. Struct. Biol.*, vol. 193, no. 1, pp. 33–44, Jan. 2016.

[114] B. Turoňová, F. K. M. Schur, W. Wan, and J. A. G. Briggs, “Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å,” *J. Struct. Biol.*, vol. 199, no. 3, pp. 187–195, Sep. 2017.

[115] D. Tegunov and P. Cramer, “Real-time cryo-electron microscopy data preprocessing with Warp,” *Nat. Methods*, pp. 1–7, Oct. 2019.

[116] K. Zhang, “Gctf: Real-time CTF determination and correction,” *J. Struct. Biol.*, vol. 193, no. 1, pp. 1–12, Jan. 2016.

[117] F. K. M. Schur *et al.*, “An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation,” *Science (80-.).*, vol. 353, no. 6298, pp. 506–508, Jul. 2016.

[118] P. B. Rosenthal and R. Henderson, “Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy,” *J. Mol. Biol.*, vol. 333, no. 4, pp. 721–745, Oct. 2003.

[119] J. Zivanov *et al.*, “New tools for automated high-resolution cryo-EM structure determination in RELION-3,” *Elife*, vol. 7, Nov. 2018.

[120] C. V Sindelar and N. Grigorieff, “An adaptation of the Wiener filter suitable for analyzing images of isolated single particles.,” *J. Struct. Biol.*, vol. 176, no. 1, pp. 60–74, Oct. 2011.

[121] E. A. Bignon, A. Albornoz, P. Guardado-Calvo, F. A. Rey, and N. D. Tischler, “Molecular organization and dynamics of the fusion protein Gc at the hantavirus surface,” *Elife*, vol. 8, Jun. 2019.

[122] C. J. Russo and L. A. Passmore, “Robust evaluation of 3D electron cryomicroscopy data using tilt-pairs,” *J. Struct. Biol.*, vol. 187, no. 2, pp. 112–118, Aug. 2014.

[123] T. A. Bowden, D. Bitto, A. McLees, C. Yeromonahos, R. M. Elliott, and J. T. Huiskonen, “Orthobunyavirus ultrastructure and the curious tripodal glycoprotein spike.,” *PLoS Pathog.*, vol. 9, no. 5, p. e1003374, 2013.

[124] A. J. Battisti, Y.-K. Chu, P. R. Chipman, B. Kaufmann, C. B. Jonsson, and M. G. Rossmann, “Structural Studies of Hantaan Virus,” *J. Virol.*, vol. 85, no. 2, pp. 835–841, Jan. 2011.

[125] Y. Modis, S. Ogata, D. Clements, and S. C. Harrison, “Structure of the dengue virus envelope protein after membrane fusion,” *Nature*, vol. 427, no. 6972, pp. 313–319, Jan. 2004.

[126] C. E. Garry and R. F. Garry, “Proteomics computational analyses suggest that the carboxyl terminal glycoproteins of Bunyaviruses are class II viral fusion protein (beta-penetrenes).,” *Theor. Biol. Med. Model.*, vol. 1, no. 1, p. 10, Nov. 2004.

[127] M. Dessau and Y. Modis, “Crystal structure of glycoprotein C from Rift Valley fever virus,” *Proc. Natl. Acad. Sci.*, vol. 110, no. 5, pp. 1696–1701, Jan. 2013.

[128] E. K. Punch *et al.*, “Potassium is a trigger for conformational change in the fusion

spike of an enveloped RNA virus," *J. Biol. Chem.*, vol. 293, no. 26, pp. 9937–9944, Jun. 2018.

[129] Y. Sugita, T. Noda, H. Sagara, and Y. Kawaoka, "Ultracentrifugation deforms unfixed influenza A virions," *J. Gen. Virol.*, vol. 92, no. 11, pp. 2485–2493, Nov. 2011.

[130] G. Kiss *et al.*, "Capturing Enveloped Viruses on Affinity Grids for Downstream Cryo-Electron Microscopy Applications," *Microsc. Microanal.*, vol. 20, no. 1, pp. 164–174, Feb. 2014.

[131] T. Heiskanen, Å. Lundkvist, R. Soliymani, E. Koivunen, A. Vaheri, and H. Lankinen, "Phage-Displayed Peptides Mimicking the Discontinuous Neutralization Sites of Puumala Hantavirus Envelope Glycoproteins," *Virology*, vol. 262, no. 2, pp. 321–332, Sep. 1999.

[132] T. Heiskanen *et al.*, "Improvement of binding of Puumala virus neutralization site resembling peptide with a second-generation phage library," *Protein Eng. Des. Sel.*, vol. 16, no. 6, pp. 443–450, Jun. 2003.

[133] T. Raymond, E. Gorbunova, I. N. Gavrilovskaya, and E. R. Mackow, "Pathogenic hantaviruses bind plexin–semaphorin–integrin c αvβ3 integrin conformers," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 4, p. 1163, 2005.

[134] I. N. Gavrilovskaya, M. Shepley, R. Shaw, M. H. Ginsberg, and E. R. Mackow, "beta3 Integrins mediate the cellular entry of hantaviruses that cause respiratory failure.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 12, pp. 7074–9, Jun. 1998.

[135] R. Acuna *et al.*, "Hantavirus Gn and Gc Glycoproteins Self-Assemble into Virus-Like Particles," *J. Virol.*, vol. 88, no. 4, pp. 2344–2348, Feb. 2014.

[136] R. K. Jangra *et al.*, "Protocadherin-1 is essential for cell entry by New World hantaviruses," *Nature*, vol. 563, no. 7732, pp. 559–563, Nov. 2018.

[137] X. Zhang *et al.*, "Cryo-EM structure of the mature dengue virus at 3.5-Å resolution," *Nat. Struct. Mol. Biol.*, vol. 20, no. 1, pp. 105–110, Jan. 2013.

[138] J. E. Voss *et al.*, "Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography," *Nature*, vol. 468, no. 7324, pp. 709–712, Dec. 2010.

[139] E. R. Mackow and I. N. Gavrilovskaya, "Cellular receptors and hantavirus pathogenesis.," *Curr. Top. Microbiol. Immunol.*, vol. 256, pp. 91–115, 2001.

[140] Y. Choi, Y.-C. Kwon, S.-I. Kim, J.-M. Park, K.-H. Lee, and B.-Y. Ahn, "A hantavirus causing hemorrhagic fever with renal syndrome requires gC1qR/p32 for efficient cell binding and infection.," *Virology*, vol. 381, no. 2, pp. 178–83, Nov. 2008.

[141] V. S. Matthys, E. E. Gorbunova, I. N. Gavrilovskaya, and E. R. Mackow, "Andes Virus Recognition of Human and Syrian Hamster 3 Integrins Is Determined by an L33P Substitution in the PSI Domain," *J. Virol.*, vol. 84, no. 1, pp. 352–360, Jan. 2010.

[142] E. Krautkrämer and M. Zeier, "Hantavirus causing hemorrhagic fever with renal syndrome enters from the apical surface and requires decay-accelerating factor (DAF/CD55).," *J. Virol.*, vol. 82, no. 9, pp. 4257–64, May 2008.

[143] A. Waterhouse *et al.*, "SWISS-MODEL: homology modelling of protein structures and complexes," *Nucleic Acids Res.*, vol. 46, no. W1, pp. W296–W303, Jul. 2018.

[144] H. Ashkenazy *et al.*, "ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules.," *Nucleic Acids Res.*, vol. 44, no. W1, pp. W344-50, 2016.

[145] F. Sievers *et al.*, “Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega.,” *Mol. Syst. Biol.*, vol. 7, p. 539, Oct. 2011.

[146] A. Kotecha *et al.*, “Rules of engagement between αvβ6 integrin and foot-and-mouth disease virus,” *Nat. Commun.*, vol. 8, no. 1, p. 15408, Aug. 2017.

[147] Z. Elgundi, M. Reslan, E. Cruz, V. Sifniotis, and V. Kayser, “The state-of-play and future of antibody therapeutics,” *Adv. Drug Deliv. Rev.*, vol. 122, pp. 2–19, Dec. 2017.

[148] C. de Carvalho Nicacio, A. Lundkvist, K. B. Sjölander, A. Plyusnin, E. M. Salonen, and E. Björling, “A neutralizing recombinant human antibody Fab fragment against Puumala hantavirus.,” *J. Med. Virol.*, vol. 60, no. 4, pp. 446–54, Apr. 2000.

[149] J. Arikawa, A. L. Schmaljohn, J. M. Dalrymple, and C. S. Schmaljohn, “Characterization of Hantaan Virus Envelope Glycoprotein Antigenic Determinants Defined by Monoclonal Antibodies,” *J. Gen. Virol.*, vol. 70, no. 3, pp. 615–624, Mar. 1989.

[150] J. Hörling and A. Lundkvist, “Single amino acid substitutions in Puumala virus envelope glycoproteins G1 and G2 eliminate important neutralization epitopes.,” *Virus Res.*, vol. 48, no. 1, pp. 89–100, Apr. 1997.

[151] M. Wang, D. G. Pennock, K. W. Spik, and C. S. Schmaljohn, “Epitope Mapping Studies with Neutralizing and Non-neutralizing Monoclonal Antibodies to the G1 and G2 Envelope Glycoproteins of Hantaan Virus,” *Virology*, vol. 197, no. 2, pp. 757–766, Dec. 1993.

[152] T. Heiskanen, Å. Lundkvist, R. Soliymani, E. Koivunen, A. Vaheri, and H. Lankinen, “Phage-Displayed Peptides Mimicking the Discontinuous Neutralization Sites of Puumala Hantavirus Envelope Glycoproteins,” *Virology*, vol. 262, no. 2, pp. 321–332, Sep. 1999.

[153] D. N. Mastronarde, “Automated electron microscope tomography using robust prediction of specimen movements.,” *J. Struct. Biol.*, vol. 152, no. 1, pp. 36–51, Oct. 2005.

8. Appendix

8.1. Sequences used

8.1.1. PUUV M segment construct 1

CGAATTGAAGGAAGGCCGTCAAGGCCGCATGAGCTCACCGGTATGGGCAAGATCATCC
CCGTGTGCCTGTACCTGCTGTTCCTGTGCCTGCTGCTGTGGAACCGGGGCGACGCCA
GAAACCTGAACGAGCTGAAGATCGAGTGCCCCCACACAATCGGCCTGGGCCAGGGCC
TGGTGATCGGCAGCGTGGAACTCCCTCCTGTGCCTCTGACCCAGGTGGAGAGCCTGA
AGCTGGAAAGCAGCTGCAACTTCGACCTGCACACCAGCACCAGCTCCCAGCAGCCCTT
CACCAAGTGGACCTGGGAGATGAAGTCCGACCTGGCCGAGAATACCCAGGCCAGCAG
CACAAGCTTCCAGACCAAGAGCAGCGAGATCAACCTGCGGGGCCTGTGTCTGGTGCC
CCCTCTGGTGATCGAGACAGCCGCCAGAACCAGAAAGACAATCGCCTGCTTCGACCTG
AGCTGCAACCAGACCGCCTGCCAGCCCACCGTGTTTCTGATCGGCCCCATCCAGACCT
GCATCACCACCAAGAGCTGCCTCCTGGGCCTGGGCGACCAGCGGATCCAGGTGAACT
ACGAGAAAACCTACTGCGTGAGCGGCCAGCTGGTGGAGGGCGTGTGCTTCAACCCCG
TGCATACCATGGCCCTGAGCCAGCCCAGCCACACCTACGACATCGTGACCGTGATGGT
GCGGTGCTTCCTGATCGCCAAGAAGGTGTCCACCGGCGACAGCATGAAGCTGGAAAA
GTCCTTCGAGACACTGGTGCAGAAAACCAGCTGCACCGGCAACGGCTTCCAGGGCTAC
TACATCTGCCTGGTGGGCAGCAGCAGCGAGCCCCTGTATATCCCCACCCTGGACGACT
ACAGATCCGCCGAGGTGCTGAGCAGAATGGCCTTCGCCCCCCACGGCGAGGATCACG
ACGTGGAGAAGAACGCCATCAGCGCCATGCGGATCATCGGCAAAGTGACCGGCAAGG
CCCCTAGCACCGAGAGCAGCGACACAATCCAGGGCGTGGCCTTCAGCGGCAACCCCC
TGTACACAAGCACCGGCGTGCTGACCGCCAAGGACGACCCCGTGTACATCTGGGCCC
CTGGCATCATCATGGAAGGCAACCACAGCGTGTGCGACAAGAAAACCCTGCCCCTGAC
CTGGACCGGCTTCATCCCCCTGCCCGGCGAGATCGAGAAAACCACCCAGTGCACCGT
GTTCTGCACACTGGCCGGACCTGGCGCCGATTGCGAGGCCTACAGCGAGACAGGCAT
CTTCAACATCAGCAGCCCCACCTGTCTGATCAACCGGGTGCAGCGGTTTAGAGGCGCC
GAGCAGCAGATTAAGTTCATCTGCCAGCGGGTGGACATGGACATCACCGTGTACTGCA
ACGGCCTGAAGAAAGTGATCCTGACCAAGACCCTGATCATCGGCCAGTGCATCTACAC
CTTCACCAGCATCTTCAGCCTGATCCCCGGCGTGGCCCATTCTCTGGCCGTGGAGCTG
TGTGTGCCTGGACTGCACGGCTGGGCCACCATGCTGCTGCTGCTGACCTTCTGCTTCG
GCTGGATCCTGATCCCCTTCTTCTCCATGATCCTGCTGAAGATCCTGATTGCCTTCGCC
TACCTGTGCAGCAAGTACAACACCGACAGCAAGTTCAAGATTCTGATCGAGAAAGTGAA
GAAAGAATACCAGAAAACCATGGGCAGCATGGTGTGCGAAGTGTGCCACTACGAGTGC
GAGACAGCCAAAGAGCTGGAATCCCACCGGAAGTCCTGCCCTATGGGAAGCTGCCCC
TACTGCCTGAATCCTGCCGAGGCCACCACAAGCGCCCTGCAGGCCCACTTCAAAGTGT
GCAAGCTGACCAGCCGGTTCCAGGACAACCTGCGGAAGTCCCTGGGCGCCTACGAGC
CCAAGCAGGGCTGCTACCGGACCCTGAGCCTGTTCCGGTACAGATCCCGGTTCTTCGT
GGGCTTCATCTGGTGTCTGCTGCTGGTGCTGGAACTGATCGTGTGGGCCGCCAGCGC
CGAGACACAGAACCTGAACAGCGGCTGGACCGATACAGCCCACGGCAGCGGCATCAT
CCCCATGCGGACCGACCTGGAACTGGACTTCAGCCTGCCCAGCAGCGCCTCCTACAC
CTACCGGCGGCAGCTGCAGAACCCCGCCAACGAGCAGGAAAAGATCCCATTCCATCTG
CAGATCAGCAAACAGGTGATCCACGCCGAGATCCAGCACCTGGGCCACTGGATGGAC
GGCACCTTCAACCTGAAAACCGCCTTCCACTGCTACGGCAGCTGCGAAAAGTACGCCT
ACCCCTGGCAGACCGCCGGCTGCTTCATCGAGAAGGACTACGAGTACGAGTCCGGCT
GGGGCTGCAATCCCCCCGATTGTCCTGGCGTGGGCACCGGATGTACAGCCTGCGGCG
TGTACCTGGACAAGCTGAAGTCCGTGGGCAAGGCCTTCAAGATCGTGTCCCTGCGGTA
CACCCGGAAGGCCTGCATCCAGCTGGGCACCGAGCAGACCTGCAAGAGCGTGGACAG

CAACGACTGCCTGGTGACCACCAGCGTGAAAGTGTGCCTGATCGGCACCGTGTCCAAG
TTCCAGCCCTCCGACACCCTGCTGTTTCTGGGCCCTCTGGAACAGGGCGGCCTGATCT
TCAAGCAGTGGTGCACAACCACCTGTCAGTTCGGCGACCCCGGCGATATCATGAGCAC
CCCCGTGGGCATGAAGTGCCCCGAGCTGAACGGCAGCTTCCGGAAGAAGTGCGCCTT
CGCCACCACCCCCGTGTGTCAGTTCGACGGCAACACCCTGAGCGGCTACAAGCGGAT
GATCGCCACCAAGGACAGCTTCCAGAGCTTCAACGTGACCGAGCCCCACATCAGCGCC
TCCAGCCTGGAATGGATCGACCCCGACAGCAGCCTGCGGGACCACATCAACGTGATC
GTGGGCCGGGACCTGAGCTTCCAGGACCTGAGCGAGACACCCTGTCAGGTGGACCTG
GCCACCACCTCTATCGATGGCGCCTGGGGAAGCGGCGTGGGCTTTAACCTGGTGTGC
AGCGTGTCCCTGACCGAGTGCAGCACCTTCCTGACCAGCATCAAGGCCTGCGACAGC
GCCATGTGCTACGGCTCCACCACCGCCAATCTGCTGCGGGGCCAGAACACAGTGCAC
ATCGTGGGCAAAGGCGGCCACAGCGGCTCCAAGTTCATGTGCTGCCACGACACCAAG
TGCAGCAGCACAGGACTGGTGGCCGCTGCCCCCCACCTGGATAGAGTGACCGGCTAC
AACCAGGCCGACAGCGACAAGATCTTCGACGATGGCGCCCCTGAGTGCGGCATCAGC
TGCTGGTTCACCAAGTCCGGCGAGTGGATCCTGGGCGTGCTGAATGGCAACTGGATG
GTGGTGGCCGTGCTGGTGGCCCTGTTCATCCTGAGCATCTTTCTGTTCACCCTGTGCT
GCCCCAGAAGGCCCAGCTACGCCAAAGAGCACAAGCCCGGTACCCTGGGCCTCATGG
GCCTTCCTTTCACTGCC

8.1.2. PUUV M segment construct 2

ATGGGAAAGTCCAGCCCCGTGTGCCTGTACCTGATCCTGCAGGGCCTGCTGCTGTTCG
ACACCGTGAACGCCCGGAACCTGAACGAGCTGAAGATGGAATGCCCTCACACCATCGG
CCTGGGCCAGGGACTGGTCGTGGGATCTGTGGAACTGCCTCCCGTGCCCATCCAGCA
GATCGAGAGCCTGAAGCTGGAAAGCAGCTGCAACTTCGACCTGCACACCAGCACAGCC
GGCCAGCAGAGCTTCACCAAGTGGACCTGGGAGACAAAGGGCGACCTGGCCGAGAAT
ACCCAGGCCAGCAGCACCAGCTTCCAGACCAAGAGCAGCGAAGTGAACCTGCGGGGC
CTGTGTCTGATCCCCACCCTGGTGGTGGAAACCGCCGCCAGAATGAGAAAGACAATCG
CCTGCTACGACCTGAGCTGCAACCAGACCGTGTGCCAGCCCACCGTGTATCTGATGGG
CCCCATCCAGACCTGCCTGACCACCAAGTCTTGCCTGCTGGGACTGGGCGACCAGCG
GATCCAGGTCAACTACGAGAGAACCTACTGCGTGTCCGGCCAGCTGGTGGAAGGCGT
GTGCTTCAACCCCATCCACACCATGGCCCTGTCCCAGCCCAGCCACACCTACGACATC
GTGACAATCATGGTCCGATGCTTTCTGGTCATCAAGAAAGTGACCAGCGGCGACAGCA
TGAAGATCGAGAAGAACTTCGAGACACTGGTGCAGAAAACCGGCTGCACCGCCAACGG
CTTCCAGGGCTACTACATCTGCCTGATCGGCAGCAGCAGCGAGCCCCTGTACGTGCCA
ACCCTGGACGACTACAGATCCGCCGAGGTGCTGAGCAGAATGGCCTTTGCCCCTCACG
GCGAGGACCACGACATTGAGAAGAACGCCGTGTCCGCCCTGCGGATCGCCGGCAAAG
TGACAGGCAAGGCCCCTAGCACCGAGTCCAGCGATACAGTGCAGGGAATCGCCTTCA
GCGGCAGCCCTCTGTACACAAGCACAGGCGTGCTGACCGCCAAGGACGACCCTGTGT
ATGTGTGGGCTCCCGGCATCATCATGGAAGGCAACCACAGCGTGTGCGAGAAGAAAAC
CCTGCCCCTGACCTGGACCGGCTTCATCCCTCTGCCCGGCGAGATCGAAAAGACCACC
CAGTGCACCGTGTTCTGCACCCTGGCTGGACCTGGCGCCGATTGCGAGGCCTACAGC
GAGACAGGCATCTTCAACATCAGCAGCCCCACCTGTCTGATCAACCGGGTGCAGAGAT
TCAGAGGCGCCGAGCAGCAGATCAAGTTCGTGTGTCAGCGGGTGGACATGGACATCA
CCGTGTACTGTAACGGCGTGAAGAAAGTCATCCTGACCAAGACCCTGGTCATCGGCCA
GTGCATCTACACCTTCACCAGCATCTTCTCCCTGATCCCCGGAATCGCCCACAGCCTG
GCCGTGGAACTGTGTGTGCCTGGACTGCACGGATGGGCTACCGTGCTGCTGCTCCTG
ACCTTCTGCTTCGGCTGGGTGCTGATCCCTACCATCACCATGATCCTGCTGAAGATCCT
GATCGCCTTCGCCTACCTGTGCAGCAAGTACAACACCGACAGCAAGTTCAAGATCCTG

GTCGAGAAAGTCAAGAAAGAATACCAGAAAACCATGGGCAGCATGGTCTGCGAAGTGT
GCCAGTACGAGTGCGAGACAGCCAAAGAGCTGGAATCCCACAGAAAGAGCTGCAGCA
TCGGCAGCTGTCCCTACTGCCTGAACCCTAGCGAAGCCACACCAAGCGCTCTGCAGGC
CCACTTCAAAGTGTGCAAGCTGACCAGCCGGTTCCAGGAAAACCTGAAGAAAAGCCTG
ACTATGTACGAGCCCATGCAGGGCTGCTACCGGACCCTGAGCCTGTTCCGGTACAGAT
CCCGGTTCTTCGTGGGCCTCGTGTGGTGCATGCTGCTGGTGCTCGAACTGATCGTGTG
GGCCGCCAGCGCCGAGACACAGAATCTGAATGCCGGCTGGACCGACACAGCCCACGG
CTCTGGCATCATCCCCATGAAGGCCGACCTGGAACTGGACTTCAGCCTGCCCAGCAGC
GCCTCCTACACCTACAGAAGGCAGCTGCAGAACCCCGCCAACGAGCAGGAAAAGATCC
CATTCCATCTGCAAATCAGCAAGCAGGTCATCCACGCCGAGATCCAGCACCTGGGCCA
CTGGATGGACGCCACCTTCAACCTGAAAACCGCCTTTCACTGCTACGGCTCTTGCGAG
AAGTACGCCTACCCTTGGCAGACCGCCGGCTGCTTCGTGGAAAAGGACTACGAGTACG
AAACCGGATGGGGCTGCAACCCTCCCGATTGTCCTGGCGTGGGCACCGGATGTACAG
CCTGCGGCGTGTACCTGGACAAGCTGAAGTCTGTGGGCAAGGTGTTCAAGATTGTGTC
CCTGCGGTACACCCGGAAAGTCTGCATCCAGCTGGGCACAGAGCAGACATGCAAGAC
CGTGGACAGCAACGATTGCCTGATCACCACCAGCGTGAAAGTGTGCCTGATTGGCACC
ATCTCCAAGTTCCAGCCCTCCGATACCCTGCTGTTCCTGGGACCTCTGCAGCAGGGCG
GCCTGATCTTCAAGCAGTGGTGTACCACCACCTGTCAGTTCGGCGACCCCGGCGATAT
CATGAGCACCCCTACCGGCATGAAGTGCCCCGAGCTGAACGGCAGCTTCAGAAAGAA
GTGCGCCTTTGCCACCACCCCTGTGTGTCAGTTTGACGGCAACACCATCAGCGGCTAC
AAGCGGATGGTGGCCACAAAGGACAGCTTTCAGTCCTTCAACGTGACCGAGCCCCACA
TCTCCACCAGCGCCCTGGAATGGATCGACCCCGACAGCAGCCTGAGGGACCACATCA
ACGTGATCGTGTCCAGGGATCTGAGCTTCCAGGACCTGAGCGAGACACCCTGCCAGGT
GGACCTGACCACAAGCGCCATTGATGGCGCCTGGGGCAGCGGCGTGGGCTTTAATCT
CGTGTGTACCGTGTCCCTGACCGAGTGCAGCGCCTTCCTGACCAGCATCAAAGCCTGC
GACGCCGCCATGTGCTACGGCAGCACCACCACAAATCTCGTGCGGGGCCAGAACACC
ATCCACGTCGTGGGAAAAGGCGGCCACAGCGGCTCCAAGTTTATGTGCTGCCACGACA
CCAAGTGCAGCTCCACAGGACTGGTGGCTGCCGCCCCACACCTGGACAGAGTGACCG
GATTCAACCAGGCCGACAGCGACAAGATCTTCGACGATGGCGCCCCTGAGTGCGGCA
TGAGCTGCTGGTTCAAGAAACTGGGCGAGTGGGTGCTGGGAGTGCTGAATGGCAACT
GGATGGTGGTGGCCGTGCTGATCGCTCTGCTGATCCTGAGCATCTTCCTGTTCGCCCT
GTGCTGCCCTCGGAGGCCCAGCTACAAGAAGGAGCACAAGCCCTGA

8.1.3. HNTV M segment construct

GAATTCGCCACCATGGGCATGTGGAAGTGGCTCGTGATTGCCAGCCTTGTGTGGCCTG
CTCTGAGCCTGAGAAACGTGTACGACATGAAGATCGAGTGCCCTCACACCGTGTCCTT
CGGCGAGAATAGCGTGATCGGCTACGTGGAACTGCCTCCTATGCCTCTGGCCGACACA
GCTCAGATGGTGCCTGAGAGCAGCTGCAGCATGGACAACCACCAGAGCATCAACACCA
TCACCAAGTACACCCAAGTGATCTGGCGCGGCAAGGCCGATCCTGGCCAAAGCAGCC
AGAACAGCTTCGAGACAGTGTCCACCGAGGTGGACCTGAAGGGCACCTGTGTGCTGAA
GCACAAGATGGTGGAAGAGAGCTACAGAAGCCGGAAGTCCATCACCTGTTACGACCTG
AGCTGCAACAGCACCTTCTGCAAGCCCACACTGTACATGATCGTGCCCATCCACGCCT
GCAACATGATGAAGTCCTGCCTGATCGCTCTGGGCCCCTACAGAGTGCAGGTCGTGTA
CGAGAGAACCTACTGCATGACCGGCGTGCTGATCGAGGGCAAGTGCTTCGTGCCTGAT
CAGAGCGTGGTGTCCATCATCAAGCACGGCATCTTCGATATCGCCAGCGTGCACGTCG
TGTGCTTTTTCGTGGCCGTGAAGGGCAACACCTACAAGCTGTTCGAGCAAGTGAAGAA
GTCCTTCGAGAGCACCTGTAACGACACCGAGAACAAGGTGCAGGGCTACTACATCTGC
ATCGTCGGCGGCAATAGCGCCCCTATCTACGTGCCAACACTGGACGACTTCCGGTCCA

TGGAAGCCTTCACCGGCATCTTTAAGAGCCCTCACGGCGAGGATCACGACCTGGCCG
GCGAAGAGATCGCCAGCTATTCTATTGTGGGCCCTGCCAACGCCAAGGTGCCACATTC
TGCCAGCAGCGATACCCTGTCTCTGATCGCCTACAGCGGCATCCCTAGCTACAGCAGC
CTGAGCATCCTGACCAGCAGCACCGATGCCAAGCACGTGTTCAGCCCTGGACTGTTCC
CCAAGCTGAACCACACCAACTGCGACAAGAGCGCCATTCCTCTGACCTGGACCGGCAT
GATTGACCTGCCTGGCTACTACGAGGCCATCCATCCTTGCACCGTGTTCTGCGTGTTGT
CTGGACCTGGCGCCTCTTGCGAGGCCTTTTCTGAAGGCGGAATCTTCAACATCACAAG
CCCCATGTGCCTGGTGTCCAAGCAGAACAGATTCCGGCTGACAGAGCAGCAAGTCAAC
TTCGTGTGCCAGCGCGTGGACATGGACATCGTGGTGTACTGCAACGGCCAGCGGAAA
GTGATCCTGACAAAGACCCTGGTCATCGGCCAGTGTATCTACACAATCACCAGCCTGTT
CAGCCTGCTGCCTGGCGTGGCCCATTCTATCGCTGTGGAACTGTGTGTGCCTGGCTTT
CACGGATGGGCTACAGCCGCTCTGCTCGTGACCTTCTGTTTTGGCTGGGTGCTGATCC
CAGCCGTGACCTTTATCATCCTGACAGTGCTGAAGTTTATCGCCAACATCTTCCACACC
AGCAATCAAGAGAACCGGCTGAAGTCCGTGCTGCGGAAGATCAAAGAAGAGTTCGAAA
AGACCAAGGGCAGCATGGTCTGCGACGTGTGCAAATACGAGTGCGAGACATACAAAGA
GCTGAAGGCCCACGGCGTGTCCTGTCCTCAGAGCCAGTGTCCTTACTGCTTCACCCAC
TGCGAGCCTACCGAAGCCGCCTTTCAGGCCCACTACAAAGTGTGCCAAGTGACCCACA
GGTTCCGGGACGACCTGAAGAAAACCGTGACACCCCAGAACTTCACCCCTGGCTGCTA
CAGAACCCTGAACCTGTTCCGGTACAAGAGCCGGTGCTACATCTTTACTATGTGGGTGT
TCCTGCTGATCCTGGAAAGCGTGCTGTGGGCCGCCTCTGCCTCTGAAACACCTCTGAC
ACCCGTGTGGAACGACAACGCTCATGGCGTGGGAAGCGTGCCCATGCACACAGACCT
GGAACTGGATTTCAGCCTGACCTCCAGCTCTAAGTACACCTACCGGCGGAAGCTGACC
AATCCTCTGGAAGAGGCCCAGTCCATCGATCTGCACATCGAGATCGAGGAACAGACCA
TCGGCGTGGACGTTCACGCTCTCGGCCATTGGTTCGACGGCCGGCTGAATCTGAAAAC
CAGCTTCCACTGCTACGGCGCCTGCACTAAGTACGAGTACCCTTGGCATACCGCCAAG
TGTCACTACGAGCGGGACTACCAGTACGAGACAAGCTGGGGCTGCAACCCTAGCGATT
GTCCTGGCGTCGGCACAGGATGTACAGCCTGCGGACTGTACCTGGACCAGCTGAAGC
CTGTGGGCAGCGCCTACAAGATCATCACCATCCGGTACTCTCGGAGAGTGTGCGTGCA
GTTCGGAGAAGAGAACCTGTGCAAGATCATTGACATGAACGACTGCTTCGTGTCCCGG
CACGTGAAAGTGTGCATCATCGGCACCGTGTCTAAGTTCAGCCAGGGCGACACCCTGC
TGTTCTTCGGACCATTGGAAGGCGGCGGACTGATCTTCAAGCACTGGTGCACAAGCAC
CTGTCAGTTCGGCGATCCCGGCGACATCATGAGCCCTAGAGACAAGGGCTTCCTGTGT
CCTGAGTTCCCCGGCAGCTTCAGAAAGAAGTGCAACTTCGCCACAACTCCCATCTGCG
AGTACGACGGCAATATGGTGTCCGGCTACAAGAAAGTGATGGCCACCATCGACAGCTT
CCAGTCCTTCAACACCTCCACCATGCACTTCACCGACGAGCGGATCGAGTGGAAGGAC
CCTGATGGCATGCTGCGGGACCACATCAACATCCTGGTCACCAAGGACATCGACTTCG
ACAACCTGGGCGAGAATCCCTGCAAGATCGGCCTGCAGACCAGCTCTATTGAAGGCGC
TTGGGGTAGCGGCGTGGGCTTTACACTGACCTGTCTGGTGTCCCTGACCGAGTGTCCT
ACCTTCCTGACCTCTATCAAGGCCTGCGATAAGGCCATCTGTTACGGCGCCGAGTCTG
TGACCCTGACCAGAGGACAGAACACCGTGAAGGTGTCCGGAAAAGGCGGCCACTCTG
GCTCCACCTTCAAGTGCTGTCACGGCGAAGATTGCTCCAAGATTGGACTGCATGCCGC
CGCTCCTCACCTGGACAAAGTGAATGGCATCTCTGAGATCGAGAACTCCAAGGTGTAC
GACGATGGCGCCCCTCAGTGCGGCATCAAGTGTTGGTTCGTGAAGTCCGGCGAGTGG
ATCTCTGGCATCTTCAGCGGCAACTGGATCGTGCTGATTGTGCTGTGTGTGTTTCTGCT
GTTTAGCCTGGTCCTGCTGAGCATTCTGTGCCCCGTGCGGAAACACAAGAAGTCTGGT
ACCCTGGAGGTGCTGTTCCAGGGCCCCGGAATGGTGTCTAAGGGCGAAGAACTGTTTA
CCGGCGTGGTGCCAATCCTGGTTGAACTGGATGGGGACGTGAACGGACACAAGTTCTC
TGTGTCTGGCGAAGGCGAGGGCGACGCCACATATGGAAAGCTGACCCTGAAGTTCATC

TGCACCACCGGCAAACTGCCCGTGCCTTGGCCTACACTGGTCACAACACTGTCTTGGG
GCGTGCAGTGCTTCGCTAGATACCCCGACCACATGAAGCAGCACGATTTCTTCAAGTC
CGCCATGCCTGAGGGCTACGTGCAAGAGCGGACCATCTTCTTCAAGGACGATGGCAAC
TACAAGACCAGGGCCGAAGTGAAGTTCGAGGGCGATACCCTCGTGAACCGGATCGAG
CTGAAAGGCATCGATTTCAAAGAGGACGGCAACATCCTCGGGCACAAACTTGAGTACA
ACTACTTCAGCGACAACGTGTACATCACCGCCGACAAGCAGAAGAACGGAATCAAGGC
CAACTTCAAGATCCGGCACAACATCGAGGATGGCGGAGTGCAGCTGGCCGACCACTAT
CAGCAGAATACCCCTATCGGAGATGGCCCCGTGCTGCTGCCCGATAATCACTACCTGA
GCACCCAGAGCAAGCTGAGCAAGGACCCCAACGAGAAGCGGGATCACATGGTCCTGC
TGGAATTTGTGACAGCCGCCGGAATCACCCTCGGCATGGACGAGCTTTACAAACATCA
CCATCACCATCATTAATGACTCGAG

8.1.4. ANDV M segment construct

GAATTCGCCACCATGGAAGGCTGGTATCTGGTGGCCCTGGGCATCTGCTACACCCTGA
CACTGGCCATGCCTAAGACCACCTACGAGCTGAAGATGGAATGCCCTCACACAGTCGG
ACTCGGCCAGGGCTATATCATCGGCTCTACAGAGCTGGGCCTGATCTCTATCGAGGCC
GCCAGCGACATCAAGCTGGAAAGCAGCTGCAACTTCGACCTGCACACCACCAGCATGG
CCCAGAAATCCTTCACACAGGTGGAATGGCGGAAGAAGTCCGACACCACCGATACCAC
AAATGCCGCCAGCACCACCTTTGAGGCCCAGACCAAGACCGTGAACCTGCGGGGCAC
ATGTATTCTGGCCCCTGAGCTGTACGACACCCTGAAGAAAGTGAAGAAAACCGTGCTG
TGCTACGACCTGACCTGCAATCAGACCCACTGCCAGCCTACCGTGTACCTGATTGCCC
CTGTGCTGACCTGTATGAGCATCCGGTCCTGTATGGCCCGGGTGTTCACCAGCAGAAT
CCAAGTGATCTACGAGAAAACCCACTGCGTGACCGGCCAGCTGATCGAGGGCCAGTG
CTTTAACCCTGCTCACACACTGACCCTGAGCCAGCCAGCTCACACCTACGATACCGTG
ACTCTGCCTATCAGCTGCTTCTTCACCCCTAAAGAGTCCGAGCAGCTGAAAGTGATCAA
GACCTTCGAGGGCATCCTGACCAAGACAGGCTGCACCGAAAATGCTCTGCAGGGCTAC
TACGTGTGCTTCCTGGGCTCTCACAGCGAGCCTCTGATCGTGCCTAGCCTGGAAGATA
TCAGAAGCGCCGAGGTCGTCAGCCGGATGCTGGTTCATCCTAGAGGCGAGGACCACG
ACGCCATCCAGAATAGCCAGAGCCACCTGAGAATCGTGGGCCCCATTACAGCCAAGGT
GCCAAGCACCAGCAGCACCGATACACTGAAGGGCACAGCCTTTGCCGGCGTGCCAAT
GTACAGCAGCCTGAGCACCCTGGTCAAGAACGCCGATCCTGAGTTCGTGTTCAGCCCC
GGCATCATCCCCGAGAGCAATCACAGCGTGTGCGACAAAAAGACCGTGCCTATCACCT
GGACCGGCTACCTGCCTATCTCCGGCGAGATGGAAAAAGTGACCGGCTGTACCGTGTT
CTGCACCCTTGCTGGACCAGGCGCTAGCTGTGAAGCCTACAGCGAGAACGGCATCTTC
AACATCAGCAGCCCCACCTGTCTGGTCAACAAGGTGCAGAGATTCCGGGGCAGCGAG
CAGAAGATCAACTTCATCTGCCAGCGCGTGGACCAGGACGTGGTGGTGTACTGCAACG
GCCAGAAAAAAGTGATCCTCACAAAGACCCTCGTGATCGGCCAGTGCATCTACACCTTC
ACCAGCCTGTTCAGCCTGATGCCTGATGTGGCCCATAGCCTGGCCGTGGAACTGTGTG
TTCCTGGACTGCACGGATGGGCCACTGTGATGCTGCTGAGCACCTTCTGCTTTGGCTG
GGTGCTGATCCCTGCCGTGACACTGATCATCCTGAAGTGCCTGCGGGTGCTGACATTC
AGCTGCAGCCACTACACCAACGAGAGCAAGTTCAAGTTCATCCTCGAAAAAGTCAAGGT
CGAGTACCAGAAAACCATGGGCAGCATGGTCTGCGACGTGTGCCACCACGAATGCGA
GACAGCCAAAGAGCTGGAATCCCACCGGCAGAGCTGCATCAATGGACAGTGCCCCTAC
TGCATGACCATCACCGAGGCCACAGAAAGCGCCCTGCAGGCCCACTACTCTATCTGCA
AGCTGACCGGCAGATTCCAAGAGGCCCTGAAAAAGTCTCTGAAGAAGCCCGAAGTCAA
GAAGGGCTGCTACAGAACCCTGGGCGTGTTCCGGTACAAGAGCAGATGTTACGTGGG
CCTCGTGTGGTGCCTGCTGCTGACATGCGAGATCGTGATTTGGGCCGCCTCTGCCGAG
ACACCCCTTATGGAATCTGGCTGGAGCGATACAGCCCACGGCGTGGGAGAGATCCCC

ATGAAGACAGACCTGGAACTGGACTTCAGCCTGCCTAGCAGCAGCAGCTACTCCTACA
GACGGAAGCTGACAAACCCCGCCAACAAAGAAGAGAGCATCAGCTTCCACTTCCAAAT
GGAAAAGCAAGTGATCCACGCCGAGATCCAGCCTCTCGGCCATTGGATGGACGCCAC
CTTCAACACCAAAACCGCCTTCCACTGCTACGGCGCCTGCCAGAAGTACAGCTACCCT
TGGCAGACCAGCAAGTGCTTCTTCGAGAAGGACTACCAGTACGAAACCGGCTGGGGCT
GCAACCCTGGCGATTGTCCTGGTGTTGGAACCGGCTGCACAGCCTGCGGAGTGTACCT
GGACAAGCTGAAGTCTGTGGGCAAAGCCTACAAGATCATCAGCCTGAAGTACACCCGG
AAAGTGTGCATCCAGCTGGGCACAGAGCAGACCTGCAAGCACATCGACGCCAACGACT
GTCTCGTGACCCCTAGCGTGAAAGTCTGCATCGTGGGCACCGTGTCCAAACTGCAGCC
TAGCGATACCCTGCTGTTCCTGGGACCTCTGGAACAAGGCGGCGTGATCCTGAAGCAG
TGGTGTACCACCAGCTGCGCCTTTGGCGATCCCGGCGATATCATGTCTACCCCAAGCG
GCATGAGATGCCCTGAGCACACCGGCAGCTTCAGAAAGATCTGCGGCTTCGCCACAAC
ACCCGTGTGCGAGTATCAGGGCAATACCATCAGCGGCTACAAGCGGATGATGGCCACC
AAGGACAGCTTCCAGAGCTTCAACCTGACCGAGCCTCACATCACAGCCAACAAACTGG
AATGGATCGACCCCGACGGCAATACCCGGGACCATGTGAACCTGGTGCTGAACCGGG
ACGTGTCATTTCAGGACCTGAGCGACAACCCCTGCAAGGTGGACCTGCATACCCAGGC
CATTGAAGGCGCTTGGGGAAGCGGAGTGGGCTTTACCCTGACATGTACCGTGGGCCT
GACAGAGTGCCCCAGCTTCATGACCTCCATCAAGGCCTGCGATCTGGCCATGTGTTAC
GGCAGCACTGTGGCCAATCTGGCCAGGGGCTCTAACACCGTGAAGGTCGTCGGAAAA
GGCGGCCACAGCGGCAGCAGCTTCAAGTGCTGTCACGATACCGACTGCAGCAGCGAG
GGACTGCTTGCCTCTGCTCCTCACCTGGAACGCGTGACAGGCTTCAACCAGATCGACA
GCGACAAGGTGTACGACGATGGCGCCCCTCCATGCACCTTCAAGTGTTGGTTTACCAA
GAGCGGCGAGTGGCTGCTGGGAATCCTGAACGGCAATTGGATTGTGGTGGTGGTCCT
GGTGGTCATCCTGATCCTGAGCATCATCATGTTCAGCGTGCTGTGTCCCAGACGGGGC
CACAAGAAAACAGTGGGTACC

8.1.5. Protocadherin-1 construct

>hPCDH-1_EC1EC2

CGCACCGGTACAAGAGTGGTGTACAAGGTGCCCGAGGAACAGCCTCCTAACACACTGA
TCGGAAGCCTGGCCGCCGATTACGGCTTTCCTGATGTGGGCCACCTGTACAAGCTGGA
AGTGGGAGCCCCTTACCTGAGAGTGGATGGCAAGACCGGCGACATCTTCACCACCGAA
ACCAGCATCGACCGCGAGGGCCTGAGAGAGTGTCAGAATCAACTGCCCGGCGATCCC
TGCATCCTGGAATTTGAGGTGTCCATCACCGACCTGGTGCAGAACGGCAGCCCTAGAC
TGCTGGAAGGACAGATCGAGGTGCAGGACATCAACGACAACACCCCTAACTTCGCTAG
CCCCGTGATCACCCTGGCTATCCCCGAGAACACCAACATCGGCAGCCTGTTTCCAATT
CCTCTGGCCAGCGACAGAGATGCTGGCCCTAATGGCGTGGCCAGCTATGAACTGCAG
GCCGGACCTGAGGCTCAAGAGCTGTTTGGACTGCAAGTGGCCGAGGACCAAGAGGAA
AAGCAGCCCCAGCTGATCGTGATGGGCAACCTGGACAGAGAGAGATGGGACAGCTAC
GACCTGACCATCAAAGTGCAGGATGGCGGCTCTCCACCTAGAGCCAGTTCTGCTCTGC
TGAGAGTGACCGTGCTGGACACCAACGATAATGCCCCAAAGTTCGAGCGGCCCAGCG
GTACCCGC

8.2. HNTV N-terminal histidine primer design

ID	Name	Direction	Template used	Translated sequence	DNA Sequence
A	HNTV NHis HexaHis 3link	Forward	original HNTV	RTGHHHHHHGSGLRNV YDMKIE	CGCACCGGTCATCACCATCACCA TCATGGATCCGGACTGAGAAACG TGTACGACATGAAGATCGAG

B	HNTV NHis HexaHis 5link	Forward	original HNTV	RTGHHHHHHGSGSGLR NVYDMKIE	CGCACCGGTCATCACCATCACCA TCATGGATCCGGATCCGGACTGA GAAACGTGTACGACATGAAGATC GAG
C	HNTV NHis DecaHis 3link	Forward	original HNTV	RTGHHHHHHHHHHGSG LRNVYDMKIE	CGCACCGGTCACCACCATCATCA CCATCACCACCACCATGGATCCG GACTGAGAAACGTGTACGACATG AAGATCGAG
D	HNTV NHis DecaHis 5link	Forward	original HNTV	RTGHHHHHHHHHHGSG SGLRNVYDMKIE	CGCACCGGTCACCACCATCATCA CCATCACCACCACCATGGATCCG GATCCGGACTGAGAAACGTGTAC GACATGAAGATCGAG
E	HNTV NHis HDecaH 3link	Forward	N-terminal His HNTV construct A	RTGHHHHHHHHHHHHH HHHGSGLRNVYDMKIE	CGCACCGGTCACCACCATCACCA TCACCACCACCATCATCATCACC ATCACCATCATGGATCC
-	HNTV Mseg rev xho1	Reverse	all	PVRKHKKS*LED	GATCCTCGAGTTACTAAGACTTCT TGTGTTTCCGCACGGG

Color key: AgeI site. Poly-Histidine. Linker peptide. HNTV native sequence. Stop codon. XhoI site.

8.3. PUUV FFU assay data

Table 8.1 Primers to introduce an N terminal His tag to HNTV VLPs

PUUV FFU assay													
Volume per well		0.2	ml										
	10^{-1}			10^{-2}			10^{-3}						
	A	B	Average	A	B	Average	A	B	Average	Dilution (10^{-x})	PFU/ml A	PFU/ml B	PFU/ml Average
0.0017 MOI	TNTC	TNTC	TNTC	6	3	4.5	0	0	0	2	3.0E+03	1.5E+03	2.3E+03
0.0035 MOI	TNTC	TNTC	TNTC	10	5	7.5	0	0	0	2	5.0E+03	2.5E+03	3.8E+03
0.014 MOI	TNTC	TNTC	TNTC	6	14	10	0	1	0.5	2	3.0E+03	7.0E+03	5.0E+03

Table 8.2 PUUV live virus FFU assay data

8.4. Grid preparation details

Sample	VLP Absorbance 280nm	Buffer	Temperature (°C)	Mixed protein	MW of mixed protein (kDa)	On-grid concentration of protein (μM)
HAZV	n/a	PBS	RT	n/a	n/a	n/a
TULV	n/a	T20N100	RT	n/a	n/a	n/a
HNTV VLP	0.47	PBS	4	n/a	n/a	n/a
ANDV VLP + PCDH1	0.52	PBS	21	PCDH1 EC1-EC2	26.1	2.62
ANDV VLP	0.57	PBS	21	n/a	n/a	n/a
PUUV VLP + gC1qR	3.98	PBS	21	gC1qR	25.1	3.17

PUUV VLP + αVβ3 integrin	3.98	PBS +1.6mM $CaCl_2$	21	αVβ3 integrin	297	0.64
PUUV VLP + fab 4G2	0.91	PBS	21	4G2 fab	43.1	2.71
HNTV VLP + fab C4	0.47	PBS	4	C4 fab	42.3	1.07
HNTV VLP + fab C4 (21°C)	0.47	PBS	21	C4 fab	42.c3	1.07

Table 8.3 Cryo-EM sample preparation details
Absorbance is measured using a nanodrop spectrophotometer. T20N100 = Tris 20mM NaCl 100m.

8.5. Data acquisition parameters

Sample	Microscope	kV	Detector	Acquisition Software	Dose per image ($e^-/Å^2$)	Pixel Size (Å)	Defocus range (μm)	Tilt range and increment (degrees)	Number of tilt series or images
HAZV	Tecnai F30 “Polara	300	K2 + GIF	SerialEM	3.5	1.35	2.5-4	-45-45	23
TULV	Tecnai F30 “Polara”	300	K2 + GIF	SerialEM	4.71	2	2.5-4	-30 to 60 (step 3)	49
TULV pH comparison	Tecnai F30 “Polara”	300	K2 + GIF	SerialEM	40	1.35	4	n/a	8+5
TULV handedness	Tecnai F30 “Polara	300	K2 + GIF	SerialEM	5.5	1.35	3-4	-45 to 45 (step 5)	7
HNTV VLP	Titan Krios	300	K2 + GIF	Tomo4	4.5	1.36	2.5-4	-30 to 60 (step 3)	69
ANDV VLP + PCDH1	Titan Krios	300	K2 + GIF	Tomo4	4.272	1.76	2.5-4	-42 to 60 (step 3)	66
ANDV VLP	Titan Krios	300	K2 + GIF	Tomo4	4.77	1.76	2.8-4	-30 to 60 (step 3)	20
PUUV VLP + gC1qR	Titan Krios	300	K2 + GIF	Tomo4	4.77	1.76	2.5-4	-30 to 60 (step 3)	114
PUUV VLP + αVβ3 integrin	Glacios	200	Falcon	Tomo4	4.5	1.955	2.5-4	-45 to 60 (step 3)	9
PUUV VLP + fab 4G2	Titan Krios	300	K2 + GIF	Tomo4	4.5	1.76	2.5-4	-30 to 60 (step 3)	68
HNTV VLP + fab C4	Titan Krios	300	K2 + GIF	Tomo4	4.77	1.76	2.5-4	-42 to 60 (step 3)	43

Table 8.4 Cryo-EM data acquisition parameters
Parameters used for data acquisition. Microscopes by Thermo Fisher/FEI. K2 = K2 Summit; Gatan. GIF = IF Quantum LS; Gatan). Falcon=Falcon (Thermo fisher). SerialEM [153].

8.6. Data processing details

Sample	Initial symmetry imposed	Final bin factor	3D classification (Number of particles before, after)	Final symmetry imposed	Final number of particles	Final resolution (Å)
HAZV	C1	4	n/a	C1	~4000	21.6
TULV	C4	2	23781, 18071 *	C4	18,064	11.4
HNTV VLP	C4	2	4954, 4761 *	C4	4748	12.6
ANDV VLP + PCDH1 (initial)	C4	2	n/a	C4	2823	14.3
ANDV VLP + PCDH1	C4	2	5149, 3023	C1	3023	16.5
ANDV VLP	C4	4	n/a	C4	339	19.8
PUUV VLP + gC1qR	C4	2	n/a	C4	7763	10.7
PUUV VLP + αVβ3 integrin	C4	4	n/a	C4	1414	18.2
PUUV VLP + fab 4G2 (lattice)	C4	4	n/a	C4	1998	16.8
PUUV VLP + fab 4G2	C4	4	7621, 1722 **	C4	1721	15.0
HNTV VLP + fab C4	C4	4	7395, 2380	C1	2380	19.0

Table 8.5 Data processing details
Details of the subtomogram averaging applied. STA in Dynamo. 3D classification performed in relion's subtomogram averaging workflow. * = Futher refinement in Dynamo post classification. ** = A special case in which a combination of patch finding and cross-correlation cut-off were used instead of 3D classification. (See section 6.2 for details).

8.7. Additional ELISA data

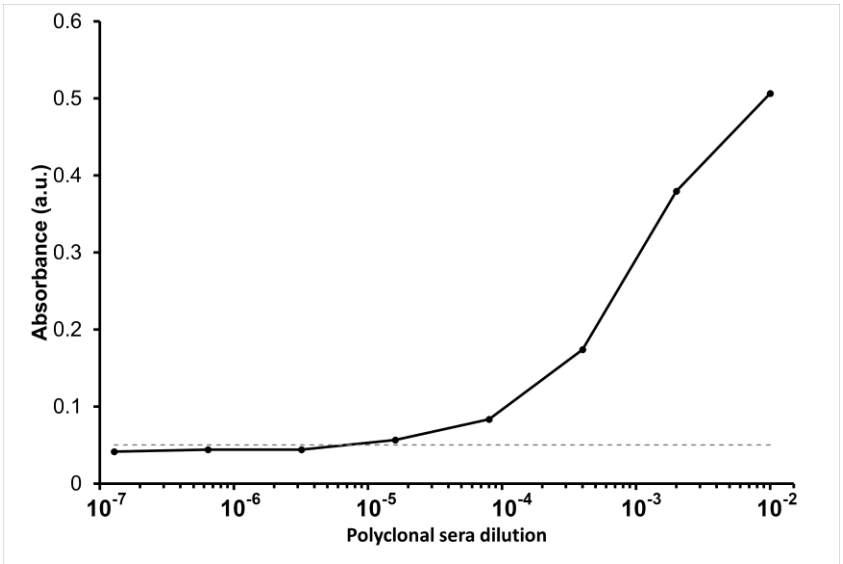


Figure 8.1 A positive control ELISA confirming the presence of TULV
A positive control ran alongside the ELISA in Figure 4.12 to confirm the presence of TULV on the plate. This ELISA utilised a serial dilution of some anti-PUUV Gn polyclonal sera which also detects TULV (a kind gift of Dr Hepojoki), detected with an HRP conjugated goat-derived anti-rabbit antibody (ab97080, Abcam). An analyte-free negative control well was included giving a background reading of 0.049 as indicated with a grey dashed line. To conserve sample, only single replicates of this control were performed.

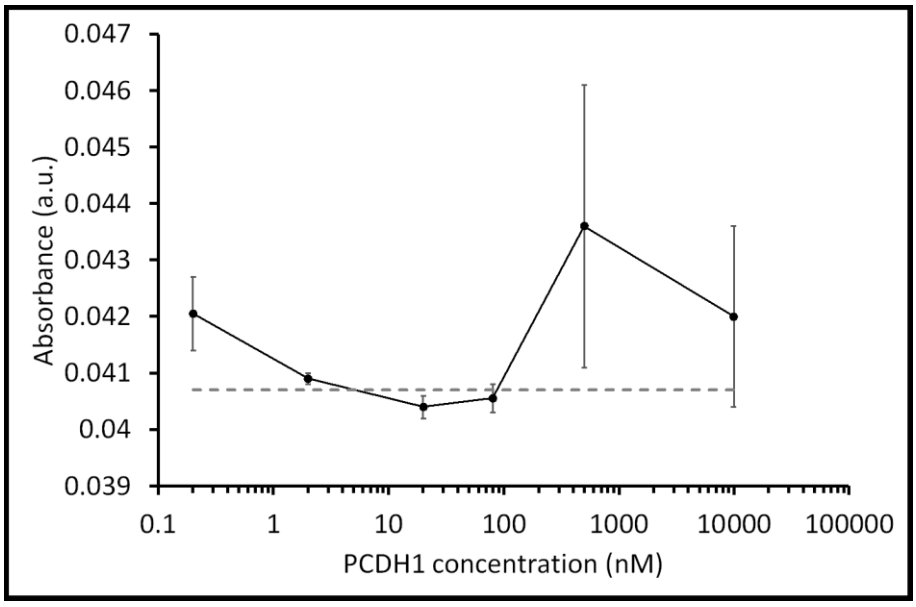


Figure 8.2 An attempt to monitor binding of ANDV VLPs to PCDH1 EC1-EC2 by ELISA
A sample purified ANDV VLPs were immobilised on a plate and probed with a dilution series of PCDH1 EC1-EC2. An analyte-free negative control well was included giving a background reading of 0.0407 as indicated with a dashed grey line. The 280nm absorbance of the original VLP sample was 0.52 a.u. which was diluted 250x before plating. An additional well was included with the same VLP sample diluted 50x. This gave a signal of 0.0475 when probed with 10000nM PCDH1 EC1-EC2. Note the scale of the y-axis. A positive ELISA signal would be expected to be > 0.1 a.u.

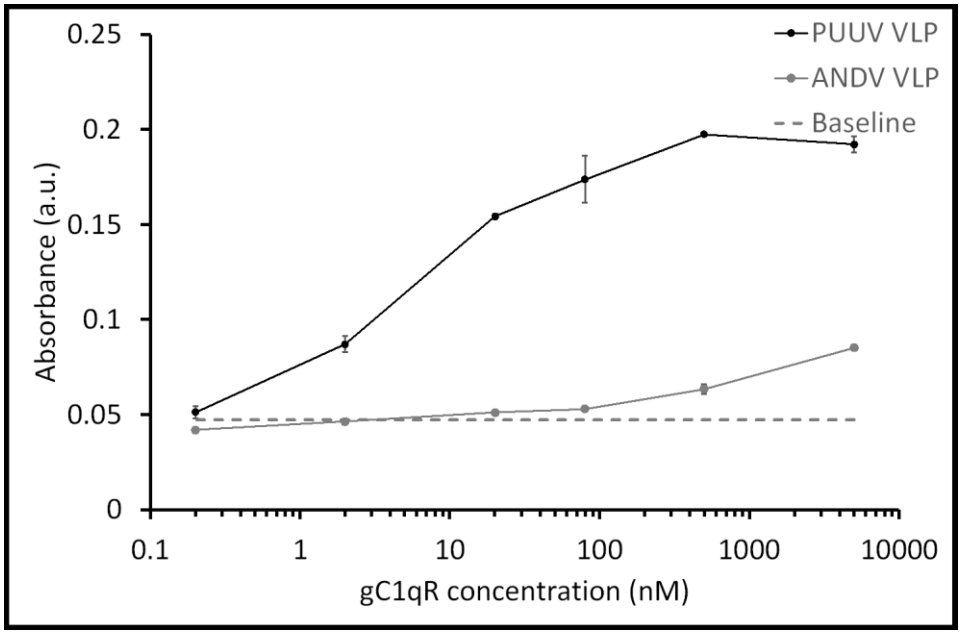


Figure 8.3 ELISA to demonstrate binding of gC1qR to PUUV VLPs with negative control displayed
Samples of purified PUUV VLPs and ANDV VLPs were immobilised on a plate and probed with a dilution series of gC1qR. An analyte-free negative control well was included giving a background reading of 0.0474 (indicated as baseline) and 0.0442 for PUUV and ANDV respectively. The 280nm absorbance of the PUUV VLP sample was 3.98 a.u. which was diluted 250x before plating. The 280nm absorbance of the ANDV VLP sample was 0.57 a.u. which was diluted 250x before plating. The final data point for ANDV VLP:PCDH1 EC1-EC2 at maximal concentration comes from a single replicate. This is included to demonstrate the limitations of the ELISA experiment utilised as discussed in section 5.4.

8.8. FSC curves

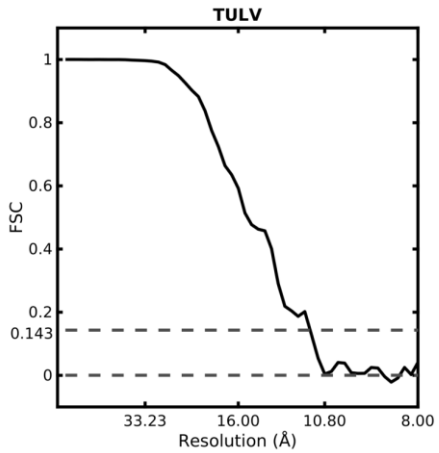


Figure 8.4 TULV FSC curve

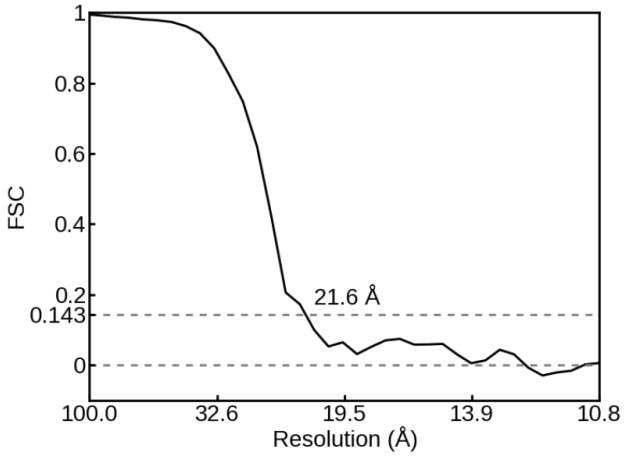


Figure 8.5 HAZV FSC curve

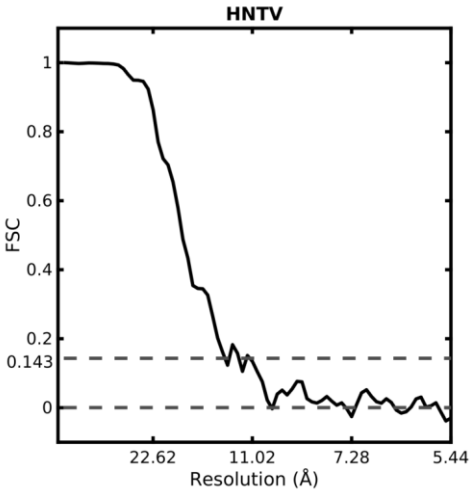


Figure 8.6 HNTV VLP FSC curve

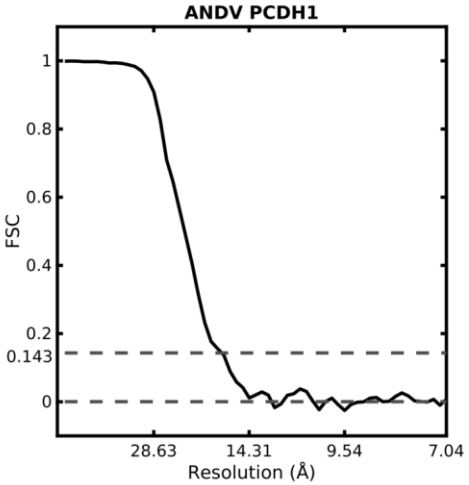


Figure 8.7 ANDV + PCDH1 EC1-EC2 FSC curve

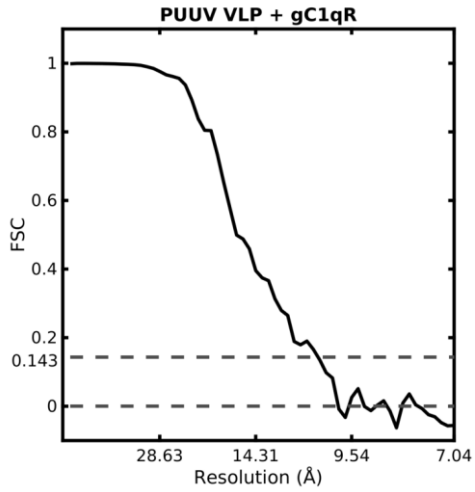


Figure 8.8 PUUV VLP + gC1qR FSC curve

8.9. SDS-PAGE

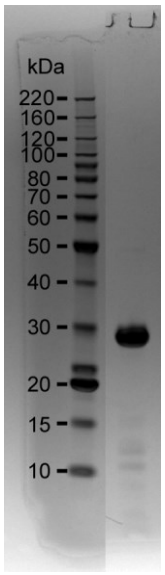


Figure 8.9 SDS PAGE of purified gC1qR
The sample was a kind gift from Dr Rissanen, Strubi. The expected MW of the protein is 25.1kDa.

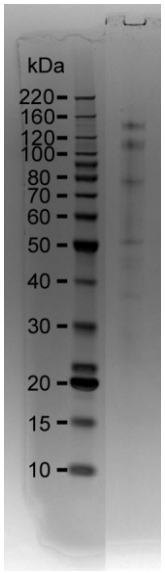


Figure 8.10 SDS PAGE of purified αVβ3 integrin ectodomain
The sample was a kind gift from Helen Duyvesteyn. The band profile is as expected compared to similar integrin samples (Personal correspondence, Helen Duyvesteyn, Strubi and Alison Burman, Pirbright).

8.10. Additional lattice quality data

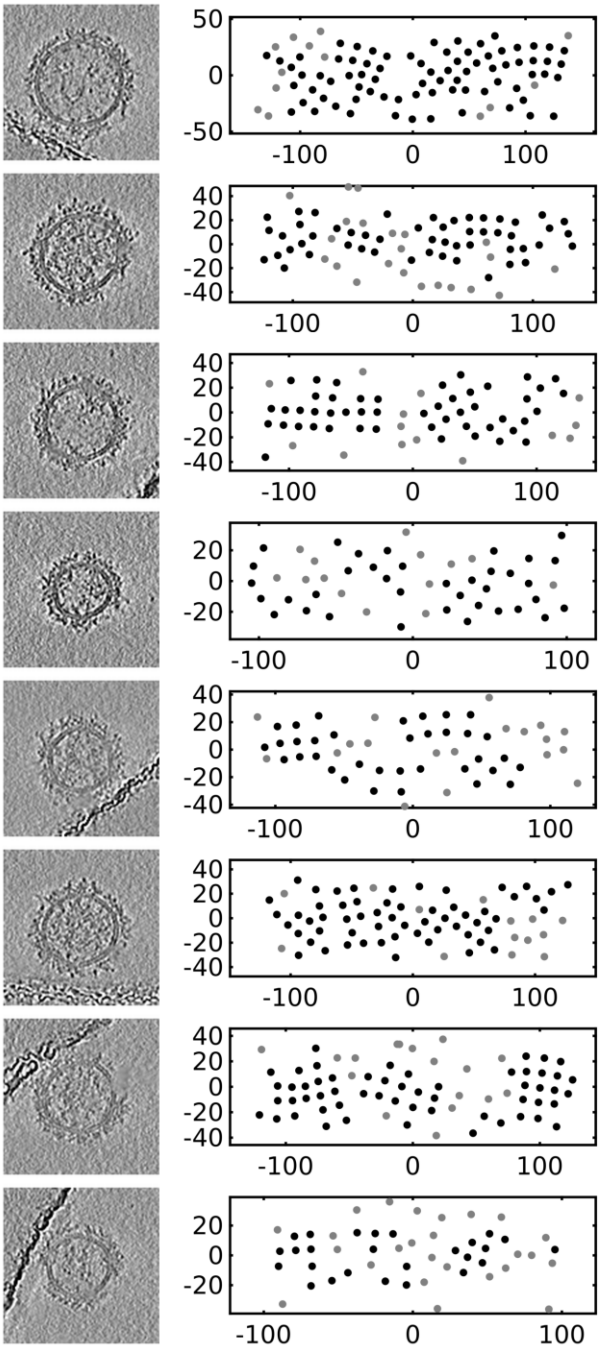


Figure 8.11 A demonstration of the lattice present on HNTV VLPs in the presence of fab C4 at 21°C
A computational experiment performed in which the eight VLPs displayed as tomographic slices on the left were refined using STA. The corresponding refined positions of the hantaviral spikes are displayed to the right as a Mercator projections (excluding particles on the top and bottom of the VLP. Units=nm). The subset of refined coordinates determined to belong to regular patches (using the patch finder script from section 3.5) are displayed in black.

8.11. Dose Weight Sharpening formula derivation

The dose weighting formula describing the dose dependant attenuation of the fourier components

$$W = e^{\frac{-d}{2(af^b+c)}}$$

Where

f = spatial frequency

d = dose

and a, b and c are experimentally derived constants from Grant and Gregorieff 2015 (a = 0.245, b = −1.665, c = 2.81)

For the nth tilt image in a tilt series this can be expressed as

$$W = e^{\frac{-(nd+p)}{2(af^b+c)}}$$

Where

d = dose per tilt image

p = pre dose applied before the series acquired (if any)

When this function is applied to all N tilts in a tilt series and finally averaged into a tomogram the

fourier components are weighted relative to each other as described by the sum (or average) of all the dose weighting functions applied.

This weighting is described by A where

$$A = \sum_{n=1}^{N} e^{\frac{-(nd+p)}{2(af^b+c)}}$$

Where

n = the tilt image number

N = the total number of images

For simplicity I will define

$$t = \frac{-1}{2(af^b + c)}$$

So the above formula can be re-written as

$$A = \sum_{n=1}^{N} e^{ndt+pt}$$

This can be further rearranged to

$$A = e^{pt} \sum_{n=1}^{N} (e^{dt})^n$$

The above sum can now be simplified using the general formula for the sum of a geometric series starting from 1 displayed below

$$\sum_{n=1}^{N} x^n = \frac{x(1 - x^N)}{1 - x}$$

So with x = e^{dt}

$$A = e^{pt}\left(\frac{e^{dt}\left(1 - e^{dtN}\right)}{1 - e^{dt}}\right)$$

A further simplification gives

$$A = \frac{e^{t(d+p)}\left(1 - e^{dtN}\right)}{1 - e^{dt}}$$

To express this sum as an average over the N tilts.

$$A = \frac{e^{t(d+p)}\left(1 - e^{dtN}\right)}{N(1 - e^{dt})}$$

Note that the sum vs the average does not change the shape of the curve so does not affect the weighting correction. It is included just to make the value easier to plot on a chart between 0 and 1.

So in summary, the artificial weighting that is applied to the Fourier components of a tomogram that has been reconstructed from dose weighted images is described by:

$$A = \frac{\sum_{n=1}^{N} e^{ndt+pt}}{N} = \frac{e^{t(d+p)}\left(1 - e^{dtN}\right)}{N(1 - e^{dt})}$$

Where

d = dose per tilt

p = pre applied dose

N = the total number of images

And

$$t = \frac{-1}{2(af^{b} + c)}$$

Where

f = spatial frequency

a, b and c are constants as defined by Grant and Gregorieff 2015 (a = 0.245, b = −1.665, c = 2.81)