Molecular determinants of phleboviral cell entry

Steinar Halldorsson

Division of Structural Biology

Nuffield Department of Medicine

Wadham 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 2017

Molecular determinants of phleboviral cell entry.

Steinar Halldorsson, Wadham College, University of Oxford
Doctor of Philosophy, Division of Structural Biology, Nuffield Department of Medicine.
Trinity term 2017.

Abstract

Phleboviruses are emerging zoonotic pathogens which constitute a global threat to human and animal health. The mosquito-borne Rift Valley fever virus (RVFV) is a widespread problem across the African continent and causes regular deadly outbreaks in ruminants. The recently emerged severe fever with thrombocytopenia syndrome virus (SFTSV) is a serious human public health concern in China which has rapidly spread to Japan and Korea with fatality rates as high as 16-30%.

Phleboviral cell entry is mediated by two viral glycoproteins: the class II fusion protein Gc and the lesser known Gn. Initial cell attachment is glycan dependent and the penetration into the cell cytoplasm is mediated by the Gc fusion protein which catalyses viral and cell membrane merger. The entry mechanism is not well understood from a structural perspective which limits mechanistic insights. The purpose of this thesis is to further our understanding of the cell entry process by filling in the missing structural information on the phleboviral glycoprotein layer. To this end, an integrated structural approach using cryo-EM and X-ray crystallography was adopted.

The crystal structure of the Gn ectodomain is presented which reveals an unprecedented structural relationship with seemingly unrelated viruses. Single-particle cryo-EM and localized reconstruction reveal the glycoprotein layer of the RVFV and a pseudo-atomic model of the RVFV is presented. The assembly shows the shielding of the Gc fusion protein and suggests that the Gn functions as a fusion chaperone. The post-fusion crystal structure of the Gc protein from SFTSV further consolidates a mechanism of membrane fusion by class II fusion proteins. Finally, preliminary data on receptor binding and mechanism of antibody mediated neutralization are presented. The work presented herein provides a novel platform for studying and understanding entry and assembly of phleboviruses as well as the design of novel therapeutics.

Word count: ~34,000 excluding abstract, table of contents, figures, references and appendix.

Acknowledgments

When it came to writing this part of a thesis, it dawned on me how many people I am grateful for having in my life and work. In a way, as it takes a whole village to raise a child, it takes a whole department to complete a DPhil. First and foremost, I would like to thank my two supervisors Juha T. Huiskonen and Thomas A. Bowden for their constant support, immense patience and great supervision throughout my time in Oxford. You both saw something in me that I wasn't sure I had. Thanks to senior members of their groups which I have thoroughly enjoyed working with and learning from: Antra Zeltina, Ilona Rissanen, Sai Li and Mengqiu Li. Thanks to my fellow DPhil students and junior scientist who have been a great source of comfort, laughter, help and gossip: Robert Stass, Alice Stelfox, Rhys Pryce, Young Kim, Simonas Masiulis, Felipe Moser, Helen excited state, Helen ground state, Weng Ng, James Hillier, Lea Sefer, Eamon Byrne, Christina Heroven, Sam Griffiths, Kirsty Sawtell, Vojtech Prazek, Michael Grange, Valerie Mordhorst and Nayab Malik. Especially thanks to those of you who frequented the afternoon coffee breaks, these were lifesavers during hard times. A special thanks to people who helped me out a lot experimentally as well as emotionally: Elizabeth Allen, Max Renner, Alex Diwa, Serban Ilca, Christiane Kowatsch, Joy Ho Shu Fen, Belinda Faust and Clare Bycroft.

I have the utmost gratitude for experimental help from these experienced people: Karl Harlos for experimental phasing, David Bitto for teaching me everything I know about viruses, Margaret Jones for keeping the wet lab running, Weixian Lu for tissue culture help, Robert Parsons for keeping the containment lab safe and Mark Langridge and Cornelia Cazey for keeping the electron microscopes in good condition. Other STRUBI people I would like to thank for their advice and help are Lindsay Baker, Cathy Whittle, Daven Vishishtan and Benjamin Vollmer.

My time in Oxford was challenging at times and I am thankful that I have wonderful people outside of the Oxford bubble who have helped me to keep my sanity in check. Thanks to my parents Ingibjörg Helgadóttir and Halldór Benediktsson who are my role models and the wisest people I know. Thanks to my brother Benedikt Halldórsson who showed me that anything is possible if you just put your mind to it. And finally, I need to thank the most special person in the world, Till Spanke, the person who has been my constant source of strength, who has given me unconditional love and filled my life with happiness. My Tillchen, you have given me more than you can ever imagine.

Table of contents

List of figures

List of tables

1. Introduction

1.1 Importance of Phleboviruses as pathogens

The *Phlebovirus* genus is a group of zoonotic, vector-borne viruses known to infect a great number of mammalian species. Many phleboviruses are pathogenic, such as Sandfly fever Naples and the Heartland virus, and can cause haemorrhagic fever [1]. Two of the most important members of this family for human health are the Rift Valley fever virus (RVFV) and severe fever with thrombocytopenia syndrome virus, both of which can be lethal to humans and livestock [2, 3].

RVFV was first discovered in Kenya in 1930 as the causative agent of an illness in sheep characterised by haemorrhagic fever, foetal deformation and abortion of pregnant ewes [4]. The virus can infect a large range of mammals including humans, mice, cattle and camels [3, 4], but the highest mortality rates reported so far have been in sheep, where 20% have died in clinical trials [5]. In humans, the disease is most commonly associated with self-limiting fever and flu-like symptoms [4], but in some cases the disease can progress into lethal haemorrhagic fever associated with hepatitis and in severe cases encephalitis [6, 7].

Several large outbreaks in humans and ruminants have occurred on the African continent. The largest human outbreak occurred in Egypt in 1977-1978 where an estimated 18,000 people developed symptoms and 600 people died [8, 9]. RVFV is spread by several species of mosquitoes, such as Culex, Anopheles and Aedes [9, 10] and as the demographic of these vectors change with a change in climate, so does the demographic of the disease they carry. In 2000 an outbreak of the virus in humans and livestock took place in Yemen on the Arabian peninsula [11, 12], confirming predictions that the virus was bound to

spread outside of the African continent and sparking fears that the virus will start to cause outbreaks in southern Europe [13]. Currently there is no approved human vaccine available against the virus but efforts are being made to develop non-replicative vaccines for a large range of domestic animals [14].

The RVFV is classified as a 'select agent' by the US Army Medical Research Institute of Infectious Diseases and as such it is considered a potential bioterrorism threat [15]. Even though this is perhaps an extreme stance, the potential for large outbreaks of RVFV and other phleboviruses in naive areas is real and large cross-national organisations such as the European Commission are funding fundamental research into these deadly viruses [16].

SFTSV is a recently discovered zoonotic phlebovirus which first emerged in the provinces of Henan and Hubei in China in 2009 [17]. The novel phlebovirus causes a disease characterised by deadly haemorrhagic fever, thrombocytopenia and leukocytopenia which has resulted in human fatality rates of 16-30% in clinical settings [18]. In the most deadly cases, encephalitis often precedes death [19]. Since its initial discovery, thousands of cases have been reported in China, South Korea and Japan [20, 21]. The virus is spread by several species of ticks and serological studies of domestic and wild animals suggests that the virus can infect a large number of mammals, including wild boars and rats [22–25]. A highly similar phlebovirus termed Heartland virus (HRTV) was discovered across the Pacific in rural Missouri, USA, causing similar clinical symptoms and is also suspected to be spread by ticks [26, 27]. Despite the high mortality rate of SFTSV in a clinical setting, most human infections may be asymptomatic [28]. Such sub-clinical cases and the large range of animal reservoir creates the perfect condition for undetected spread of SFTSV and HRTV. Currently there is no specialised treatment against these novel phleboviruses and

we have yet to learn more about these mysterious but deadly viruses. At least, we can expect more cases of SFTSV and HRTV infections in East Asia and North America.

1.2 Phleboviral molecular biology and entry mechanism

1.2.1 Overview

Ten distinct species have been defined within the *Phlebovirus* genus which is now a part of the newly defined the *Phenuviridae* family, within the large and diverse *Bunyavirales* order [29]. The genus can be further separated into two serogroups, the Uukuniemi virus-like group (including SFTSV) and the Sandfly fever virus group (including RVFV) [30]. The major distinguishing factors between these two groups is the absence of the NSm virulence factor in the Uukuniemi virus-like group, and viruses from the Sandfly fever virus group are generally transmitted by mosquitoes and phlebotomes while the Uukuniemi-like viruses are transmitted by mites [30].

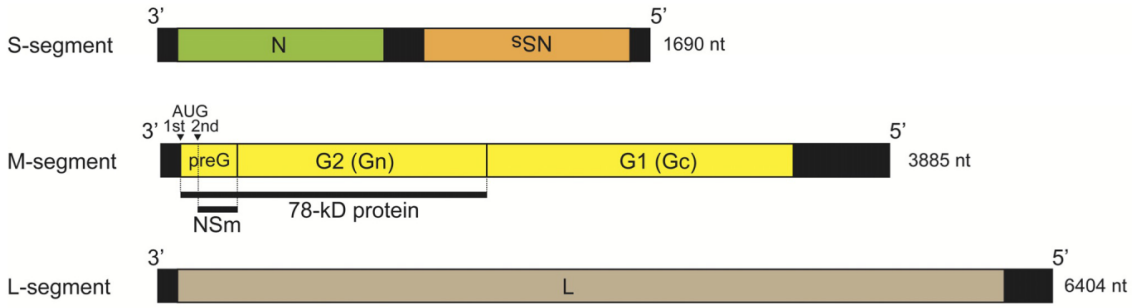


Figure 1.1. Genome organisation of RVFV.
The phleboviral genome is three segmented as exemplified by the RVFV genome. Figure is adopted from [2]. The S-segment codes in and ambi-sense strategy for the nucleoprotein N and the virulence factor NSs. The NSs is the only gene directly translated from the genome. The M-segment codes for the two glycoproteins Gn and Gc, and the virulence factor NSm. The large L-segments codes for the RNA-dependent RNA-polymerase.

The phlebovirus genome is three-segmented, negative sense RNA which codes for six or five proteins (Figure **1.1**). The S-segment codes for a virulence factor NSs, which

inhibits interferon signalling, and the nucleoprotein N which encases the viral genome [31, 32]. The S-segment encoded NSs is the only gene directly translated from the genome, i.e. the S-segment is an ambi-sense RNA molecule. The M-segment contains the two glycoproteins, Gn and Gc, which are expressed as a single polyprotein precursor and are cleaved by host signal peptidase to produce mature glycoproteins [33, 34]. Viruses of the Sandfly fever virus group additionally contain the NSm virulence factor, an anti-apoptosis factors that blocks the caspase cascade [35]. The L-segment codes for the RNA-dependent RNA-polymerase which uses a cap-snatching mechanism to synthesise positive-sense mRNA from the negative-sense RNA genome.

1.2.2 Entry pathway

The glycosylation pattern of pheboviruses plays a crucial role in the entry of phleboviruses into mammalian cells. RVFV, SFTSV and the Uukuniemi virus (UUKV, a non-pathogenic phlebovirus) have been shown to use a C-type lectin called dendritic cell specific ICAM-3 grabbing non-integrin (DC-SIGN) as an entry-receptor [36, 37]. DC-SIGN is a type-II transmembrane protein which is almost exclusively expressed on dendritic cells and is involved in the recognition and binding of oligo-mannose motifs [38]. The glycans found on the surface of enveloped viruses originating from mosquitoes, such as dengue virus (DENV) and Sindbis virus (SINV), are mostly composed of foreign oligo-mannose type glycans [39, 40]. Like phleboviruses, DENV and SINV can use DC-SIGN to enter mammalian cell and the entry is greatly enhanced if the virus is derived from mosquito cells [39, 40]. Similar studies have not yet been performed in phleboviruses, but the glycan content of UUKV and RVFV derived from mammalians cells has been shown to contain a population of oligo-mannose type glycans which is consistent with DC-SIGN binding [41, 42]. However, these two studies had different results regarding which of the two

glycoproteins is compatible with DC-SIGN binding. The UUKV study showed that only the Gc contained oligo-mannose type glycans while the RVFV study showed that both Gn and Gc contained such glycans. In addition to DC-SIGN, heparin sulfate acts to facilitate entry of RVFV and endocytosis is ultimately mediated by a clathrin or a caveolae-dependent pathway [43–45].

Once a phlebovirus is endocytosed, membrane fusion occurs to release the virus genome and the RNA-dependent RNA-polymerase into the host cell. The fusion pathway is summarised in Figure **1.2**. Fusion is pH-dependent and is mediated by the class II fusion protein Gc [45, 47]. The fusion mechanism of UUKV has been studied in some depth and has been shown to take place in the late endosomes at a pH of ~5.6 [45]. Furthermore, the fusion is greatly enhanced in vitro by the presence of the anionic lipid bis(monoacylglycero)phosphate (BMP) [48], which is highly abundant in the late endosomal membrane [49]. SFTSV entry on the other hand, has been shown to be dependent on glycolipid biosynthesis [50]. When glucosylceramide synthase was inhibited with a small molecule or knocked down, SFTSV accumulated in endosomal compartments, suggesting a role for glucosylceramide lipids in virus fusion and/or trafficking. Additionally, SFTSV has been shown to avoid receptor mediated entry altogether by utilising extra-cellular vesicles [51]. This mechanism allows the virus to infect neighbouring cells without interference from the host immune system.

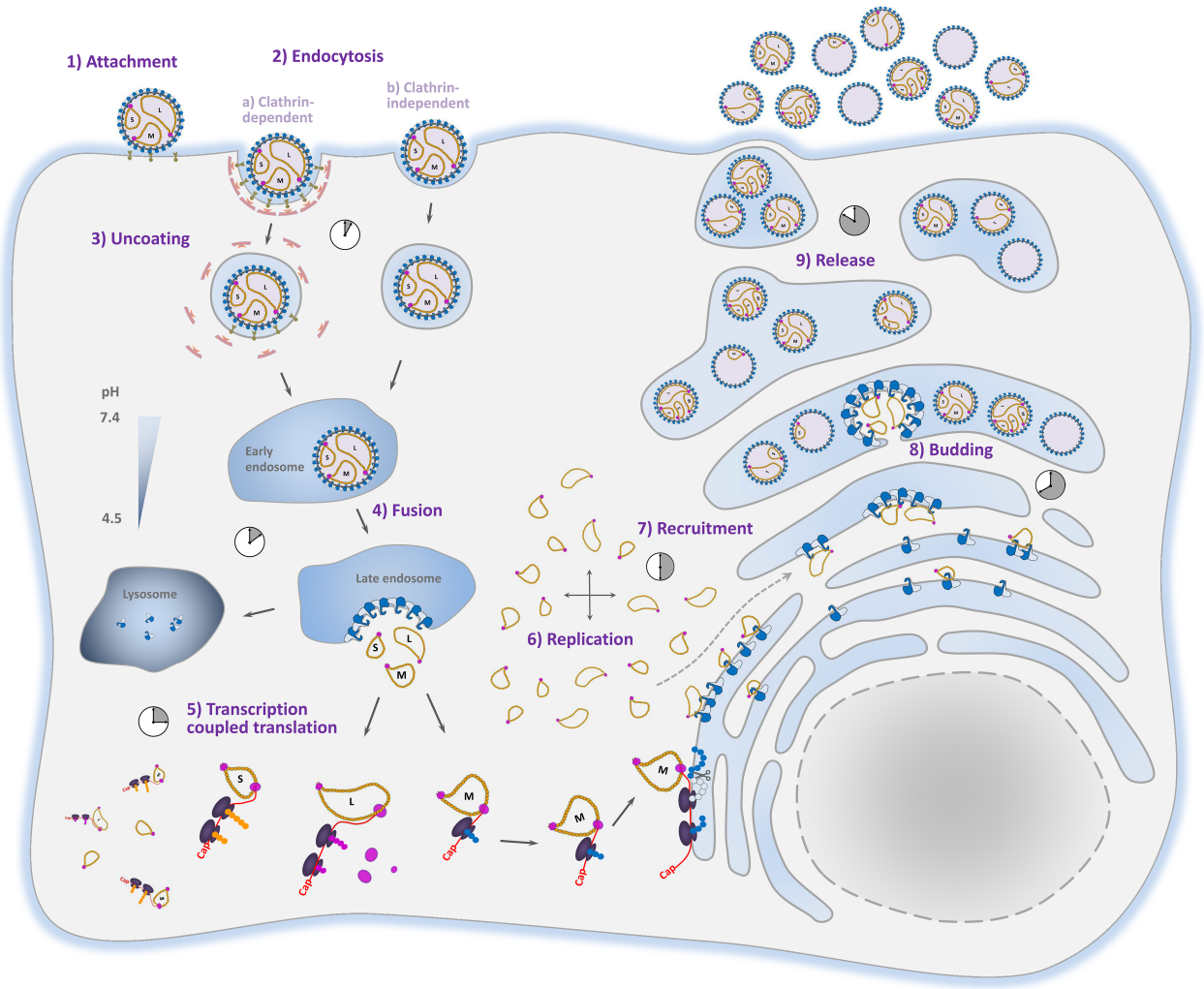


Figure 1.2 The life cycle of phleboviruses.
Phleboviruses such as UUKV enter cells through the endocytic pathway and assemble via budding in the Gogli network. Adapted from [46]. Steps **1–4** relate to virus entry while steps **5–9** relate to virus biosynthesis and budding. UUKV, RVFV and SFTSV attach to a cell via DC-SIGN on the cell surface (**1**). The virus is endocytosed via a clathrin or a caveolae mediated pathway, which brings the virus into the early endosome (**2–3**). As UUKV progresses through the endocytic pathway the pH drops to the fusogenic level (pH 5.6 is the half-point for UUKV fusion) which leads to merger of viral and late endosomal membranes (**4**). The presence of the late endosomal lipid BMP greatly enhances fusion for UUKV while glucosylceramide lipids have been indicated in SFTSV entry. Transcription, translation and replication of virus genome take place in the cytoplasm (**5-6**). The glycoproteins Gn and Gc are translated on ER-associated ribosomes where the cytoplasmic tails of the glycoproteins associate with the coated viral genome segments (**7**). The Gn contains a Golgi–localisation signal which brings the glycoproteins to the site of budding into the Golgi apparatus (**8**). The newly budded viral particles are finally released to the extra-cellular milieu via exocytosis (**9**).

1.2.3 Virus biosynthesis

Once fusion has occurred, genome replication and mRNA production take place in the host cell cytoplasm [52] (Figure **1.2**). The virulence factors NSs and NSm help to suppress the host cell response to the infection while replication and translation of viral proteins takes place. The whole M-segment is expressed as a single polyprotein which is cleaved into Gn and Gc (and NSm if present), presumably co-translationally in the ER [33, 34]. The Gn of UUKV and RVFV contains a Golgi-localisation signal which guides the protein, along with the Gc, to the Golgi [53–55]. If Gc is expressed in the absence of Gn it will accumulate in the ER, suggesting that Gn and Gc form a heterodimer as early as in the ER [53]. Budding of UUKV particles takes place in the Golgi, where the cytoplasmic tails of the Gn and Gc engage with the viral genome to ensure proper virion packaging [55–57]. The viral genome is packaged into ribonuclear particles which takes the shape of pan-handles and includes the viral RNA-polymerase [58, 59]. Upon budding of the virion into the Golgi, the particle is trafficked towards the plasma membrane and new particles are released to the extra cellular milieu [52].

1.2.4 Immune response to phleboviral infection

A key player in the early stages of RVFV infection is the NSs protein which acts as a suppressor of the innate immune response [60]. The protein downregulates general and interferon specific transcription [61, 62] and induces the degradation of protein kinase R [63]. Less is known about the immune response to SFTSV, but there is some evidence that suggests that the inability to mount a proper innate immune response against the virus is a determining factor for pathogenicity [64]. However, the major determinant of RVFV clearance and survival outcomes in infection models is a robust antibody mediated B-cell response [7]. Antibodies raised only against the Gn and Gc of RVFV confer full protection

when vaccinated livestock animals are challenged with the live virus [65]. Additionally, monoclonal antibodies (mAbs) raised against either the Gn or Gc neutralise effectively *in vitro* [65]. Hence, a vaccine which elicits a good B-cell response may be an excellent strategy in preventing and combatting outbreaks of RVFV and potentially other phleboviruses.

1.3 Class II fusion proteins and their chaperones

1.3.1 The class II fusion protein fold

The class II fusion protein fold is found in several RNA viruses and eukaryotes. The first structure of such a protein was the crystal structure of the ectodomain of the E protein from tick-borne encephalitis virus (TBEV, genus *Flavivirus*, family *Flaviviridae*), a negative sense RNA virus of the *Flavivirus* genus [66]. The 396 amino acids long protein formed as a head-to-tail dimer in the crystal with each protomer measuring ~120 Å in length (Figure **1.2A&B**). This dimer form is referred to as the 'pre-fusion' state since it is the form the protein adopts on the mature flaviviral virion surface [67]. The E protein was divided into three distinct domains simply termed I, II and III according to the genomic sequence, although domains I and II are segmented and interwoven. Domain I has a central eight-stranded up-and-down β-barrel and links the other two domains together. Domain II is involved in the homo-dimerization of the protein but also contains the fusion loop at the tip of a five-stranded β-sandwich. Domain III has an immunoglobulin like fold and is linked to domain I via a fifteen amino acid long linker, which was presumed to be largely flexible.

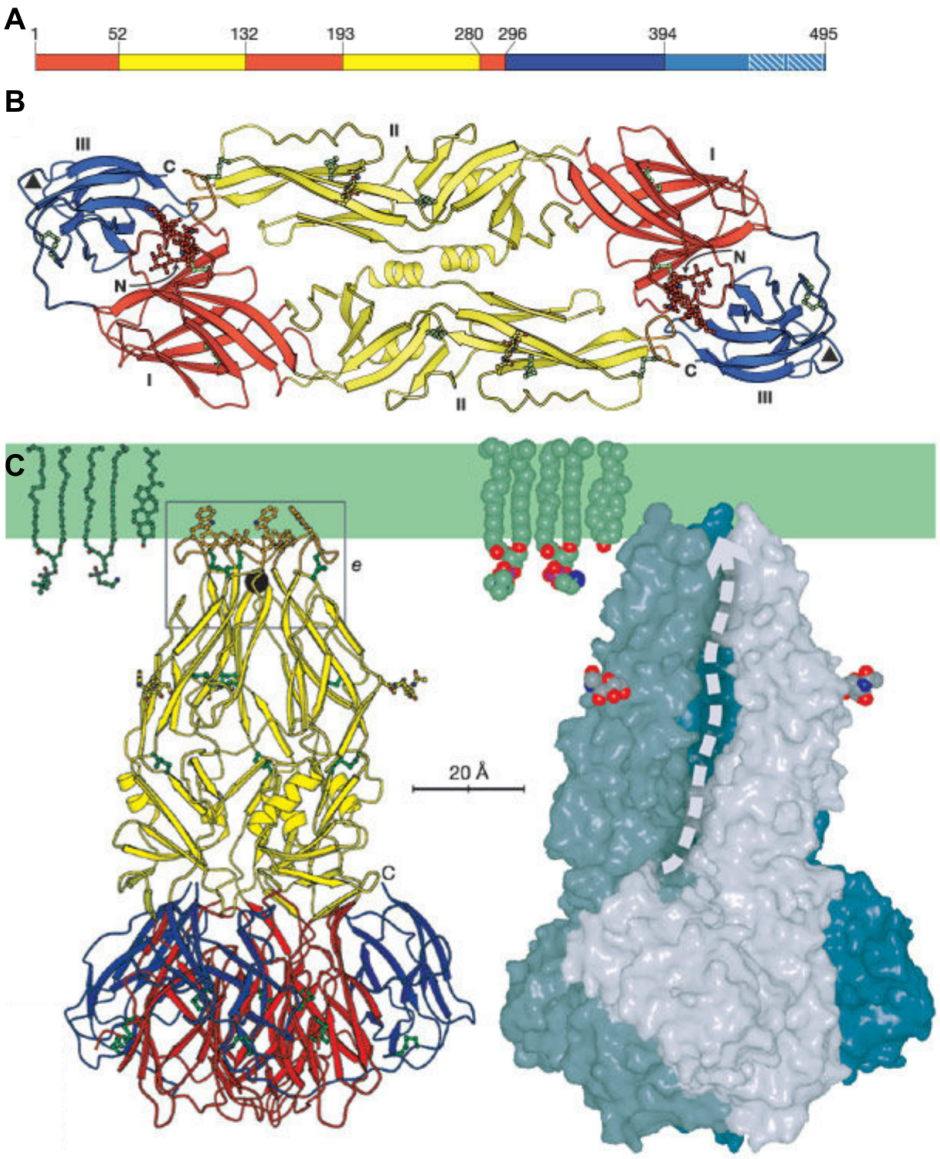


Figure 1.3. The flaviviral E protein.
The flaviviral fusion protein E has been captured in two distinct states by X-ray crystallography. Adapted from [68] A) Domain organization of the DENV E protein. Domain I is red, domain II is yellow and domain III is blue. The stem region connecting the C-terminus of domain III to the transmembrane region is in light blue and the transmembrane/intralumenal part is shown striped blue and white. B) The pre-fusion head-to-tail dimer of DENV E protein (PDB code 1OAN) shown in cartoon. Domains are coloured as in panel *A*. C) The trimeric post-fusion conformation of DENV E protein (PDB code 1OK8). On the right, shown in cartoon and domains are coloured as in panel *A* and fusion loop is shown in orange. The block in green represents the target membrane. On the left, the same trimer shown as electrostatic surfaces.

The first 'post-fusion' structure of a class II fusion protein was of the E protein from dengue virus [68] (Figure **1.3C**) which crystallised in a trimeric conformation as predicted

by previous biochemical studies [69]. The post-fusion trimer was captured in a purification scheme which involved exposing soluble E protein to low pH in the presence of liposomes followed by detergent purification to solubilize the E trimers from the liposomes [68]. The conformational changes involved in the mechanism of fusion are described in more details in **1.3.3**, but the major difference between the conformation of the pre- and post-fusion states is the dramatic translational shift which domain III undergoes. The shift brings the C-terminus of domain III almost 40 Å closer to the fusion loop at the tip of domain II.

Many more crystal structures of class II fusion proteins have been captured by X-ray crystallography in the pre- and post-fusion states (Figure **1.4**). This fold has been found in other positive sense RNA viruses such as SFV (genus *Alphavirus*, family *Togaviridae*) and Rubella virus (RUBV, genus *Rubivirus*, family *Togaviridae*) [70, 71]. Surprisingly though, the structure of E2 protein from bovine viral diarrhoea virus (BVDV, genus *Pestivirus,* family *Flaviviridae*) revealed a novel fold with little similarity to class II fusion proteins, despite the close phylogenetic relationships between pestiviruses and flaviviruses [72]. Several negative sense RNA viruses of the *Bunyavirales* order also contain a class II fusion protein such as RVFV (genus *Phlebovirus,* family *Phenuiviridae*) and Puumala virus (PUUV, genus *Orthohantaviridae*, family *Hantaviridae*) [47, 73].

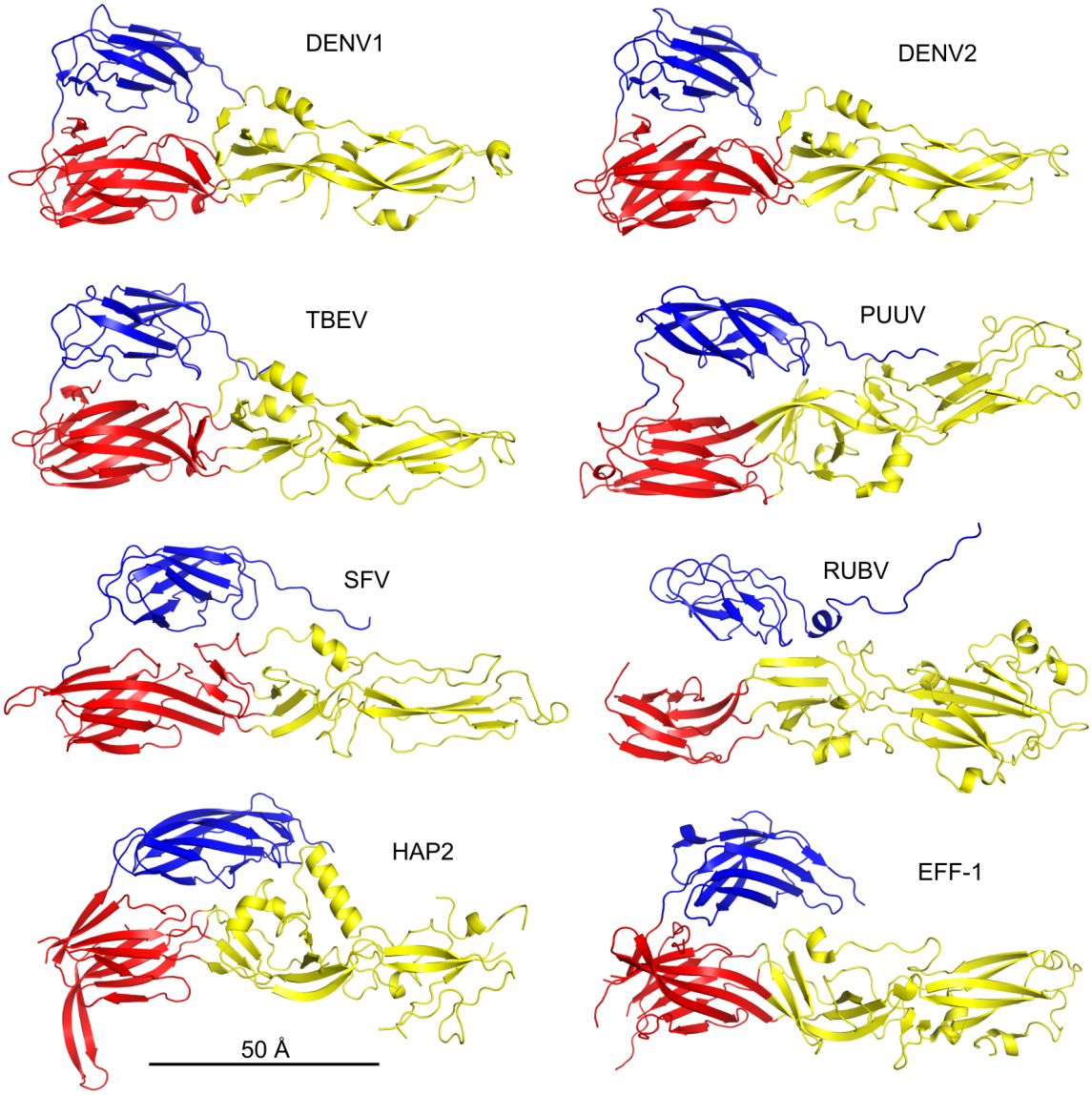


Figure 1.4. The class II fusion fold is found in a variety of viruses and eukaryotes. Eight representatives of the class II fusion fold are shown, all are protomers from a post-fusion trimer. The structures are coloured as in **Figure 1.3**, apart from the fusion loop which has not been highlighted. Structures have been solved from several RNA viruses such as: Dengue virus type 1 (DENV1, PDB code 3G7T), Dengue virus type 2 (DENV2, PDB code 1OK8), tick-borne encephalitis virus (TBEV, PDB code 1URZ), Puumala virus (PUUV, PDB code 5J9H), Semliki Forest virus (SFV, 1RER) and Rubella virus (RUBV, PDB code 4ADI). The same fold has also been found in eukaryotes: HAP2 (PDB code 5MF1) from the unicellular algae *C. reinhardtii* and EFF-1 (epithelial fusion failure 1, PDB code 4OJD) from the worm *C. elegans*.

The class II fusion fold has often been speculated to originate from a non-viral origin and indeed two class II fusion protein structures from eukaryotes have been solved. EFF-1 from the multicellular *C. elegans* is involved in normal cell-cell fusion and is a

unique class II fusion protein which lacks the canonical hydrophobic fusion loop residues [74]. The mechanism of fusion has been hypothesised to differ from regular class II fusion and may require EFF-1 to be present on both membranes that are to fuse. Another class II fusion protein was found in a single-cell green algae *C. reinhardtii* [75]. The HAP2 fusogen is involved in sexual reproduction and homologues of this protein are found throughout the eukaryotic taxon, including higher plants and unicellular protozoa. Unlike EFF-1, HAP2 has the canonical features associated with class II fusion proteins, such as fusion loops at the apex of domain II, but the trigger of fusion is not yet known. The similarity between HAP2 and other viral class II fusion proteins is so striking that the probability of convergent evolution is negligible. Currently, the most plausible explanation is horizontal gene transfer by reverse transcription. Transposable elements have been found in *C. elegans* which are of a phleboviral origin and contain a putative class II fusion protein sequence [76]. As ever, it is impossible to discern whether the fold originated from a virus or from a cell, but we can be certain that the popular class II fusion fold is likely to be found in surprising places.

1.3.2 The chaperones of class II fusion proteins

All viruses known to use a class II fusion protein for fusion also have an additional envelope-bound protein, which is indispensable for the virus. In the case of DENV (flavivirus), the M protein is present on the viral envelope and has been proposed to act as a 'latch' which keeps the E protein locked on the virion surface [77]. The small M protein of ~72 amino acids, contains 20 amino acids which forms a tight set of interactions with E at the interface between E homodimers on the mature virion (Figure **1.5**). Three histidines of M form key contacts at this interface and have been suggested to act as a pH sensing switch which releases the E protein and allows it to interact with a target membrane at the appropriate pH. The M protein is also important in the maturation of the virion and has

been suggested to 'spring load' the E protein when the immature spiky virion changes into the mature smooth virion.

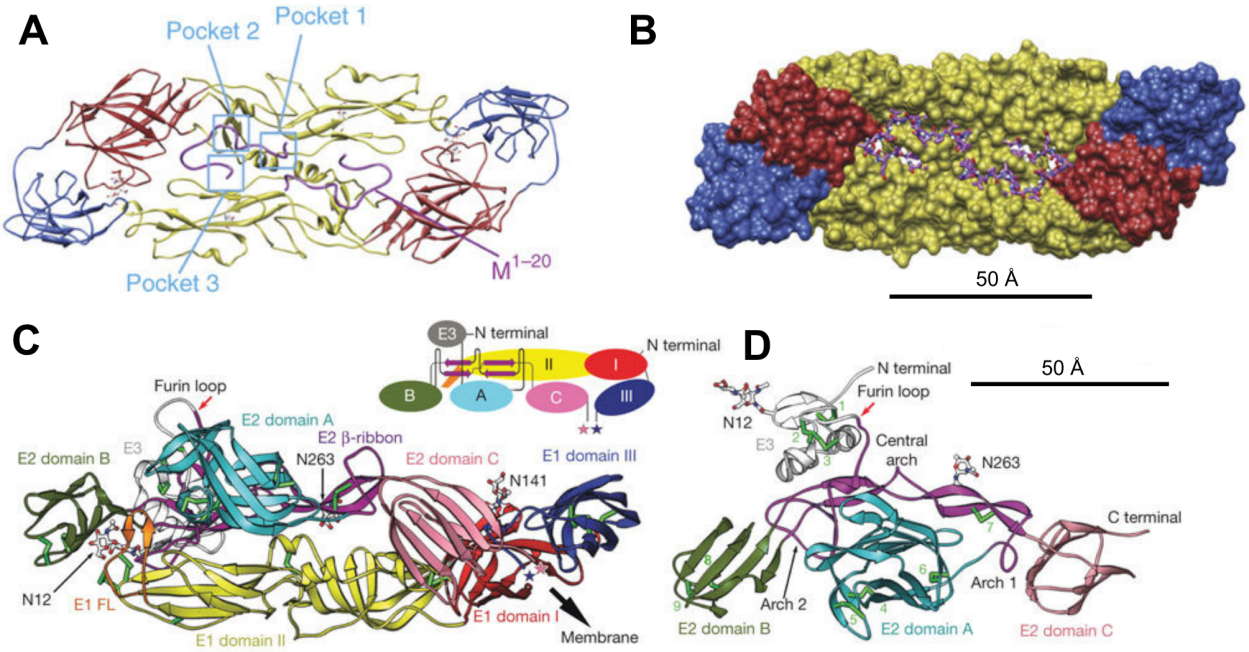


Figure 1.5. Class II fusion protein chaperones of flavi- and alphaviruses.
The fusion proteins of alpha- and flaviviruses require a fusion chaperone protein for proper function. Figures in panels *A* and *B* are adapted from [77]. Figures in panels *C* and *D* are adapted from [78]. In all figures the class II fusion protein is coloured according to domain as in Figure **1.3A**. A) The atomic model of the ectodomains E and M from DENV represented in cartoon, as revealed by cryo-EM [77]. The homodimer of the fusion protein E is held together by two M proteins (coloured magenta), which forms tight interactions at three sites (pockets). B) Same structure as in panel *A* but shown with E in surface representation and M in sticks. C) The atomic structure of the E1, E2 and E3 heterotrimer from CHIV shown in cartoon. E2 is coloured according to domains, domain A is teal, domain B is forest green, domain C is pink and the β-ribbon is purple. E3 is coloured light grey. The inset shows a schematic diagram of the trimer coloured according to domains. The E1 and E2 make extensive contacts which include all the different domains from each protein and the fusion loop of E1 is shielded in a groove between domains A and B of E2. D) The structure of E2 and E3 (without E1) shown as in panel *C*.

Alphaviruses also have a chaperone protein present on the virion surface. The crystal structure of the E1–E2 heterodimer from Chikunguya virus (CHIV) provided the first atomic level detail of the interaction between the fusion protein E1 and its chaperone E2 [78]. The long fusion loop in alphaviruses (~14 amino acids) is sequestered from solvent

at the interface between domains A and B in E2. The 421 amino acid long E2 protein makes much more extensive contacts with the E1 than the flaviviral M protein makes with E protein. The E1 and E2 interaction involves all the different domains of each protein and buries a surface area of ~2500 Å^2. As is the case for flaviviruses, histidine side chains from E2 stabilise the E1–E2 heterodimer. The E2 is also important for the higher-order assembly of the virion where E2–E2 contacts between different E1–E2 heterodimers are crucial for the formation of the asymmetric unit [78–80].

Another putative chaperone protein structure exists, the Gn from Puumala virus (PUUV), but the exact nature of the interaction with Gc, its class II fusion protein partner, is unknown [81]. Hantaviruses are largely pleiomorphic so high-resolution single-particle cryo-EM is not applicable. The PUUV Gn was fitted into a cryo electron tomography (cryo-ET) reconstruction of Tula virus, which showed that Gn is membrane distal and most likely sequesters Gc from solvent. However, at the time of publication there was no Gc crystal structure available and the Tula virus (TULV) density was only at 16 Å resolution which greatly limits the accuracy of such predictions. The hantaviral Gn may not even function as a chaperone for Gc, but a recent review established a structural link between the the hantaviral Gn and the alphaviral E2 which may be evidence to support a fusion chaperone function of the Gn [82]. These two seemingly unrelated proteins share some common secondary structural features and have a similar domain organisation and may have evolved from a common ancestor.

1.3.3 Mechanism of fusion for class II fusion proteins

Most of what is known about the mechanism of fusion of class II fusion proteins comes from studies of flaviviruses. A mechanism was proposed in 2004 when the post-fusion structure of DENV E protein was published and that mechanism has largely remained

unchallenged since [68, 83]. From X-ray crystallographic studies, the starting state and the end state of the fusion proteins are known but the intermediates have been difficult to capture, presumably as these are unstable. A crystal structure of RVFV Gc in an extended conformation may constitute an extended intermediate and arrays of extended intermediates of the UUKV Gc inserted into a target membrane have been observed by cryo-ET, which are consistent with the length of the RVFV Gc 'intermediate' structure [47, 48].

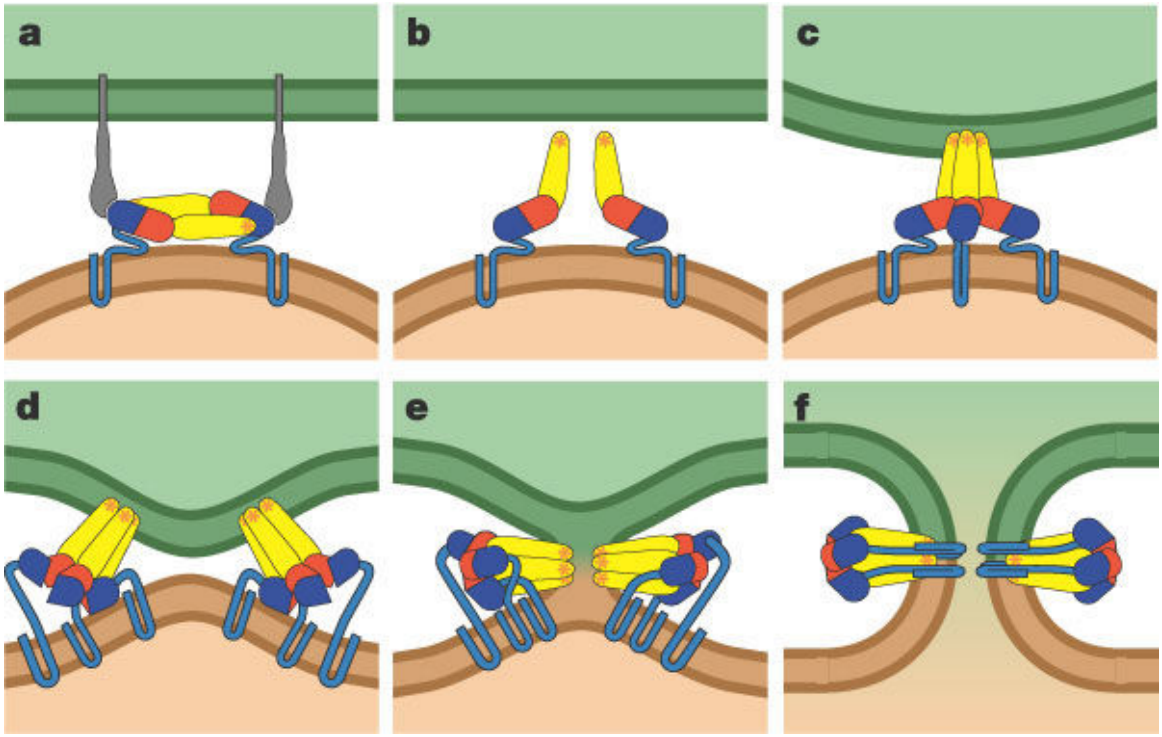


Figure 1.6 Mechanism of class II viral fusion.
The six stages of membrane fusion hypothesised for class II fusion proteins, exemplified by DENV E protein. Figure is adopted from [68]. The figure shows an artistic model of E protein coloured according to domains as in **Figure 1.2A**. a) The fusion protein is in the pre-fusion state on the virion surface prior to activation by low pH or potentially a protein on the target membrane. b) The pre-fusion dimers dissociate and the fusion protein rises to interact with a target membrane. b) Three E protein molecules come together and form a trimeric intermediate. d) The stem region and domain III start to snap back and the concerted action of two or more trimers brings the two membranes closer together. e) As the post-fusion trimer start to take shape the first rate-limiting step of hemi-fusion takes place. f) Full fusion is believed to be the result of the formation of the post-fusion state when the transmembrane regions and the fusion loops are drawn close together.

The fusion protein exists on the virion surface in a metastable, pre-fusion state (Figure **1.6a**), which is stabilised by a fusion chaperone protein. Low pH environment in endosomal compartments leads to protonation of key histidine side chains on the fusion protein and its chaperone, which destabilises the interaction between the two and frees the fusion protein [84]. The fusion protein is now free to interact with a target membrane and initial fusion loop insertion of the extended intermediate takes place (Figure **1.6b**). The next intermediate steps have not been observed directly by structural methods, but biochemical and kinetic studies of flaviviral virus like particles (VPLs) have lent evidence to support their existence. The extended intermediates cluster together and most likely form trimers (Figure **1.6c**). Domain III and the C-terminal stalk of the fusion protein may start to re-arrange and begin to draw the two membranes closer together (Figure **1.6d**) and several trimers may cluster together to form dimples in the two membranes. Hemi-fusion appears to be a major rate-limiting step and has been suggested to require the energy from at least two trimers (Figure **1.6e**) [85, 86]. Full fusion is the second rate-limiting step and is believed to be achieved when all three domain III in a trimer have arranged into the post-fusion state (Figure **1.6f**) and the C-terminal transmembrane helices have arrived at the fusion loops.

Based on structural and functional evidence, the mechanism of fusion of class II fusion proteins is believed to be much the same for different groups of viruses [87]. All viral class II fusion proteins have a known chaperone partner, class II fusion is low-pH triggered and the fusion proteins exist in at least the two well-characterised pre- and post-fusion states.

1.4 Cryo-EM and class II fusion proteins

1.4.1 Same protein, different icosahedral assemblies

The first low resolution cryo-EM structure of an enveloped virus was that of SINV in 1987 [88]. It took until 2001 for the first atomic-level model of an enveloped virus to be published, that of SFV at 3.5 Å resolution [70]. A crystal structure of the E2 protein was not available and de-novo model building from cryo-EM models was not well established at that time and as a result the model lacks an atomic structure of E2. Additionally, resolution estimation in cryo-EM had not converged on the 'gold-standard Fourier shell correlation', which was only introduced a couple of years later [89], so the resolution may have been over-estimated. Nonetheless, the map shows clearly that the E1 fusion protein does not assemble into a head-to-tail pre-fusion dimer as expected from the crystal structure of the flaviviral E protein [66]. Further cryo-EM and crystallographic studies of CHIV and Venezuelan equine encephalitis virus (VEEV) allowed for an atomic level structure of the entire glycoprotein layer [78–80]. Interestingly, the matrix proteins bellow the envelope assemble into a regular T=3 icosahedron while the surface glycoproteins assemble into a T=4 quasi-icosahedron of four E1-E2 heterodimers in each asymmetric unit. The heterodimers form trimers which are held together by contacts involving both proteins while the trimers of heterodimers interact exclusively via E1-E1 contacts.

The first cryo-EM map of a flavivirus was published in 2002 but was only at 24 Å resolution [67]. However the fitting of the available crystal structure of the E protein homodimer from TBEV into the DENV map proved to be fairly accurate when compared to a higher resolution structure [77]. DENV assembles into a T=3 icosahedron with 3 quasi-equivalent E proteins in each asymmetric unit. Each E-E homodimer is stabilised by two M proteins but in contrast to the alphavirus assembly, most contacts on the flaviviral surface are between E proteins. As a result, the flaviviral virion has a very smooth surface

made up of flat E homodimers while the alphavirus virion has characteristic trimeric spikes of E1-E2 heterodimers which face upwards at an angle of ~25° from the virion membrane.

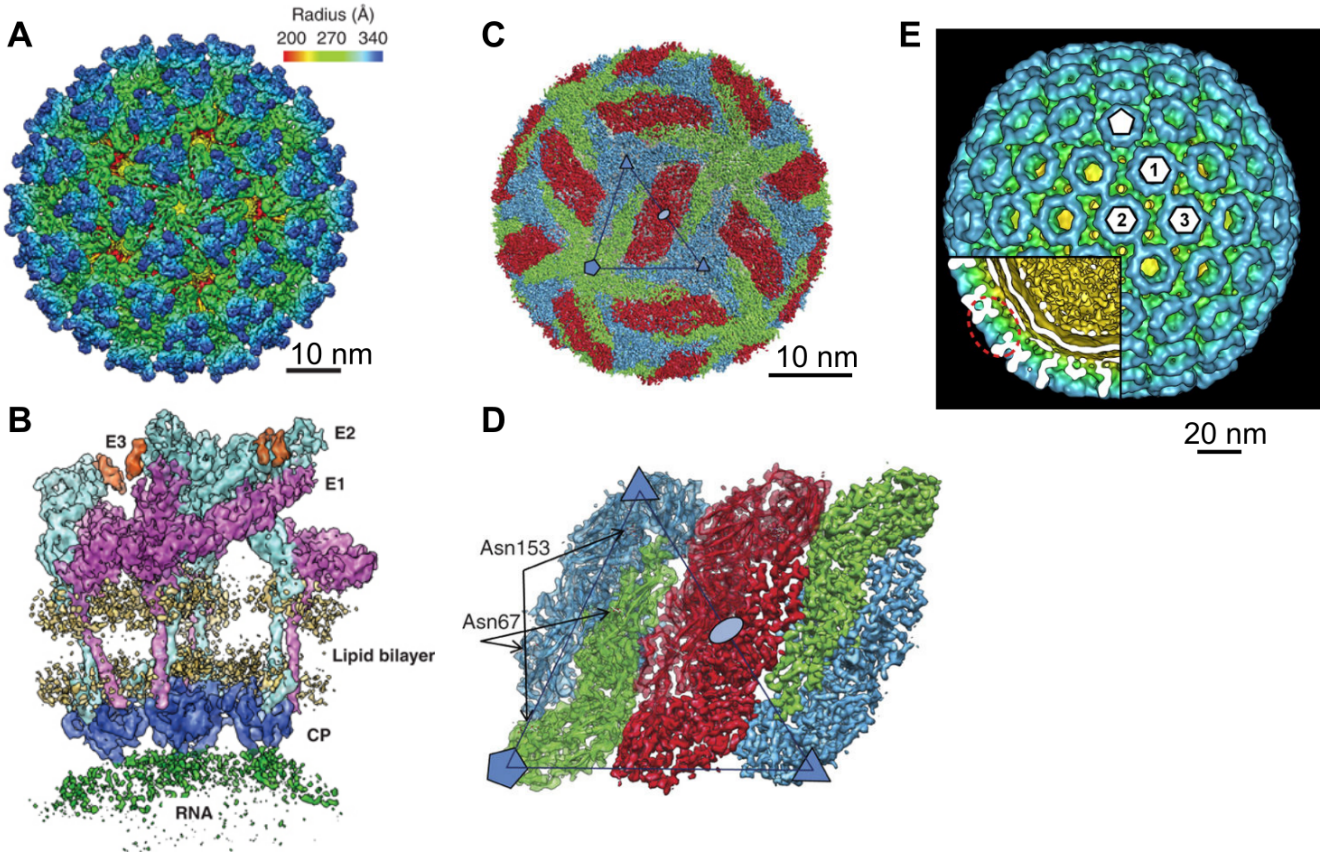


Figure 1.7. Cryo-EM structures of viruses containing a class II fusion protein. Single-particle cryo-EM has been a key technique in understanding the organisation of enveloped viruses which contain a class II fusion protein. Panels *A* and *B* are adapted from [80], panels *C* and *D* are adapted from [77] and panel *E* is adapted from [90]. A) Map of the alphavirus VEEV at 4.4 Å resolution showing the characteristic trimeric spikes on the virion surface made up of E1-E2 heterodimers. The map is coloured according to distance from the virion center. B) A section of the density from panel *A* showing the asymmetric unit from the side. E1 is coloured in magenta, E2 is teal, E3 is orange and the capsid protein (CP) is blue. The glycoproteins (E1, E2 and E3) interact with the CP under the lipid bilayer via the endodomains. E1 and E2 ectodomains form tight interactions which help to organise the trimeric spikes. C) A map of the flavivirus DENV at 3.5 Å resolution. The E protein is shown in three different colours (red, blue and green) according to quasi-equivalence. D) A detail of the asymmetric unit from panel *C*. The three non-equivalent E proteins lie flat on the virion surface and arrange in the canonical pre-fusion dimer as seen by X-ray crystallography [91]. E) Map of the phlebovirus RVFV at 22 Å resolution. The map shows the hexamers (marked from 1 to 3) and the pentamer on the virion surface which are made up of the Gn and Gc glycoproteins, the location of which have not been determined.

Lastly, the group of viruses which are the main subject of this thesis, and also assemble into icosahedral particles, are phleboviruses. There are currently only low resolution structures available of RVFV and UUKV [90, 92, 93] and from these we know that phleboviruses form large ~110 nm particles with a T=12 icosahedral symmetry from three distinct hexamers and a pentamer. As previously mentioned, the Gn and Gc assemble into heterodimers before budding, which then somehow form four distinct types of capsomers on the virion surface.

Even though the fusion proteins of phlebo- alpha- and flaviviruses are very similar they do not lead to the same icosahedral assembly. The presence of very different fusion chaperones is undoubtedly a contributing factor.

1.4.2 Pleiomorphic viruses with class II fusion proteins.

Much of the focus of cryo-EM on viruses has been on icosahedral particles as they are highly symmetric and relatively easy to reconstruct. However, there are of course many important pleiomorphic viruses which are more challenging subjects for cryo-EM. Viruses of the *Bunyavirales* order are predicted or have been shown to contain a class II fusion protein [47, 73, 94, 95]. Interestingly though, viruses of different families within the *Bunyavirales* order assemble very differently. Orthobunyaviruses contain tripodal spikes that form local three-fold assemblies on the virion surface [96] (Figure **1.8A).** Hantaviruses are largely pleomorphic but their glycoproteins form a local tetragonal assembly with a local four-fold symmetry of Gn-Gc heterodimers (Figure **1.8B**) [81, 97]. And as mentioned above, phleboviruses assemble into spherical particles with icosahedral symmetry [90].

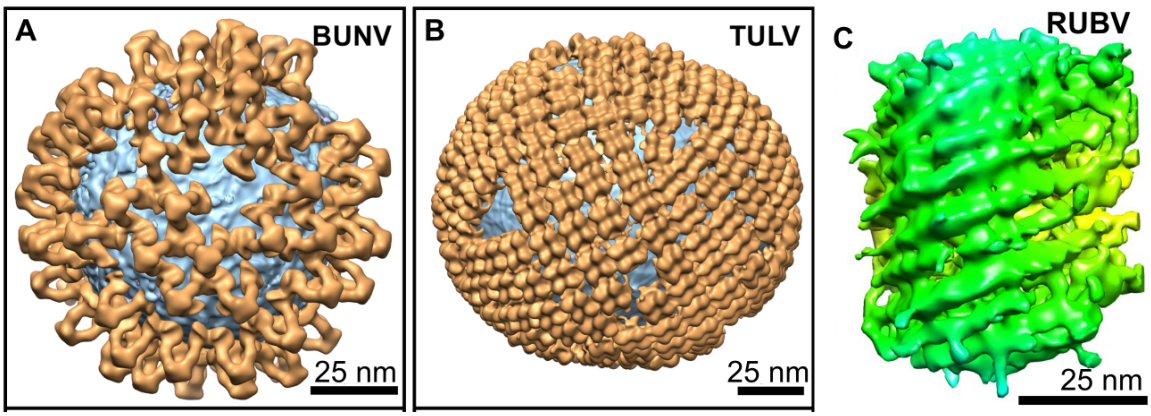


Figure 1.8. Pleomorphic viruses which contain a class II fusion protein.
Despite the well conserved nature of the class II fusion protein fold, the ultrastructure of viruses which contain such a fusion protein is highly varied. Panels *A* and *B* are adapted from [96], the membrane is coloured in light blue and the glycoprotein layer is in orange. Panel *C* is adapted from [98]. A) Bunyamwera virus (BUNV, genus *Orthobunyavirus*, order *Bunyavirales*) has tripodal spikes on the surface with a 3-fold symmetry. B) Tula virus (TULV, genus *Orthohantavirus*, order Bunyavirales) displays four-fold symmetrical spikes which form local two-dimentional curved sheets of repeating units. C) Rubella virus (RUBV, family *Togavirales*, order unassigned) has recently been shown to form a helical particle made up of rows of glycoproteins that twist around the envelope.

RUBV is a positive sense RNA virus that is phylogenetically related to alphaviruses, contains a class II fusion protein [71] and is pleiomorphic [99]. A recent cryo-ET study of RUBV showed that the virus glycoprotein layer has a loose helical symmetry made from rows of E1-E2 heterodimer [98] (Figure **1.8C**), an organisation very distinct from the icosahedral envelope of alphaviruses. The E1 fusion protein structure of RUBV is quite divergent from the alphaviral E1 [71] (Figure **1.4**) and most likely the E2 of RUBV is distinct from the E2 of alphaviruses based on the sequence divergence and viral assembly differences. Since single-particle cryo-EM is not easily applicable to such pleomorphic viruses, it may require a mixture of low resolution cryo-ET and high resolution X-ray crystallography to reveal the atomic-level organisation of their glycoproteins.

1.5 Aims

The major aim of this thesis was to structurally characterise the phleboviral glycoprotein layer and understand how the Gn and Gc contribute to cell entry. Previous work within the Huiskonen group has uncovered some of the fundamental biochemical requirements for membrane fusion [48] and other work has identified DC-SIGN as the entry receptor for phlebovirses [37, 100], but structural details of these processes are lacking. X-ray crystallography has thus far provided a structure of Gc in the pre-fusion conformation [47] but a structure of the post-fusion conformation is lacking. Until very recently a structure of a phleboviral Gn was lacking [101]. The best current model of a whole phleboviral virion by single-particle cryo-EM is only at 22 Å resolution [90] and needs to be improved to understand the interactions between the Gn and Gc.

The first part of the thesis was to use X-ray crystallography to fill in the structural gaps of our knowledge of the phleboviral glycoproteins. I sought to solve the crystal structure of the Gn and to capture the Gc in a post-fusion conformation. Once both Gn and Gc structures are known one can start to build a model of the entire virion. The second part of the thesis was to obtain an atomic model of an entire phleboviral virion. I sought to obtain a high-resolution map of the RVFV by single-particle cryo-EM and use the crystal structures of the Gn and Gc to build a model of the glycoprotein layer. When the phleboviral glycoprotein arrangements are known, one can start to explore the molecular details of functional processes. In the last part of this thesis I used a combination of X-ray crystallography and single cryo-EM with the aim of understanding DC-SIGN interaction and antibody neutralisation in the context of the whole virus. I aimed to obtain cryo-EM structures of RVFV bound to DC-SIGN and neutralising antibodies, and crystal structures of Gn-antibody complexes.

The work presented in this thesis is the first comprehensive structural study of the phleboviral glycoproteins and helps to consolidate a class II fusion protein mediated mechanism of entry of phleboviruses.

2. Materials and Methods

2.1 Molecular cloning

2.1.1 Plasmids and cDNA

The pHLsec plasmid vector [102] was used for recombinant protein expression in mammalian cells with a secretion signal and a hexa-histidine tag for purification (Figure **2.1**). The vector is based on the pLEX, vector and contains an ampicillin selection marker, a pBR322 origin of replication, a chick beta-actin promoter, a rabbit beta-globin polyA signal and a multiple cloning site with a Kozak sequence, a secretion signal, a hexa-histidine tail and a termination codon. The *Age* I and *Kpn*I restriction sites were used to insert an in-frame construct 3' of the secretion signal and 5' of the hexa-histidine tail. Genes of the M-segment of RVFV (UniProt accession no. P21401) and SFTSV (UniProt accession no. R4V2Q5), and the human DC-SIGN gene (UniProt accession no. Q9NNX6) were synthesized by GeneArt (Life Technologies, Regensburg, Germany) and codon optimised for expression in human cells lines. See Appendix **10.1** for a list of GeneArt sequences.

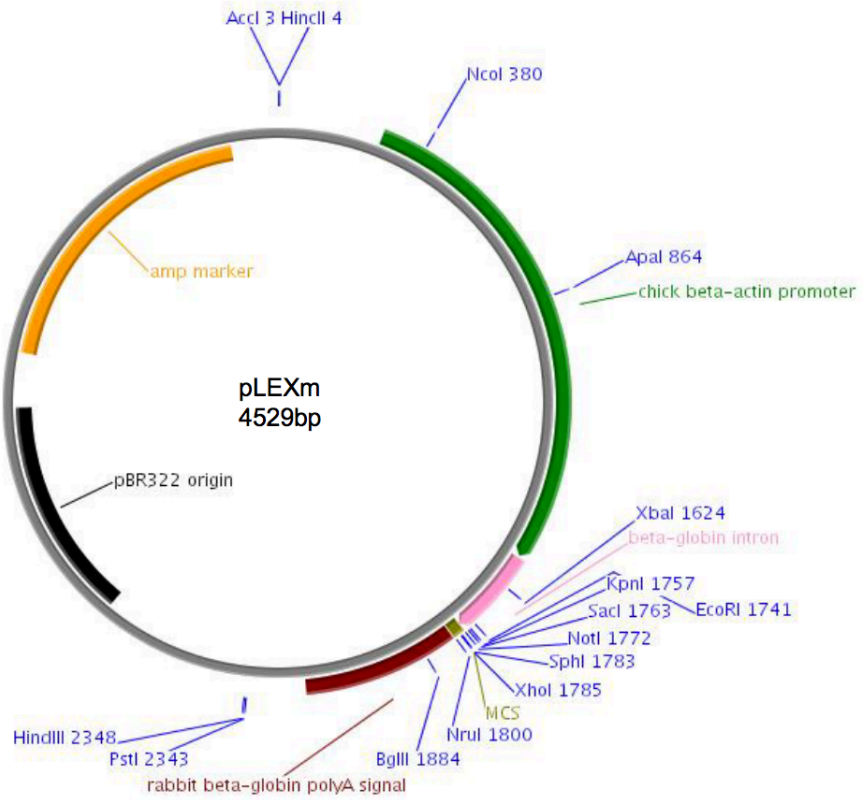


Figure 2.1 The pHLsec vector.
Outline of the pHLsec vector, adopted from Aricescu *et al.* [102]

2.1.2 Primers and construct design

Primers were synthesised by Eurofins Genomics (Ebersberg, Germany), see list of primers in **10.2** of Appendix. Melting temperature (*T*m) of primers was between 58 and 60 ˚C for optimal PCR. Primers were designed with a stabilizing region of three nucleotides, restriction site and the beginning or the end sequence of construct to be cloned. Restriction sites were chosen to match the pHLsec vector such that the forward primer contains an *Age* I site and the reverse primer contains a *Kpn*I site.

2.1.3 Polymerase chain reaction

Polymerase chain reaction was performed with the Pyrobest DNA Polymerase kit (Takara Co. Ltd., Madison, USA). The reaction mixture contained the following components: oligonucleotide forward primer (200 nM), oligonucleotide reverse primer (200 nM), cDNA

template (100-200 pg), dNTP pre-mixture (0.4 mM), Pyrobest buffer II, distilled H_20 and Pyrobest Polymerase (5.0 U). Final volume was 50 µl per reaction. DNA amplification took place in a thermal cycler using the following scheme (Figure **2.2**): A) Initial hot-start of 96 ˚C for 2 min, B) 30-32 cycles through 30 s at 96 ˚C for denaturing, 30 s at 55-60 ˚C for annealing (depending on primers) and 45 s of 72 ˚C for extension, C) Denaturing at 96 ˚C for 5 mins and a final 4 ˚C holding step. PCR products were visualised using agarose gel electrophoresis.

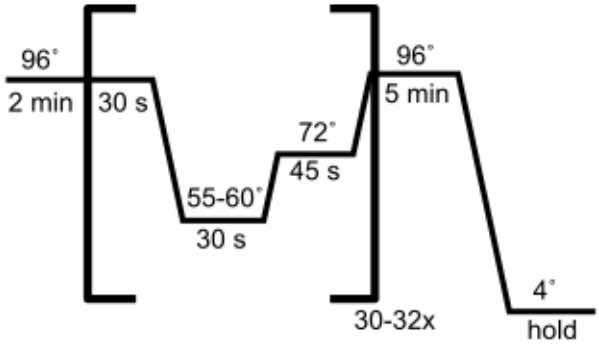


Figure 2.2 PCR cycle outline

2.1.4 Agarose gel electrophoresis

Agarose gels were composed of 1.2 % agarose in a TAE (Tris acetate EDTA) buffer and 1x SYBR Safe DNA gel stain (Invitrogen, Paisley, UK). DNA samples were mixed with a 6x Gel Loading Dye (New England Biolabs, Massachusetts, USA) and Hyperladder 1 (Bioline, London, UK) was run alongside samples. Gels were run at 100 V for 35-45 minutes and imaged using the ChemiDoc XRS+ system (Bio-Rad Laboratories, California, USA).

2.1.5 Restriction digestion

PCR products were gel purified with the QIAquick PCR Purification Kit (Qiagen, Crawley, UK) prior to enzymatic digestion. Vector and fragments were digested separately using

*Kpn*I-HF and *Age*I-HF endonucleoases (New England Biolabs, Massachusetts, USA) with recommended procedures and the CutSmart buffer as provided with enzymes. Following digestion, products were re-purified using the QIAquick PCR Purification Kit (Qiagen, Crawley, UK) before ligation.

2.1.6 Ligation of plasmid and insert

Enzymatic ligation of digested vector and insert was carried out using the NEB Quick Ligation Kit (New England Biolabs, Massachusetts, USA). An access of insert (approximately 70 ng) and an amount of vector (approximately 50 ng) were incubated with 1 µl of Quick Ligase and 10 µl of Quick Ligase buffer for 15 min at RT.

2.1.7 Transformation of plasmid

A 1:10 volumetric ratio of newly ligated plasmid and competent library efficiency DH5-α *E. coli* cells (Invitrogen, Paisley, UK) was prepared on ice, typically 2 µl : 20 µl. The mixture was allowed to settle for 30 mins on ice before a 45 s heat shock at 43 ˚C, followed by 1 h incubation with agitation at 37 ˚C. Cells were plated out on LB-agar plates containing 100 µg mL^{-1} ampicillin for selection of clones containing plasmid and incubated overnight at 37 ˚C.

2.1.8 Plasmid isolation and analysis

Colonies grown on LB-agar plates were picked and used to inoculate 5 ml of LB supplemented with 100 µg mL^{-1} ampicillin and grown at 37 ˚C with agitation overnight. Resulting cell culture was pelleted and plasmids from cells were isolated using QIAprep Spin Miniprep kit (Qiagen, Crawley, UK). Plasmids were analysed by restriction analysis, where plasmids were enzymatically digested with endonucleases *Kpn*I and *Age*I (see **2.1.5**) and run on an agarose gel (see **2.1.4**) to check for the presence of plasmid and inserts of the correct size. Inserts were further analysed by DNA Sanger sequencing provided commercially by Source BioScience (Nottingham, UK).

2.1.9 Large scale DNA production

Subcloning Efficiency BH5α Competent *E.coli* cells (Life Technologies, UK) were transformed with the plasmid of interest (see **2.1.7**). 2 L of LB containing 100 μg mL^{-1} ampicillin was inoculated with transformed colonies and cells were grown for 16-20 hours at 37 ˚C in a shaking incubator. Plasmids were isolated from the cells using the Plasmid Giga kit (Qiagen, UK) following the standard protocol. The resulting DNA pellet was resuspended in 20 mM Tris-HCl pH 8.0 and stored at -20 ˚C.

2.1.10 Site-directed mutagenesis

Single and double mutants of SFTSV Gc (construct V996) were generated by a PCR-driven overlap extension strategy [103]. In strategy no.1 (Figure **2.3**) two DNA halves of a gene were generated by PCR as a first step. They overlapped at the site of the mutation and both contained the desired mutation within the overlapping region. In the second step, the two fragments were 'annealed' together to form a single fragment, i.e. the gene of interest containing the mutation. The first PCR step was carried out as described above (**2.1.3**, with the addition of 2-4% DMSO to lower polymerase specificity) resulting in fragments AB and CD, where A, B, C and D denote the primers used (see Appendix **10.3**). Primers B and C overlap with each other and contain the desired mutation while primers A and D are specific for the start and the end of the gene respectively. Products AB and CD were gel purified (see **2.1.4**) and used as primers (final amount of product per reaction ~100 ng) for a second PCR step along with primers A and D. Final product AD was gel purified, enzymatically digested (see **2.1.5**) and ligated into the pHLsec vector (see **2.1.6**). Competent cells were transformed with the new plasmid (see **2.1.7**), plasmid was isolated and analysed (see **2.1.8**).

In a second strategy, a primer, 'O' was introduced to help with the second PCR step in 'gluing' together fragments from the first PCR reaction. In these cases, primers B and C do not overlap but primer O overlaps with both B and C.

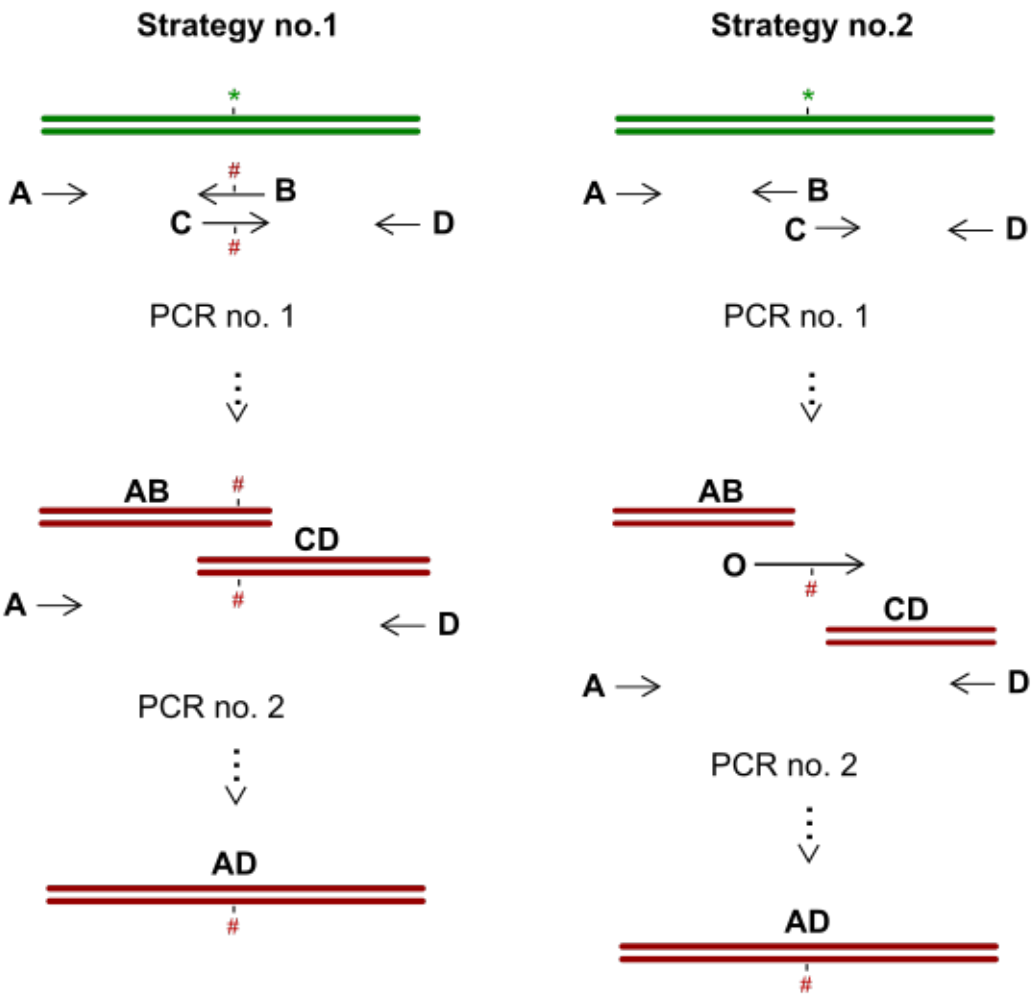


Figure 2.3 Site-directed mutagenesis cloning strategies.
Two related strategies were used for making site-directed mutants. The green gene is the template DNA containing gene of interest with the wild-type base (denoted here by *). In the first strategy, the mutation (denoted here by #) is introduced within primers B and C in the first PCR step. In the second strategy, the mutation is introduced on primer O in the second PCR step.

2.1.11 Cloning of monoclonal antibody sequences from hybridoma cells

Hybridoma cells stably expressing a rabbit anti-RVFV Gn antibody were commercially obtained from Epitomics Inc (Burlingame, California, USA). Cells from 49-6, 55-3 and 93-

4 clones were grown in RPMI media (Sigma-Aldrich, UK) supplemented with 10% fetal calf serum (FCS, Sigma Aldrich, UK) at 37 °C in the presence of 5% CO_2. Poly-A mRNA was extracted from ~190,000 cells using the RNeasy kit (Qiagen, Crawley, UK) and reverse transcription was performed using the SuperScript III First Strand kit (Invitrogen, UK). Rabbit IgG variable regions were amplified by PCR (see **2.1.3**) using primers (Appendix **10.4**) kindly provided by Dr. Katie Doores (King's College London, UK). The resulting PCR products were sequenced commercially by Source Bioscience (Nottingham, UK) using the same primers as were used for the PCR amplification. For a list of rescued sequences, see Appendix **10.6.**

2.1.12 Creation of a RabFab vector

To express Rabbit IgG Fab in a mammalian expression system, the pHLsec vector was modified to include a rabbit constant Fab region between the *Kpn*I and *Xho*I cloning sites. (Figure **2.4**). Only the heavy chain vector contains a His-tag for purification. In this vector, any variable region can be inserted between the *Age*I and *Kpn*I sited. Two such vectors were designed for both a heavy chain and a kappa chain using the constant region from a previously crystallised rabbit Fab [104] and the vector was named RabFab. A GeneArt DNA String containing the variable region from hybridoma 55-3 and the afore mentioned constant region was synthesised commercially (Life Technologies, Regensburg, Germany) and inserted into the pHLsec vector between the *Kpn*I and *Xho*I cloning sites (see **2.1.5**, **2.1.6** and **2.1.7**).

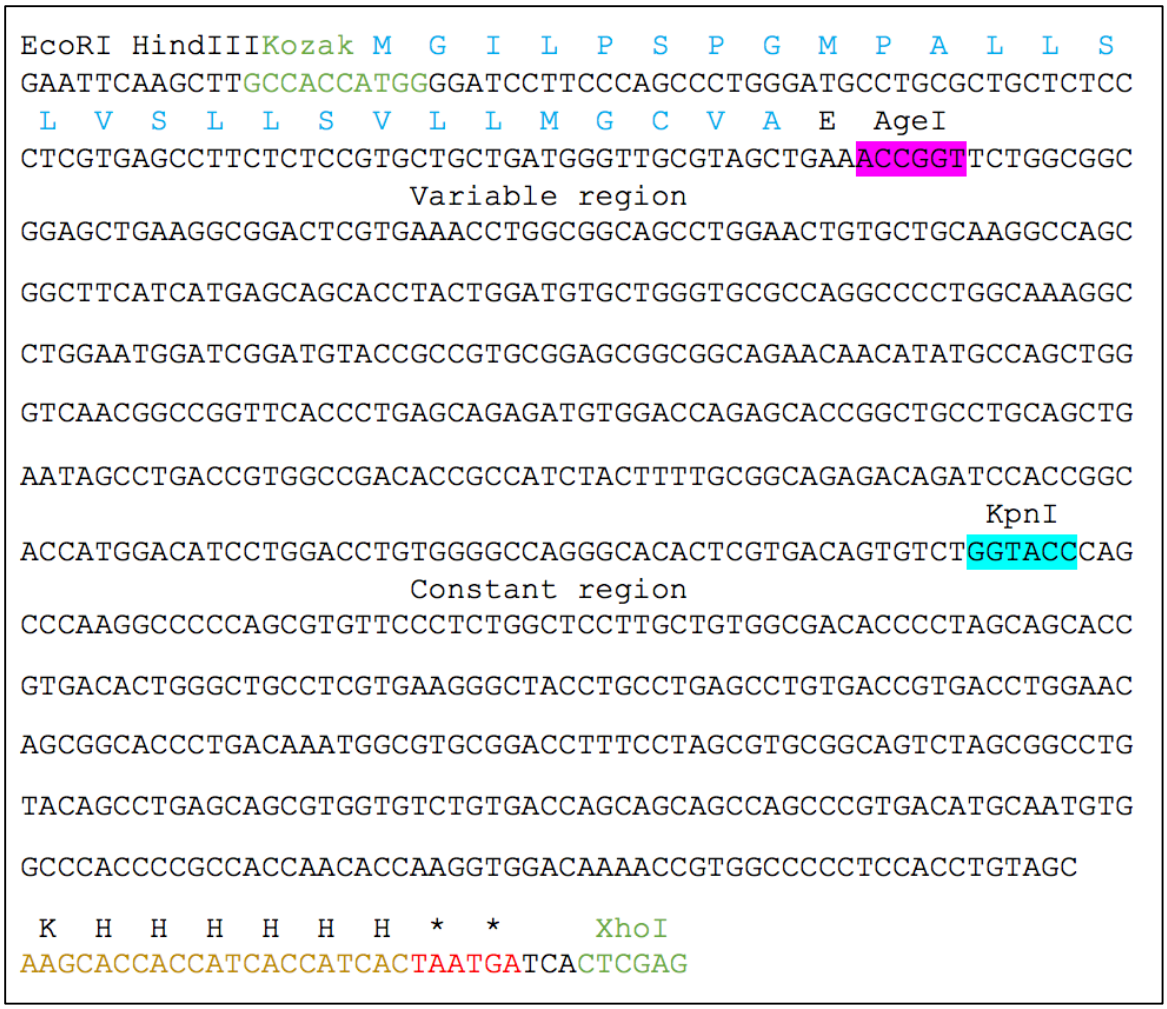


Figure 2.4 RabFab heavy chain vector.
RabFab Heavy chain vector based on the pHLsec vector [102] with the constant Fab region from a crystallised construct [104] and the variable region from clone 55-3. The variable region is flanked by *Age*I and *Kpn*I restriction sites which allows for easy cloning of new variable regions. The constant region is between *Kpn*I and *Xho*I restriction sites. Other parts of the multiple cloning site are the same as in pHLsec. The kappa chain vector only differs from the heavy chain vector in lacking a His-tag at the end of the constant region.

2.1.13 Sub-cloning into the RabFab vector using InFusion

Antibody sequences isolated from hybridoma cells were PCR amplified (see **2.1.3**) with primers designed specifically for InFusion cloning (Appendix **10.5**). These contain 15 bp which are complimentary to the vector backbone for the recombinase reaction. The RabFab vector was digested with restriction enzymes *Age*I and *Kpn*I (see **2.1.5**). A lyophilized enzyme mix from the In-Fusion® Dry-Down PCR Cloning Kit (Clonetech) was mixed with

~30 ng of vector and ~20 ng of insert in a final volume of 10 µL. The mix was incubated first at 37 °C for 15 min followed by another 15 min incubation at 50 °C. The resulting plasmids were used for bacterial transformation (see **2.1.7**) and analysed (see **2.1.8**). For the sequences of the RabFab constructs see Appendix **10.7**.

2.2 Mammalian protein expression and purification

2.2.1 Tissue culture

HEK 293 T cells were propagated in Dublecco's modified Eagle media (DMEM, Sigma Aldrich, UK) supplemented with 10 % FCS, L-glutamine and non-essential amino acids (Life technologies, UK). Cells were grown at 37 ˚C in the presence of 5 % CO_2.

2.2.2 Small-scale transient protein expression

Protein expression was performed at small-scale to estimate expression levels prior to large-scale protein expression, as described before [102]. HEK 293 T cells were grown to 80-90% confluency in a 6-well plate, with each well containing 1 ml volume of media. 5 µg of plasmid DNA containing the construct of interested was mixed with 250 µL of DMEM and incubated for 5 min at RT. Separately, 10 µL of Lipofectamine 2000 (Thermo Fisher Scientific, UK) was mixed with 250 µL of DMEM and incubated for 5 min at RT, before mixing with the DMEM containing the plasmid DNA. The Lipofectamine-plasmid mixture was incubated at room temperature (RT) for 20 min. During the incubation, the media of the cells in the 6-well plate was replaced by fresh DMEM containing 2 % FCS. The cells were transfected by the addition of the Lipofectamine-plasmid mixture to the cells media. In some cases, expression was performed in a 24 well plate using ¼ of above-listed reagents. Cells were incubated for 2-3 days at 37 ˚C in the presence of 5 % CO_2. An aliquot of the cell supernatant was collected and subject to Western blotting.

2.2.3 SDS-PAGE

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed to assess protein purity. Each sample was mixed 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) and incubated at RT for 5 min. NuPAGE 4-12 % Mini poly-acrylamide gels (Thermo Fisher Scientific, UK) were placed in the NuPAGE Mini Gel tank (Thermo Fisher Scientific, UK) filled with MES SDS Running buffer (Thermo Fisher Scientific, UK). Samples were pipetted into the wells of the gel along with a protein standard. For SDS-PAGE, 5 µL of BenchMark protein ladder (Thermo Fisher Scientific, UK) were added to one lane and for Western blotting, 1 µL of BenchMark protein ladder and 5 µL of BenchMark Pre-Stained protein ladder (Thermo Fisher Scientific, UK) were added together in a single lane. Samples were run at 180 V for 40-50 min. The gel was removed from the cassette and stained with Simply Blue safe stain (Life Technologies, UK) and imaged using ChemiDoc XRS+ (BioRad, UK).

2.2.4 Western blotting

Western blotting was performed to assess protein expression levels. An SDS-PAGE was first performed (see **2.2.3**). When loading samples to the SDS-PAGE gel, 1 µL of BenchMark protein ladder and 5 µL of BenchMark Pre-Stained protein ladder (Thermo Fisher Scientific, UK) were added together in a single lane. After completion of the SDS-PAGE, the unstained gel was placed onto a nitrocellulose membrane and protein was transferred to the membrane using the wet-blot apparatus of the NuPAGE Mini Gel tank (Thermo Fisher Scientific, UK) with the NuPAGE Transfer buffer (Thermo Fisher Scientific, UK). The blotting was run at 30 V for 60 min. The membrane was blocked for 60 min in PBS containing 5% w/v milk powder. After blocking, a primary mouse PentaHis antibody (Qiagen, Crawley UK) was used to probe for constructs and was incubated for 60

min with the membrane in a 1/1,000 dilution in PBS with 5% milk. Excess antibody was removed by washing the membrane 3x for 5 min in PBS containing 0.5% Tween-20. A secondary anti-mouse IgG Fc antibody conjugated to a horseradish peroxidase (Sigma Aldrich, UK) was incubated with the membrane for 30 min in a 1/2000 dilution in PBS containing 5% milk. Excess antibody and non-specific background were removed by washing membrane 5x for 5 min in PBS containing 0.5% Tween-20. For visualisation, the membrane was incubated with 2 mL of Clarity Western ECL substrate (BioRad, UK) for 1 min prior to imaging with a ChemiDoc XRS+.

2.2.5 Large scale transient protein expression

A similar strategy to the small-scale expression described above was adopted to a roller bottle system. HEK 293T cells were grown to 80-90% confluency in roller bottles with 200 ml of media in each bottle. 0.5 mg of plasmid DNA was used for each roller bottle and transfected with 0.75 mg of polyethylenimine (PEI). Kifunensine (Cayman Chemicals, Michigan USA) at 5 μM was added to the roller bottles during expression. Kifunensine is a mannosidose I inhibitor and limits N-linked glycans to $Man_9GlcNAc_2$ [105]. Cells were incubated for 4-6 days for optimum protein yield. Cell supernatant was collected and centrifuged at 3000g to remove cell debris and filtered through a 0.22 μm filter prior to purification.

2.2.6 Protein purification

Cell supernatant from expressing cells was first dialysed against purification buffer (10 mM Tris-HCL pH 8.0, 150 mM NaCl). Either a QuixStand or an Äkta Flux (both GE Healthcare, UK) were used for the dialysis, equipped with a hollow fibre or a membrane cassette with a 10 kDa molecular weight cut-off. The supernatant was concentrated down to ~200 ml before ~1500 ml of buffer was dial-filtered against the supernatant. The dialysed supernatant was collected and filtered through a 0.22 μm filter.

His-tagged protein was isolated from the dialysed supernatant with a nickel-containing HisTrap immobilised metal affinity chromatography (IMAC) column (GE Healthcare, UK). Protein-containing media was flown over the column to bind the protein to the column via the his-tag. The column was then washed with several column volumes of purification buffer. Impurities were washed off the column by flowing purification buffer containing imidazole in a step-wise manner, such that 15-20 column volumes each of 10 mM, 20 mM, 30 mM, 40 mM and 50 mM of imidazole were flown over the column. To elute the protein of interest, 250 mM imidazole containing purification buffer was flown over the column.

If glycan removal was desired, the protein was partially deglycosylated to a single N-acetylglucosamine at this stage and achieved by the enzymatic cleavage of a high-mannose-specific endoglycosidase F1 (endo F1), kindly provided by Dr Ilona Rissanen (Div Structural Biology, Oxford University). The protein of interest was incubated at RT overnight with endo F1 in an approximate 1:50 mass ratio of endo F1 to glycoprotein.

The final step in purification was size exclusion chromatography. Protein was concentrated using a 10 kDa MWCO Amicon protein concentrator (Millipore, Watford, UK) by centrifugation at 2,500 g at 4 ˚C. A Superdex 200 10/30 column was equilibrated in purification buffer and concentrated protein was injected onto the column. Eluted fractions were examined on an SDS-PAGE gel (see **2.2.3**).

2.3 X-ray crystallography

2.3.1 Crystallisation and cryo-cooling

Protein crystallisation was achieved by high-throughput robotic crystallisation screening using the sitting-drop vapour method in a 96-well plate format provided in-house [106]. Protein was concentrated using a 10 kDa MWCO Amicon protein concentrator (Millipore,

Watford, UK) by centrifugation at 2,500 g at 4 ˚C. The protein was filtered through a 0.22 μm filter prior to high-throughput screening. Screens were set up in a 1:1 ratio of 100 nl of precipitant solution to protein. Plates were stored at 21 ˚C and imaged at regular time intervals using the automated storage and imaging system RI 1000 (Formulatrix, Bedford, MA). To preserve crystals during X-ray data collection, the crystals were cryo-cooled. Crystals were transferred to a solution containing a cryo-protectant, or the drop containing the crystal was supplemented with cryo-protectant solution, and crystals were subsequently flash-frozen in liquid nitrogen. A list of crystals described in this thesis can be found in **Table 2.1**.

Table 2.1 List of protein crystals

Protein construct	Protein concentration	Precipitant solution	Time to grow (days)	Cryo protection
SFTSV Gc V996	3.5 mg/ml	45% pentaerythriotol 426, 0.1 M sodium acetate at pH 4.6[a]	5	Precipitant
RVFV Gn FL1	8 mg/ml	20% w/v PEG 6000 and 100mM HEPES pH 7.0[b]	238 - 291	25 % glycerol
RabFab 49-6 RVFV Gn complex	9.5 mg/ml	20 % w/v PEG Monomethyl Ether 5000, 0.1 M bis-tris pH 6.5[c]	132 - 185	25 % glycerol

[a]From the Pentaerythriotol screen (Jena Bioscience, Jena, Germany)
[b]From the Grid Screen PEG 6000 (Hampton Research, California USA)
[c]From the Index HT Screen (Hampton Research, California USA)

2.3.2 X-ray data collection and processing

All X-ray diffraction data was collected at Diamond Light Source (Harwell, UK) synchrotron on PILATUS 6M detectors with crystals at liquid nitrogen temperatures. A crystallographic table summarising data collection and refinement statistics for each crystal is found in relevant chapters. Data was indexed, integrated and scaled using XIA2 [107].

2.3.3 Experimental phasing and molecular replacement

Initial phases for SFTSV Gc V996 and RVFV Gn FL1 were obtained from crystals soaked in platinum derivatives, K_2PtCl_6 and K_2PtCl_6 respectively. The platinum atom served as an anomalous scattered and X-ray data was collected from crystals soaked with the platinum derivatives in a single anomalous wavelength dispersion (SAD) experiment. X-ray data was collected at the L-III absorption edge of platinum (~1.07 Å) and away from the absorption edge (~1.01 Å). Phases from the SAD data were obtained using autoSHARP [108] and the initial model was built with Buccaneer [109], as implemented in autoSHARP. For the RabFab-49-6-RVFV-Gn crystals, phasing was achieved by molecular replacement with Phaser [110] using the crystal structure of a rabbit monoclonal antibody R56 [104] (PDB no. 4JO1) as the search model.

2.3.4 Model building and refinement

For SFTSV Gc V996, Refmac5 [111] was used for initial rounds of refinement and PHENIX refine [112] was used for later stages of refinement with translation-libration-screw-rotation (TLS) restraints, but no non-crystallographic symmetry (NCS) restraints. For all other models, PHENIX refine was used with NCS restraints. Manual model building was done in Coot [113]. The final structure was validated using MolProbity [114].

2.3.5 Structure based phylogeny

Eleven class II fusion protein structures were obtained from the Protein Data Bank (PDB) [115]. Dengue virus 2 E (PDB code 1OK8), Semliki forest virus E2 (PDB code 1RER), Tick-borne encephalitis virus E (PDB code 1URZ), Dengue virus 2 E (PDB code 3G7T), Rubella virus E1 (PDB code 4ADI), Saint Louis encephalitis virus E (PDB code 4FG0), *C. elegans* EFF1 (PDB code 4OJD), sever fever with thrombocytopenia syndrome virus Gc (PDB code 5G47), Puumala virus Gc (PDB code 5J9H), Hantaan virus Gc (PDB code 5LJZ), *C. reinhardtii* HAP2 (PDB code 5MF1). Three separate packages were used to

estimate structural phylogeny of these proteins. SHP [116] was run locally and generated a structural alignment matrix. Evolutionary distances were estimated from the matrix using the PHYLIP 3.5 package [117]. MUSTANG 3.2.2 [118] was run locally to generate a structure based alignment. Evolutionary distances were estimated from the alignment using the PhyML 3.0 [119]. The online DALI server [120] was used to align structures to generate a relationship matrix and evolutionary distances based on a Z-score for structure similarity. All data on evolutionary distances were fed into Drawtree [117] to generate an unrooted evolutionary tree.

2.4 Rift Valley fever virus

2.4.1 Tissue culture and virus propagation

Vero cells were kindly provided by Prof Kay Grünewald (Div Structura Biology, University of Oxford). Cells were grown and maintained at 37 ˚C in a 5% CO_2 v/v atmosphere with DMEM (Sigma Aldrich, UK) supplemented with 10% FSC (Life Technologies, UK), non-essential amino acids and L-Glutamine (Life Technologies, UK). Cells were grown to ~70-80% confluency prior to infection with virus. The RVFV clone 13 vaccine strain was kindly provided by Prof Friedemann Weber (Philipps University in Marburg, Germany) and handled at biosafety containment level 3. Cells were infected with a 1 h inoculation at 37 ˚C of a serum-free growth media containing virus at a multiplicity of infection (m.o.i.) of 0.1, after which the inoculum was replaced with fresh serum-free growth media. Cells were incubated for ~43-46 h at 37 ˚C and the resulting virus-containing supernatant was clarified by low speed centrifugation at 3,000 g for 30 mins at 4 ˚C. The clarified supernatant was either stored at -80 ˚C for future inoculation or further purified.

2.4.2 Plaque assay

To estimate virus titer, Vero cells were grown to ~90% confluency in a 24-well plate and infected with a serial dilution of virus sample for 1 h followed by incubation in DMEM supplemented with 2% FCS and 0.8% w/v carboxymethylcellulose (CMC, Sigma Aldrich, UK) for 3 days. The CMC-containing media was then removed and cells washed 2x in phosphate buffered saline (PBS, Sigma Aldrich, UK) before being fixed in PBS containing 4% formaldehyde (Sigma Aldrich, UK) for 20 min at RT. The excess formaldehyde-PBS was subsequently removed, cells washed 2x in PBS and stained in double distilled H_2O containing 10% ethanol and 2% crystal violet for 5-10 min. Cells were finally washed 2x with PBS or until plaques were visible.

2.4.3 Purification of virus for cryo-EM

Native virus was purified by pelleting through a 20% w/v sucrose cushion using ultra centrifugation at 100,000 g (25,000 rpm in a Beckman SW32 rotor) for 2 h at 4 ˚C. Virus pellet was resuspended overnight at 4 ˚C in T20N100 buffer (20 mM Tris-HCl pH 7.4 and 100 mM NaCl). For single-particle cryo-EM, virus was chemically fixed with formaldehyde. Virus supernatant was placed into a Slide-a-Lyzer (Thermo Fisher, UK) dialysis flask with a 10 kDa molecular weight cut-off and dialysed against 10x volume of fixation buffer (20mM borate pH 9.0, 200 mM NaCl and 0.2% v/v formaldehyde (Sigma Aldrich, UK)) for 3-4 h at RT, following further dialysis in fresh 10x volume fixation buffer for 14-16 h. Typically 150 ml of media was dialysed against 1500 ml buffer. To remove excess formaldehyde, the fixed virus supernatant was dialysed against buffer without formaldehyde. The virus-containing supernatant was then stored at 4 ˚C and tested for viable virus by plaque assay (see **2.4.2**) prior to the next step. The supernatant was placed in a dialysis tubing (Size 9, Inf Ida 36/32’’ - 28.6mm, Medicell International, UK) with a 12-14 kDa molecular weight cut-off and concentrated by reverse dialysis. Solid granules

of polyethylene glycol with an average molecular weight of 35 kDa were applied to the membrane until the total volume of liquid reached ~6-8 ml. The concentrated virus was then placed on a 20-60 % linear sucrose gradient (containing T20N100) and centrifuged at 150,000xg (30k rpm in SW32 rotor, Beckman) for 14 h. Gradient was made using a Gradient Master (BioComp Instruments Inc., Canada). Gradient was fractionated using a Piston Gradient Fractionator (BioComp Instruments Inc., Canada) peak fractions were determined using Bradford Reagent (Thermo Fisher Scientific, UK).

2.5 Cryo electron microscopy

2.5.1 Plunge freezing of fixed RVFV particles

Holey carbon coated copper grids (C-flat 2/2; Protochips, Raleigh, NC) were glow discharged in a plasma glow discharger (Harrick, USA) for 15 seconds at high intensity. For gradient purified and fixed RVFV particles, a 3 µl aliquot was applied to a glow discharged grid and excess sucrose was removed by floating the grid over 100 ul of T20N100 buffer prior to vitrification. For pelleted and live RVFV particles, a 3 µl aliquot was applied to glow discharged grids which were vitrified by plunge-freezing using a vitrification apparatus (CryoPlunge 3; Gatan, Pleasanton, CA). Grids were blotted from both sides for 3 s prior to plunging into an ethane/propane liquid mixture. Grids were kept in liquid nitrogen storage prior to imaging.

2.5.2 Single particle cryo electron microscopy

Micrographs for single particle data processing were acquired on a Technai ‘Polara’ electron microscope (FEI, Netherlands) at liquid nitrogen temperature using 300 kV. Sample mounting followed standard procedures as previously described [121], except when handling live virus. Grids containing live virus were loaded onto a Polara multi-specimen rod inside a bespoke safety cabinet designed by Total Containment Oxford

Limited (Oxford, UK) fitted with a grid loading station to comply with containment level 3, other standard procedures for specimen mounting were followed. Micrographs were collected at a calibrated magnification of 37,037x corresponding to a pixel size of 1.35 Å with a -3 to -0.6 μm defocus. An energy filter (GIF Quantum LS, Gatan) with a slit width of 20 eV was applied during data acquisition and micrographs were recorded on a direct electron detector (K2 Summit, Gatan) in electron counting mode. 88 movie frames were collected for each micrograph over 17.2 s at a dose rate of 2.5 $e^{-}s^{-1}pix^{-1}$ resulting in a total dose of ~24 $e^{-}Å^{-2}$.

2.6 Single particle image processing

2.6.1 Icosahedral reconstruction of fixed RVFV particles

Movie frames from micrographs collected for single particle analysis were aligned using MotionCorr [122], and the contrast transfer function and defocus were estimated using CTFFIND3 [123] as implemented in the RELION 1.4 software package [124, 125]. All following steps were carried out in RELION, unless otherwise stated. Micrographs and CTF estimation were inspected by eye and bad quality micrographs were discarded. 4,336 particles were picked manually, normalized, contrast inverted and extracted using a box size of 1024x1024 pixels. Particles were down-sampled by a factor of 2 to a box size of 512x512 pixels. 2D classification was carried out using 25 classes and the best 3 classes showing clear features such as separation between the two leaflets of the viral envelope were chosen. An initial 3D refinement was carried out using fine angular sampling, application of icosahedral symmetry and using a previous structure of RVFV as an initial model (EMD-1550) [90]. The resulting parameter for particle orientation and contributions to the output model were used for 3D classification into 5 classes, but without any alignment (skip_align parameter). 2,995 particles were chosen from 3 classes and subject

to a final 3D refinement (using same parameters as in previous 3D refinement). Post-processing was performed as previously described [126]. An *ad-hoc* B-factor of -800 Å^2 was used for map sharpening and density-threshold based masking of the final model. These were carried out using a spherical internal mask to hide contributions of the viral genome and an external mask to hide contributions of the solvent and neighbouring particles. Gold standard FSC was corrected by high-resolution noise substitution [126], resulting in a final map of 13.3 Å resolution. A summary of data collection and refinement statistics can be found in Table **4.1** and a summary of the work flow is outlined in Figure **2.5**.

2.6.2 Localized reconstruction applied to fixed RVFV particles

The Localized Reconstruction method (http://github.com/OPIC-Oxford/localrec) [127] developed locally within the Huiskonen group was used to refine individual capsomers from RVFV resulting in 4 distinct capsomers: pentons, type 1, type 2 and type 3 hexamers. Additionally, partial signal subtraction was used to remove all but one capsomer density from the images [128, 129]. First, a vector defining the location of each capsomer type was defined using the reconstruction of the full viron in UCSF Chimera [130]. A soft-edged spherical mask was generated to define the capsomer boundaries and used to subtract density corresponding to other capsomers, membrane and genome. Sub-particles, corresponding to the projections of individual capsomers in the subtracted particle images, were extracted. The extracted sub-particles for each capsomer type were processed separately in RELION 1.4 using a starting model generated for each capsomer type with relion_reconstruct [124]. The sub-particles were subjected to 3D classification with 4 classes using C5, C3, C2 or C1 symmetry for pentamers, type 3, 2 and 1 hexamers respectively. Good quality 3D classes were selected by resolution estimates and visual inspection based on the presence of structural features (such as transmembrane helices) and

a final 3D refinement was performed without alignment (skip_align parameter). Final maps were post-processed, as described above (see **2.6.1**) with an *ad-hoc* B-factor of -200 Å^2. A summary of the work flow is outlined in Figure **2.5**.

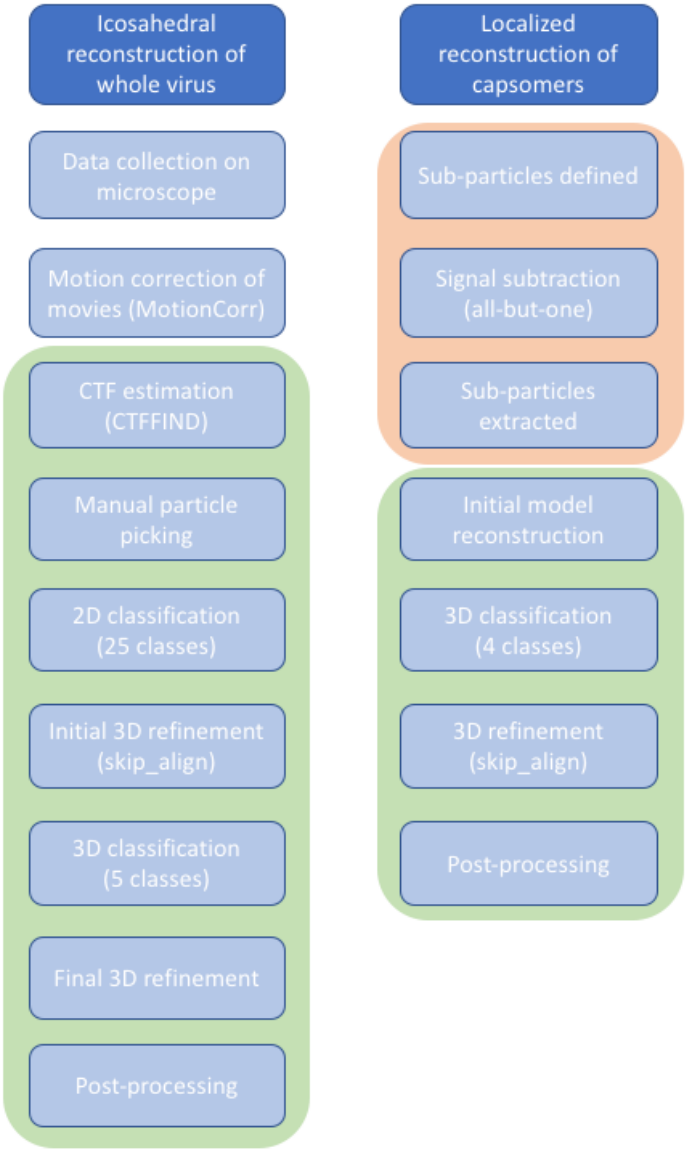


Figure 2.5 Data processing work flow.
Outline of data processing for icosahedral reconstruction of whole virus and localized reconstruction of virus capsomers. Shaded in green are steps carried out in RELION 1.4 [124] and shaded in orange are steps carried out using Localized reconstruction [127].

2.6.3 Molecular Dynamics Flexible Fitting (MDFF) and model building

A pseudo-atomic model of the RVFV was built using Molecular Dynamics Flexible Fitting [131] as implemented in NAMD 2.12 [132] by fitting the crystal structures of RVFV Gn and Gc (PDB no. 4HJ1) [47] into EM reconstructions of the capsomers. Firstly, a single Gc was fit into a segmented density corresponding to Gc from the pentamer map by a simple rigid body fitting tool in UCSF Chimera [130]. The fit was used as a starting point for a 0.7 ns simulation in MDFF. The resulting fit of Gc and the crystal structure of Gn were then rigidly fit into densities of all the capsomers, such that 5 copies of Gc and Gn were fit into the density of the pentamer and 6 copies of Gc and Gn were fit into the densities of each of the different hexamer types. These fits were then used as starting points for 5 ns simulations in MDFF, with symmetry restraints applied. All simulations were carried out at constant temperature and pressure of 310 K and 1 bar in full solvent with 0.1 M NaCl at pH 7.4. A particle mesh Ewald method with a grid spacing of 1 Å was used to compute long-range Coulomb forces in time steps of 1 fs where non-bonded interactions were evaluated every 2 fs and full electrostatics were evaluated every 4 fs. A scaling factor $\zeta = 0.5$ kcal/mol was used for the initial Gc fit and a scaling factor of $\zeta = 0.3$ kcal/mol was used for all other simulations. The geometry of the resulting MDFF fitted models was improved by geometry minimization in PHENIX [112] and validated by Molprobity [114]. To reconstruct the entire virion, an asymmetric unit was generated in UCSF Chimera by rigid body fitting minimal components (1 heterodimer from the pentamer, 2 heterodimers from the type 3 hexamer, 3 heterodimers from the type 2 hexamer and all 6 heterodimers from the type-1 hexamer) and an icosahedron symmetry matrix was generated with the *sym* command in UCSF Chimera.

3. The crystal structure of RVFV Gn

3.1 Foreword

Chapters 3–5 seek to explain the structural arrangements of the glycoprotein shell of RVFV. This chapter introduces the crystal structure of the Gn ectodomain, the missing structural piece of the glycoprotein layer of RVFV. A paper reporting the same structure and the structure of the Gn from SFTSV was published shortly before submission of this thesis (August 2017) [101]. The PDB coordinate files for these structures were not available in time to include a comparative analysis. Hence, this chapter is written largely independent of the two Gn structures. The r.m.s.d. between the published RVFV Gn structure and the structure presented in this chapter is 0.4 Å over 261 Cα-atoms and contains one more C-terminal residue.

3.2 Summary

The soluble ectodomain of RVFV Gn was transiently expressed in mammalian cells, purified and crystallised. The structure of RVFV Gn was solved to 1.6 Å resolution using the SAD method. Approximately one quarter of the expressed protein was missing in the crystal, most likely due to proteolytic cleavage of the glycoprotein during crystallogenesis. Bioinformatics analysis reveals that RVFV Gn forms a unique fold not previous observed. Interestingly, despite exhibiting a novel fold, RVFV Gn shares some architectural and secondary structure features with the Gn of PUUV (family *Hantaviridae*) and E2 protein from CHIV (family *Togaviridae*). RVFV Gn has a similar domain organization to the CHIV E2 and PUUV Gn, which consists of domain A, domain B and a β-ribbon. An equivalent to domain C observed in CHIV E2, was not crystallised but may have been cleaved during crystallogenesis. An analysis of phleboviral glycoprotein sequences reveals

that domain A exhibits the greatest level of sequence diversity and encodes the majority of glycosylation sequons. These observations may be relevant to the phleboviral ultrastruture and suggest that domain A is the domain most exposed to the immune system.

3.3 Expression and purification of the ectodomain of Gn

3.3.1 Construct design and small-scale expression trials

RVFV Gn possesses no detectable sequence homology to any currently known protein structure, where a BLAST search [133] against the Protein Data Bank (PDB) yielded no matches with a significant score or with suitable sequence coverage. Additionally, the protein fold recognition server PHYRE2 did not provide any significant predictions for the structure of Gn [134]. Due to the absence of a suitable structure to define domain boundaries and guide construct design, several constructs were screened for soluble expression, which spanned the entire soluble region of the Gn (Figure **3.1** and for a full list see Appendix **10.2**). Soluble expression in mammalian cells was assessed by Western blot analysis (Figure **3.1**). Several of the full-length constructs exhibited high levels of expression with respect to a positive control and none of the constructs that included a significant C-terminal or N-terminal truncation showed significant expression. For the full-length constructs, however, a pattern emerged, where minor truncations to the C-terminus resulted in increased levels of expression. Minor truncations to the N-terminus only marginally compromised expression. The longest construct which demonstrate high levels of expression, termed ‘FL1’ (residues 154-560), was chosen for large-scale expression.

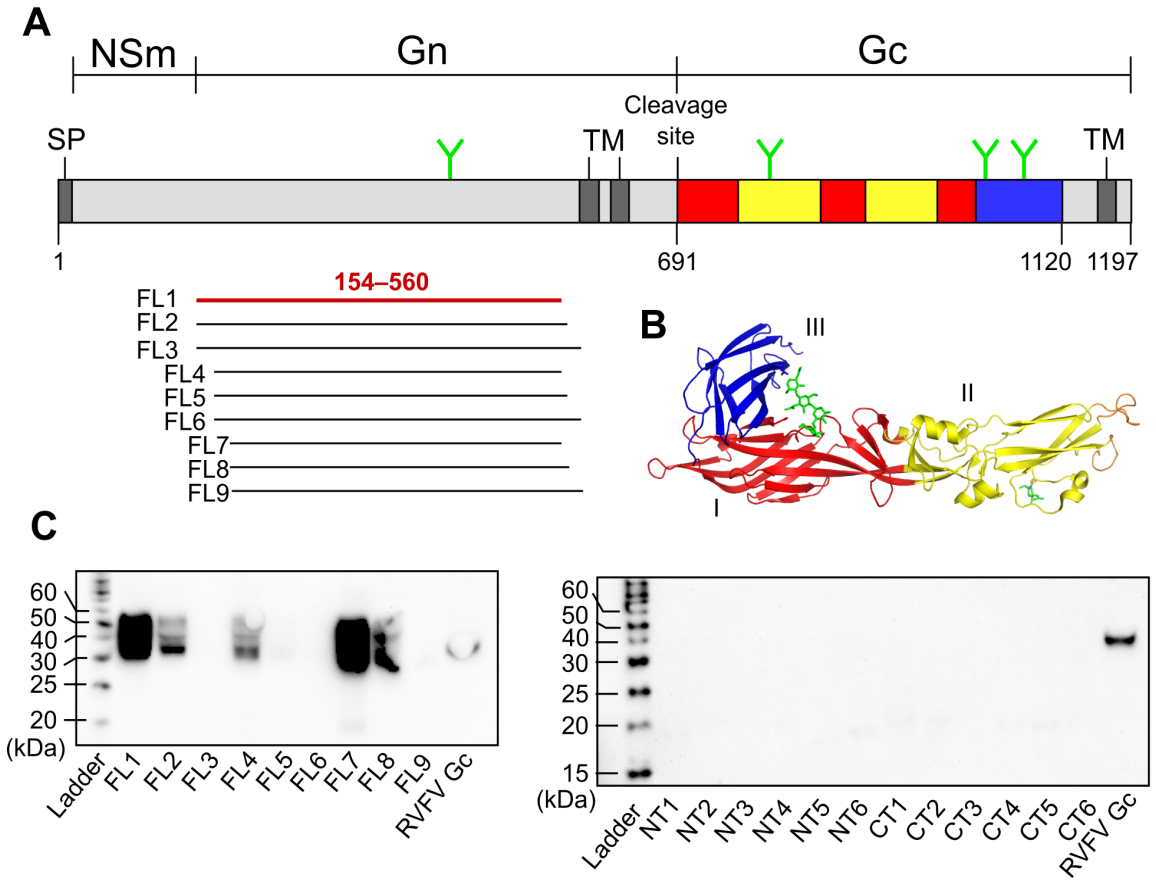


Figure 3.1 Construct design and small-scale expression trial.
A) A domain diagram of the RVFV M-segment shows the arrangements of NSm, Gn and Gc within the genome. The domain organisation of RVFV Gc has been determined previously [47] and is illustrated beneath the gene diagram in panel *B*. N-linked glycosylation sequons are annotated with a green 'Y', transmembrane regions (TM) and the signal peptide (SP) are highlighted. The full-length (FL) constructs of RVFV Gn are shown below the domain diagram and are represented by a line. Details of the N-terminal (NT) and C-terminal (CT) constructs can be found in Appendix **10.2**. B) The crystal structure of RVFV Gc is shown in cartoon and is coloured according to domain: I is red, II is yellow, III is blue and fusion loop is orange. C) A Western blot of cell supernatant from small-scale expression trials of FL constructs (left-hand side), and NT and CT constructs (right-hand side). None of the NT or CT constructs expressed to levels sufficient for large-scale production. However, several of the FL constructs demonstrated robust expression, particularly those that had a shorter C-terminus. FL1 was chosen for large-scale expression. A construct mimicking the crystallised RVFV Gc construct was used as a positive control for expression.

3.3.2 Large scale expression and purification

RVFV Gn FL1 was transiently expressed in HEK 293T cells in a roller bottle format. After 5 days the protein was purified from the supernatant by diafiltration followed by

immobilized nickel affinity purification. The protein was partially deglycosylated with endoglycosidase F1 (endoF1) and subjected to a final SEC purification step (Figure **3.2A**). The yield of purified FL1 was ~3.0 mg/L tissue culture. The protein eluted as a single peak on SEC, corresponding to a putative monomer of ~45 kDa. However, two species were visualized by SDS-PAGE analysis (Figure **3.2B**), one of the expected molecular weight (~45 kDa) and a second species of a lower molecular weight (~30 kDa). At this stage, it was unclear if the second species was a contaminant co-purified with the Gn or a break-down product of the Gn itself.

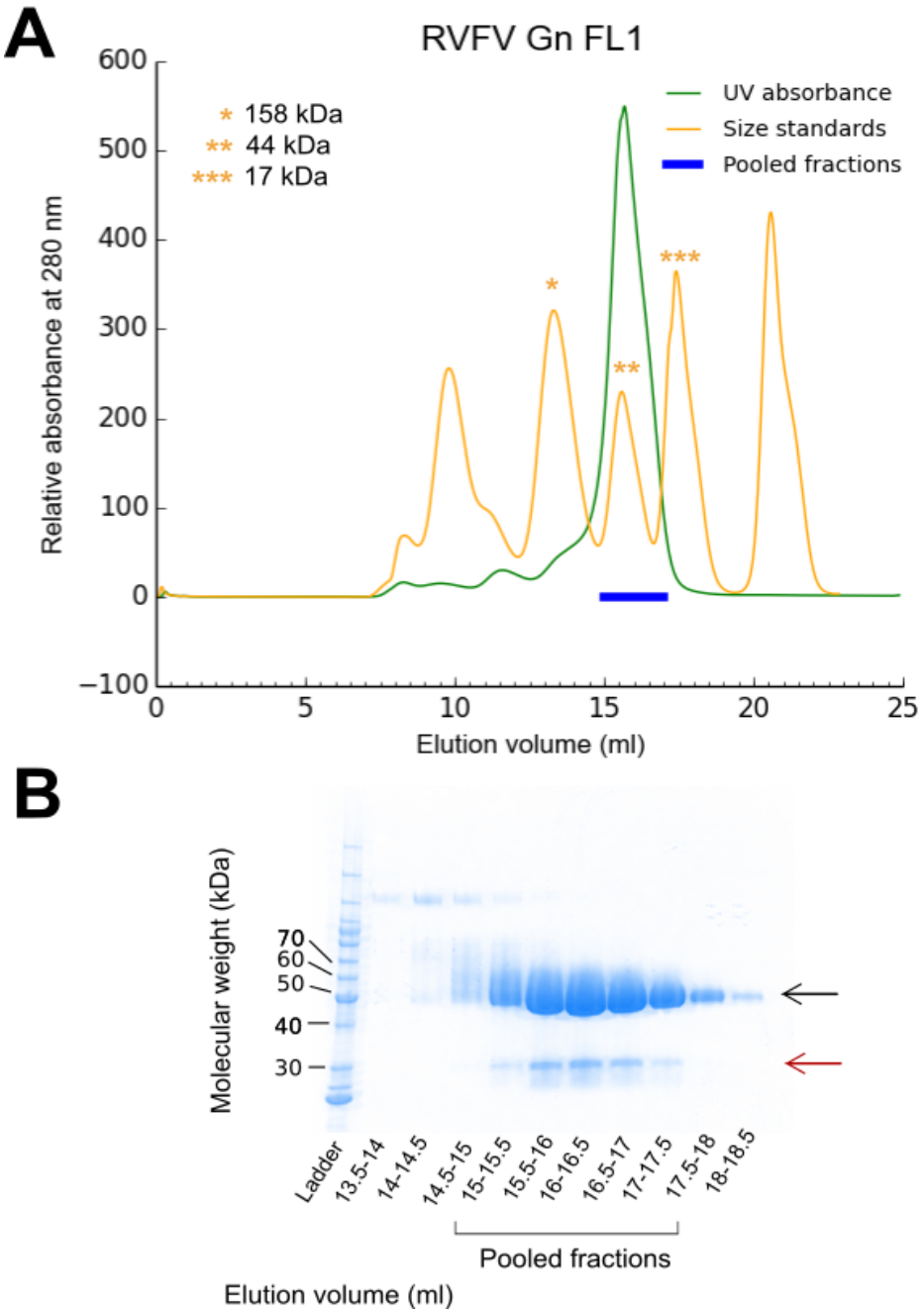


Figure 3.2 RVFV SEC.
A) Elution profile from SEC of RVFV Gn FL1. UV trace at 280 nm is shown in green and pooled fractions in blue. A size standard is shown in orange and the molecular weight of the species forming each peak is shown on the left side of the graph. B) SDS-PAGE of fractions from SEC run with a molecular weight standard (Ladder). The bands highlighted with the black arrow correspond to the expected molecular weight of RVFV Gn FL1. The bands highlighted with the red arrow correspond to an unexpected species in the purification.

3.3.3 Crystallisation and structure determination

Despite the multiple species visualized by SDS-PAGE analysis (Figure **3.2B**), fractions from the single SEC peak were pooled, concentrated to ~8.0 mg/ml, filtered through a 0.2 µm filter and subjected to high-throughput crystallisation screening using the sitting-drop vapour method [106]. RVFV Gn crystals grew in a precipitant containing 20% w/v PEG 6000 and 100mM HEPES pH 7.0. The crystals were broken up into smaller pieces before freezing (Figure **3.3A**).

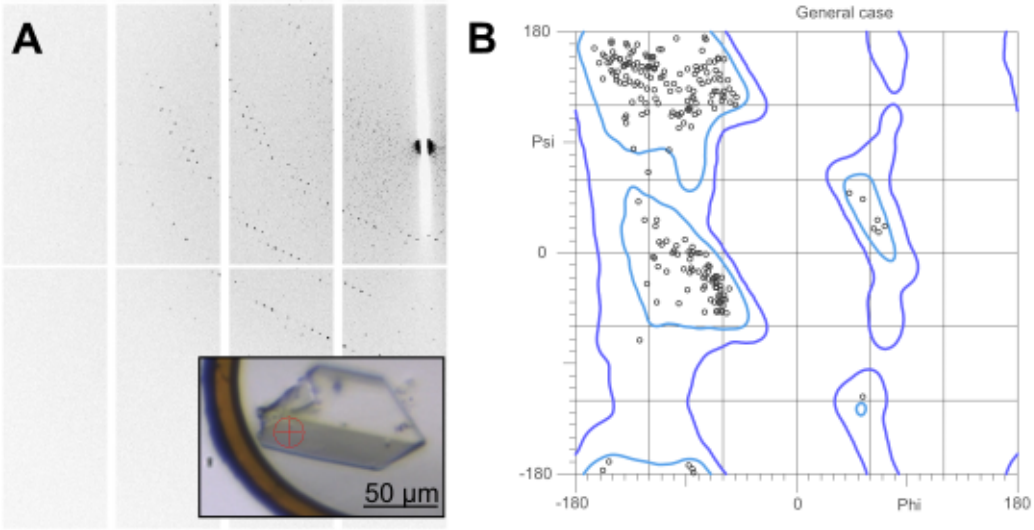


Figure 3.3 Crystal and diffraction.
A) An example of diffraction data collected for RVFV Gn FL1. The inset shows the crystal on the beamline and the aperture (red circle) used. Crystals were tetragonal and grew to ~200 um in the longest dimension and ~80-60 um in the other dimensions. B). A Ramachandran plot of the refined model (generated by MolProbity [114]). All the residues in the structure are either in the allowed or most favoured regions.

As no suitable structural homologue of RVFV Gn was known that could be used for molecular replacement (attempts with PUUV Gn failed), a SAD experimental phasing strategy was adopted. To this end, one crystal was soaked for ~100 min in a cryo-protecting solution saturated with K_2PtCL_4. X-ray data from the native and the heavy metal soaked

crystal were collected at beamline IO3 at Diamond Light Source (Figure **3.3**). A summary of crystallographic and refinement statistics can be found in Table **3.1**.

Table 3.1 Crystallographic data collection and refinement statistics of RVFV Gn FL1

Parameters	K_2PtCl_4 SAD data	Native data
Data collection		
Beamline	I03, DLS	I03, DLS
Resolution range (Å)	67.90–.46 (2.52–2.46)	48.99–16.0 (1.64–1.60)
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, b, c (Å)	37.0, 94.4, 97.8	37.4, 94.2, 98.0
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	1.07146	0.97625
Unique reflections	10,034 (333)	46,607 (3,369)
Completeness (%)	76.7 (35.3)*	99.9 (99.6)
R_{merge} (%)[a]	16.7 (146.0)	8.6 (104.4)
I/σI	35.7 (2.6)	12.35 (2.0)
Avg. redundancy	129.5 (40.6)	9.7 (9.5)
Refinement		
Resolution range		48.99-16.0 (1.64-1.60)
Number of reflections		46,537 (4,565)
R_{work} (%)[b]		16.5
R_{free} (%)[c]		19.9
RMSD		
Bonds (Å)		0.017
Angles (°)		1.52
Molecules per a.s.u.		1
Atoms per a.s.u		
(protein/water)		2,331/336
Average *B*-factors (Å^2)		
(protein/water)		34.3/40.5
Ramachandran plot (%)		
Most favoured region		99.0
Allowed region		1.0
Outliers		0.0

Numbers in parentheses refer to the relevant outer resolution shell.
r.m.s.d.: root mean square deviation from ideal geometry.
[a]$R_{merge} = \Sigma_{hkl} \Sigma_i |I(hkl;i) - \langle I(hkl) \rangle| / \Sigma_{hkl} \Sigma_i I(hkl;i)$, where $I(hkl;i)$ is the intensity of an individual measurement and $\langle I(hkl) \rangle$ is the average intensity from multiple observations.
[b]$R_{factor} = \Sigma_{hkl} ||F_{obs}| - k|F_{calc}|| / \Sigma_{hkl} |F_{obs}|$
[c]R_{free} equals the *R*–factor as calculated above but using against 5% of the data removed prior to refinement.
*Completeness at 3.5 Å was 98 %.

3.4 Structural analysis

3.4.1 Structural overview

A single molecule of RVFV Gn was observed in the asymmetric unit (Figure **3.4**). RVFV Gn resembles the shape of a triangular prism composed of two flat surfaces of 70 x 50 Å and a thickness of 30 Å (Figure **3.4**). Of the 407 amino acids in the original construct (residues 154–560 of the M-segment), 92 C-terminal amino acids were not visible in the crystal structure. Analysis of the electron density and crystal packing around the C-terminus shows that the missing residues were likely cleaved during crystallogenesis. This may explain the presence of the second species observed on the SDS-PAGE gel of the SEC fractions (Figure **3.2B**), which may have formed between the SEC run and the SDS-PAGE analysis. This species appears to migrate around 30 kDa, which matches the size of the crystallised fragment. A requirement for proteolysis may also explain the length of time required for the protein to crystallise (292 days).

Although RVFV Gn is predominantly β-stranded, alpha-helical secondary structure is also observed throughout the glycoprotein. The protein contains fifteen β-sheets, five α-helices and four 3_{10}-helices. Interestingly, analysis with the DALI server [120], a program that detects structural similarity across the Protein Data Bank, failed to detect structural homologues that resemble RVFV Gn. These data indicate that RVFV Gn likely adopts a novel fold.

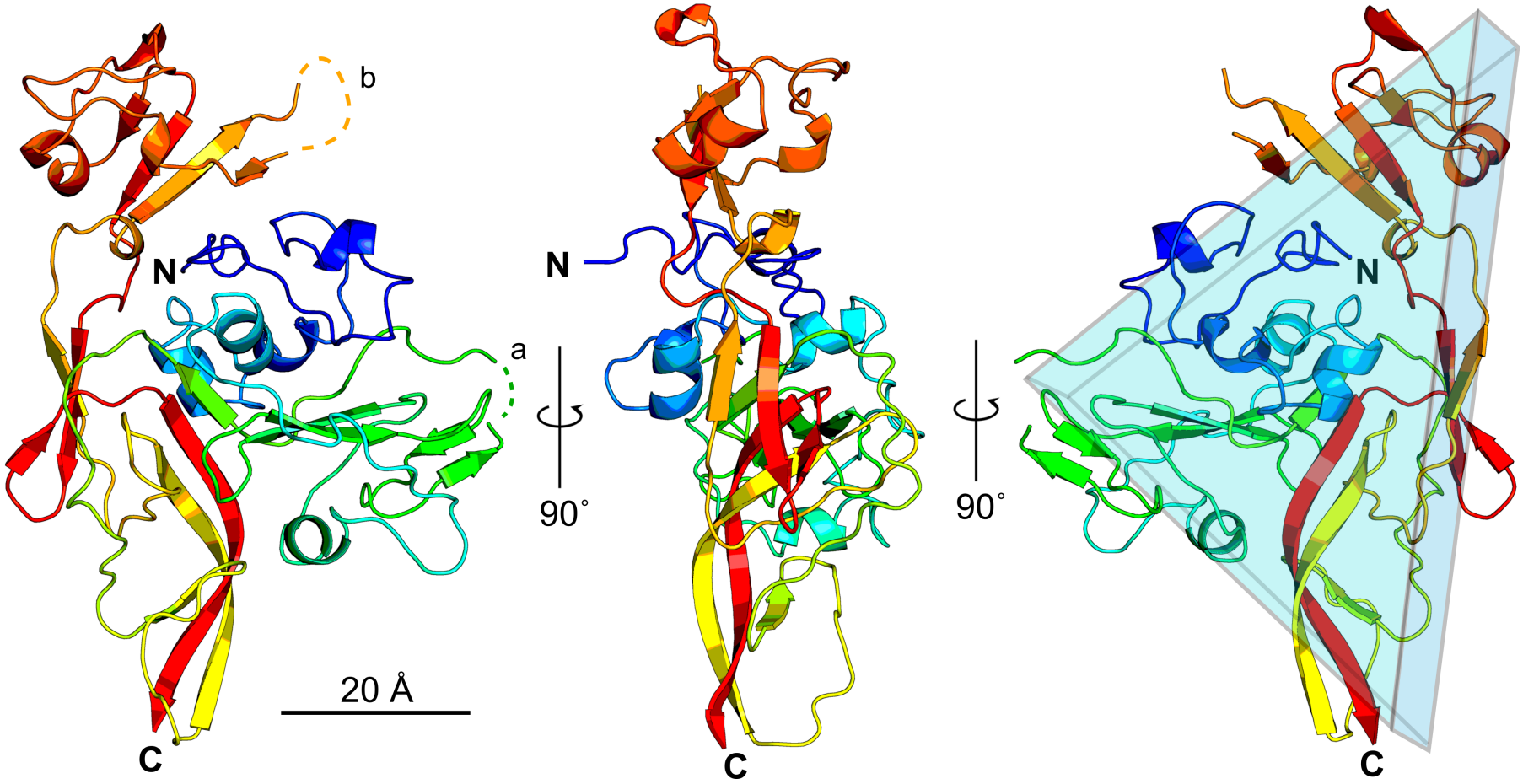


Figure 3.4 Crystal structure of the RVFV Gn ectodomain.
RVFV Gn protein is shown in cartoon representation and coloured as a rainbow ramped from blue (N-terminus) to red (C-terminus). The triangular prism shape of the Gn is illustrated in the right-hand panel. 92 amino acids of the C-terminus are missing from the original expressed construct. Two loops are unstructured, shown on the left-hand side: a) residues 288-289 b) residues 380-392.

3.4.2 Searching for structural homologues

A recent review on the envelope glycoproteins of bunyaviruses [82] highlighted the structural similarity between the Gn protein of Puumala virus (PUUV, negative-sense ssRNA, family *Hantaviridae*, order *Bunyavirales*) and E2 protein from the Chikungunya virus (CHIV, positive-sense ssRNA, family *Togaviridae*, order unassigned). Although RVFV is also a part of the *Bunyavirales* order, RVFV Gn and PUUV Gn exhibit little obvious structural similarities. However, a topology diagram comparison of the RVFV Gn, PUUV Gn, and CHIV E2 reveals a similar domain organisation and some common secondary structural elements, which suggests that these three proteins may have arisen from the same ancestral glycoprotein. (Figure **3.5**).

The E2 protein from alphaviruses has been proposed to act as a fusion chaperone [135], which helps to stabilise the E1 protein (a class II fusion protein) prior to fusion. The CHIV E2 ectodomain crystal structure is primarily composed of β-elements, which are divided into four domains: A, B, C and the β-ribbon (Figure **3.5A**). The β-ribbon domain is highly distinctive and organises the overall structure by linking the other domains together. A β-ribbon element is also present in RVFV Gn, which similarly organises the structure.

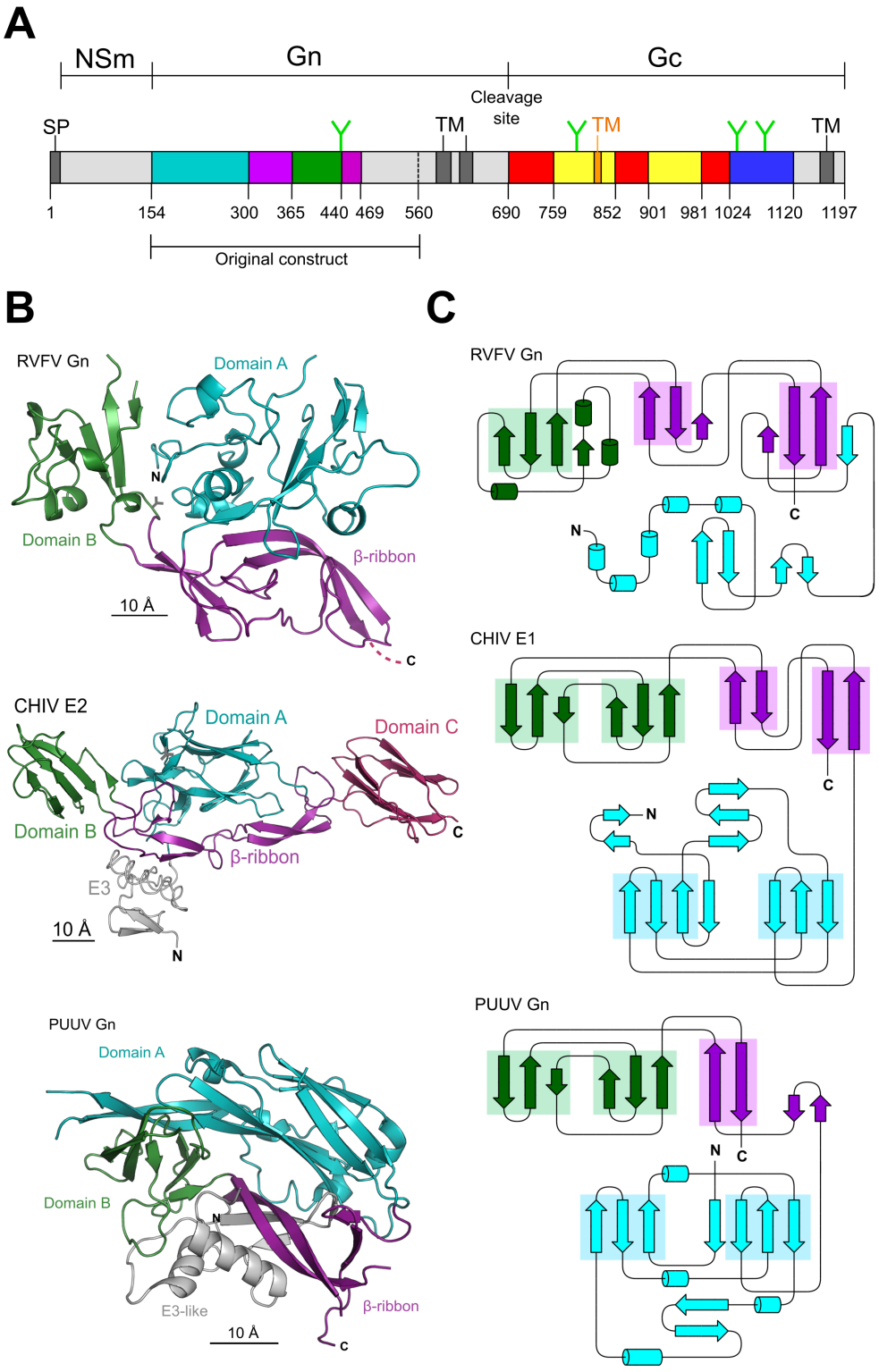


Figure 3.5 Structural homology of class II fusion chaperones.
A) A domain diagram of the RVFV M-segment as in Figure **3.1** with the addition of Gn coloured according to domains as in panel *B*. B) The crystal structures of RVFV Gn, CHIV E2 (PDB code 4N30) [78] and PUUV Gn (PDB code 5FXU) [81]. The structures are presented in cartoon and coloured according to the domain organisation: A is in teal, B is in forest green, C is in dark pink, the β-ribbon is in deep purple and E3 (or E3-like) is in

light grey. C) Topology diagrams of the structures from B. The diagrams are coloured by domain as in panel *B.* Areas of similarity are highlighted by a coloured block behind each area of similarity.

Domain A in RVFV Gn and CHIV E2 do not share any obvious secondary structure elements and have very different tertiary structures. Domain B on the other hand, has some shared elements in the two viruses. In E2, domain A is a six stranded β-sandwich, but in Gn one half of the sandwich has been replaced by α-helical elements, resulting in Gn becoming more of a β-slice with α-helical toppings. Additionally, given the similarities between CHIV E2 and RVFV Gn, it also seems likely that the ~90 amino acids missing from the C-terminus of RVFV Gn may resemble the ~80 amino acid domain C of CHIV E2.

When PUUV Gn is included in the comparison, it emerges that the PUUV Gn and CHIV E2 have similarities located in different regions of the proteins compared to the similarities between the RVFV Gn and the CHIV E2. Domain A has the same β-sandwich core in both PUUV Gn and CHIV E2, and domain B in PUUV Gn is also a six stranded β-sandwich. These proteins differ at the β-ribbon, which has been reduced in PUUV Gn. Overall, it appears that PUUV Gn and RVFV Gn are both similar to CHIV E2 in their own ways, but interestingly the two Gn's do not appear to be similar to each other.

3.5 Diversity of N-linked glycosylations and the primary sequence of the phleboviral Gn

Host cell entry of RVFV and other phleboviruses has been shown to be glycan-mediated [37, 100]. The C-type lectin, DC-SIGN, binds to oligo-mannose type sugars on the envelope glycoproteins of RVFV and SFTSV, which allows for endocytosis of the virion [37, 100]. Only one N-linked glycosylation site, Asn438, is predicted (sequon NXT/S,

where X ≠ P) to be present on the crystallized Gn and analysis of electron density around Asn438 shows no discernible density for a glycan (Figure **3.6**) despite the Asn438 side chain density being well resolved. The Gn was expressed in the presence of kifunensine and the resulting oligomannose was not cleaved by F1 endoglycosidase prior to crystallisation. The Asn438 side chain points into a solvent-filled space in the crystal which may allow the glycan some flexibility, if present.

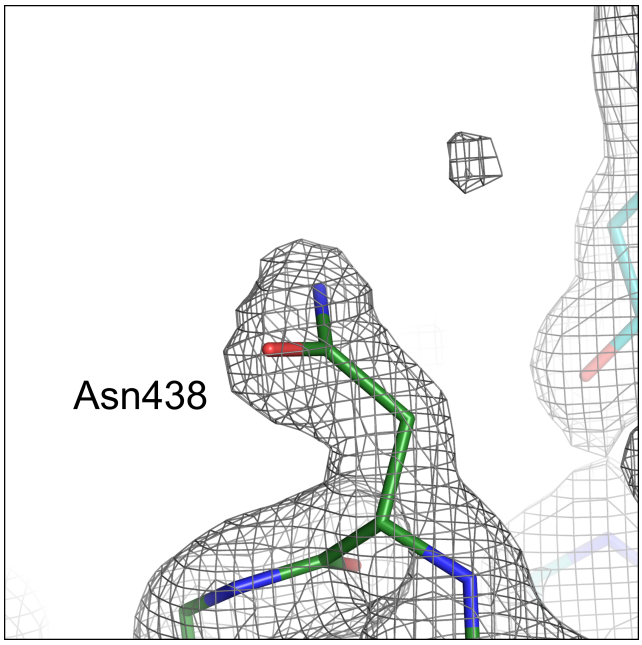


Figure 3.6 Electron density at Asn438 shows no evidence for N-linked glycosylation. Only one N-linked glycosylation is predicted on the RVFV Gn at Asn438. The grey mesh is a *2Fo-Fc* map at 1.0 σ and the model is shown in sticks. The density for the side chain is clear but there is no evidence in the electron density for a glycan.

The number of N-linked glycosylation sequons on the Gn glycoproteins of other phleboviruses is greater in most cases, with phleboviruses such as UUKV encoding five sequons. Sequence alignment of phleboviral Gn sequences with RVFV Gn reveals a low level (<20%) of sequence conservation (Appendix **10.10**). While the level of sequence conservation at the N-terminal domain A is particularly poor, it becomes elevated at the beginning of Domain B (Figure **3.7B**). If the alignment is split in two parts, the sequence identity at residues 154-295 (domain A) is between 12 and 32%, while the sequence

identity at residues 296-689 is 19-38%. This shows that domain A is much less conserved than the β-ribbon, domains B and the putative domain C.

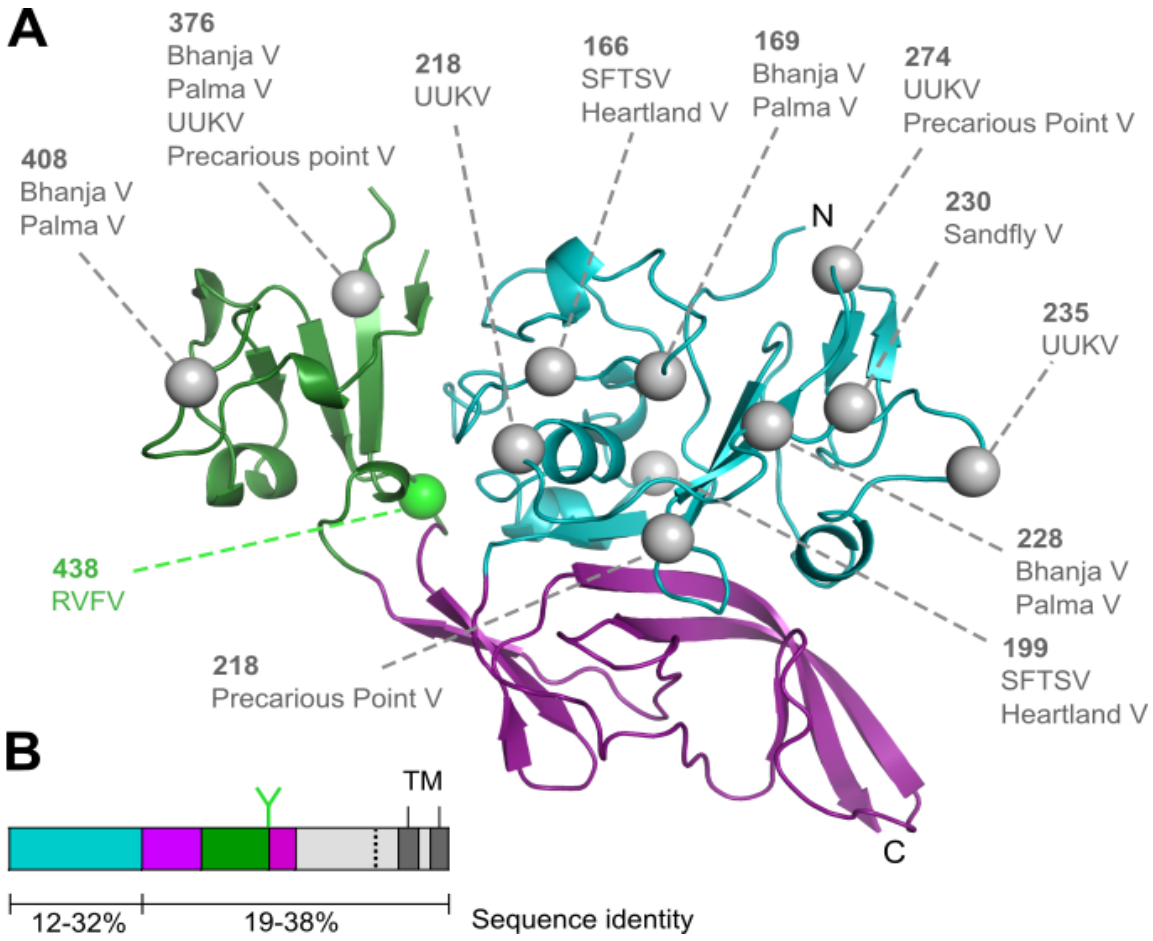


Figure 3.7 Phleboviral N-linked glycosylation and sequence conservation of the Gn. A) The location of N-linked glycosylation sequons are mapped as spheres onto the structure of the RVFV Gn ectodomain (as determined in Appendix **10.10**). Green spheres correspond to RVFV glycans and grey for glycans from other viruses. B) Domain diagram of the RVFV Gn, coloured according to domain. Regions in grey were not crystallised. The dashed line indicates the end of the original FL1 construct. The green Y shows the N-linked glycosylation site and TM highlights the transmembrane regions. Sequence identity was calculated from the alignment of the representative phleboviral Gn's (Appendix **10.10**). Sequences that mapped onto domain A (teal) exhibit a level of sequence conservation ranging between 12 and 32%. Sequence conservation outside of Domain sequences that mapped onto other parts of the Gn had a sequence identity between 19 and 38%.

The twelve predicted N-linked glycosylation sites from the selected phleboviruses were mapped onto the structure of RVFV Gn (Figure **3.7A**). One predicted glycosylation site from the UUKV is absent from the structure as it maps onto the putative domain C.

Interestingly, all the other glycosylation sites map onto domain A and domain B, with most sites located on domain A. Despite the lack of conservation in this region, all the sites are solvent exposed on the RVFV Gn and therefore likely to be glycosylated. It seems that most of the glycosylation is focused onto the part of the Gn which has low sequence conservation and this may have a functional role as discussed in the next section.

3.6 Concluding remarks: A conserved class II chaperone protein architecture.

The crystal structure of RVFV Gn is a stepping stone towards understanding the function of the Gn and the organisation of the phlebovirus glycoprotein shell. Although the folds are different, the secondary structure elements of RVFV Gn, PUUV Gn, and CHIV E2 share similar features, indicating that RVFV Gn may also act as a class II 'fusion chaperone'.

Sequence alignment of a panel of representative phleboviral sequences and mapping of N-linked glycosylation sites onto the RVFV Gn structure reveals that domain A is the most glycosylated. What are the function(s) of glycosylation sites amongst phleboviruses? Glycosylation sites are assumed to occur most often in solvent exposed regions of the protein. As a result, the location of glycosylation may inform on the arrangement of the Gn in the context of the whole virus. Given the high level of glycosylation in Domain A, it seems unlikely that this region will be involved in extensive protein-protein contacts within the higher order Gn-Gc spike structure observed in low resolution EM studies. Thus, it seems more likely that the less glycosylated and more sequence conserved β-ribbon and the putative domain C may be responsible for forming these protein-protein contacts. Additionally, if we assume that all phleboviruses assemble into an icosahedron in a similar way as the RVFV and UUKV [90, 93], one would predict a conserved mode of Gn assembly across the phleboviruses generally. However, without a

higher resolution structure of a whole phleboviral virion, these are only hypotheses and need to be confirmed.

Sequence divergence and variation in glycosylation location may also play a role in the context of a viral infection. If the glycosylated domain A is largely solvent exposed, it will also be exposed to the host immune system. This would create conditions where it would be favourable for the virus to evade the humoral immune response by introducing N-linked glycosylation sites in the domain. Indeed, the generation of N-linked glycosylation on domain A could effectively make it harder for the immune system to recognise the virion particle. Immune evasion by the generation N-linked glycosylation has precedents, and has be observed in other virus systems including HIV-1, influenza virus, and arenaviruses [136–138].

The secondary structure similarity between RVFV Gn and CHIV E2 may provide clues about the nature of interaction between the Gn and Gc on the phlebovirus surface. In CHIV, the E1 and E2 interact mainly via domain II of E1 and the β-ribbon of E2 [78]. However, all the other domains of E2 are involved in some contacts with the E1, notably with the fusion loop sitting in a groove between domain A, B, and the β-ribbon. If this is extrapolated onto the RVFV, one can expect that all the domains are involved in some form of protein-protein contacts but that the β-ribbon is integral to such contacts. Even though this is speculative, it is consistent with the observed conservation of the β-ribbon motif and its lack of glycosylation.

Less is known about the PUUV Gn and how hantaviruses assemble, but a 16 Å resolution EM density map of Tula virus and a fitting of the PUUV Gn into the membrane-distal region of the virion [81]. At the time of that report, there was no structure available of any hantavirus Gc so interactions between the two have not been seen or proposed. The PUUV Gn was placed into the density in such a way that domain A would primarily be

involved in protein-protein interaction. Since domain A of RVFV Gn appears to have very little homology to domain A of PUUV Gn, it is unlikely that they will make similar contacts.

Despite the genetic and pathobiological distinction between CHIV and RVFV, there are several properties that these viruses share, which RVFV and PUUV do not share despite being part of the same *Bunyavirales* order. Firstly, RVFV and CHIV form icosahedral particles [79, 90] while PUUV is largely pleomorphic with patches of a 2D lattice [81]. RVFV and CHIV are transmitted by mosquitoes and can use DC-SIGN for entry [37, 40] whereas PUUV is transmitted by rodents and may enter through a number of receptors that have been implicated for hantaviruses, such as integrins, decay accelerating factor and complement receptor gC1qR-p32 [139–141]. Perhaps these differences in host and receptor tropism may have led the hantaviruses down an evolutionary path which differed from that of phleboviruses and alphaviruses. However, this does not explain why the CHIV E2 seems to be related to both the RVFV Gn and the PUUV Gn, while the two Gn proteins don't seem to be related.

The structure of the RVFV Gn determined here provides a platform to address many important questions on phleboviral structure, assembly and pathobiology. Further work will undoubtedly focus on the Gn-Gc interaction, the structure of the putative domain C and how the Gn plays a role in virion assembly, maturation and fusion. By looking at the structure of the whole virus one may start to answer some of these questions.

4. Single particle cryo-EM analysis and localized reconstruction reveal structural plasticity of the RVFV.

4.1 Foreword

This chapter deals with the experimental and computational side of obtaining a cryo-EM model of the RVFV. It proved a bit tricky to produce a homogenous virus sample suited for collecting single-particle data, hence some of this chapter deals with that optimisation process. The tomography data presented in this chapter was all processed by dr. Sai Li, a post-doc in the Huiskonen group. The data is presented with Dr. Li's permission and is meant for comparison only. Some of the resolution estimates for the single-particle data may be overestimated as discussed in the conclusions.

4.2 Summary

To further our understanding of the RVFV glycoprotein layer we sought to solve the structure of the whole virus by cryo-EM. To this end, a new purification method was developed which includes chemical fixation and reverse dialysis. A single-particle approach resulted in a relatively low 13.3 Å resolution map of the whole virion. A localized reconstruction method was adopted which pushed the resolution of individual capsomers to 7.7–8.6 Å. Comparison with a tomographic reconstruction of an unfixed, live virus showed that the maps obtained from fixed virus are physiologically relevant. Flexibility must play an important role in the assembly of the RVFV since only two glycoproteins are able to form four distinct capsomers on the virion surface.

4.3 Purification and optimisation of RVFV particles for single-particle cryo-EM

4.3.1 Containment considerations

A major part of this project was establishing a method to purify structurally homogenous RVFV particles to high enough concentrations for single-particle cryo-EM. A major limitation for this project was the containment considerations for the RVFV. All work was carried out using a vaccine strain, clone 13, which lacks the NSs virulence factor and is routinely used for vaccination in sub-Saharan Africa [142]. Despite using a safe vaccine strain, the virus had to be handled at containment level 3 (CL3) since it is a replicative vaccine and if an outbreak would occur in the local domestic livestock it could be difficult to distinguish between the wild-type virus and the vaccine strain.

4.3.2 Finding a cell type to optimise particle number

An initial screen of several cell types was undertaken. A cell type should produce virus to a high titer and structurally intact particles. Vero cells, BHK21 cells from Helsinki (BHK-Hel) and BHK21 cells from St. Andrews (BHK-And) were infected and the cell supernatant was tested for virus titer in a plaque assay. Plaques were visible and easily detectable from Vero and BHK-And cells with titers at ~$2x10^6$ and ~$3x10^6$ PFU ml^{-1} supernatant, respectively. BHK-Hel cells seemed unaffected by the presence of the virus and no cytopathic effects were visible. BHK-Hel cells are used locally to produce Uukuniemi virus (UUKV) which normally grows to titers of ~$2x10^6 ml^{-1}$ of supernatant. However, the infection of UUKV does not produce any cytopathic effects and detection of the virus requires a focus-forming unit (FFU) assay. Thankfully, several hybridoma cell lines which secrete a monoclonal antibody against the RVFV Gn were available from Thomas Bowden. The supernatant from these hybridoma cells was used in an FFU assay to assess the titers

of BHK-Hel produced virus and it was found that indeed the cells do propagate the virus but only to titers of ~$3x10^5$ FFU ml^{-1} of supernatant. Therefore, Vero cells and BHK-And cells were chosen for further studies.

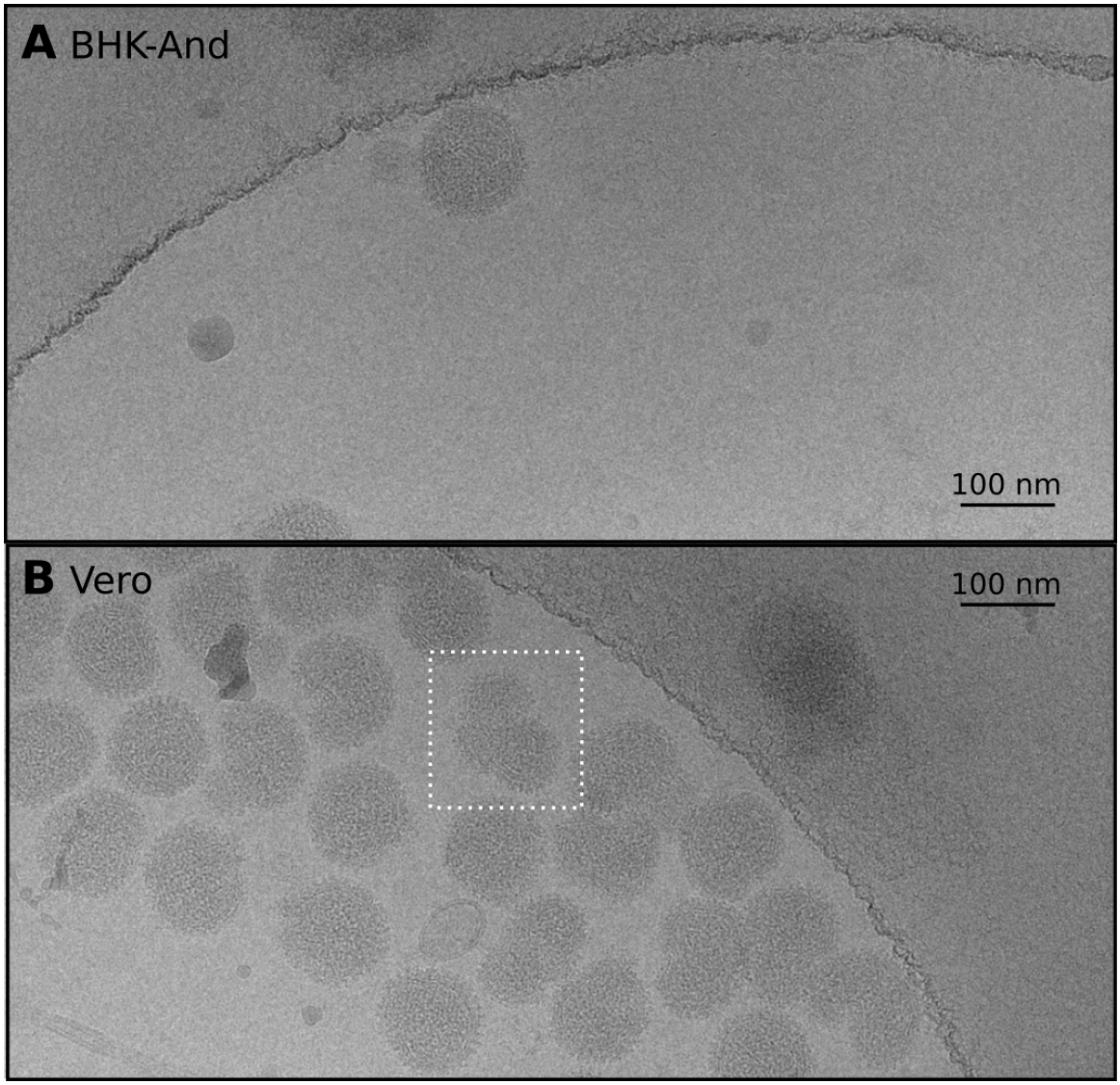


Figure 4.1 Comparing virus production in BHK-And and Vero cells.
Raw micrographs taken on an FEI Technai F30 'Polara' microscope, equipped with a K2 Summit detector and a GIF Quantum energy filter (see Methods **2.5.2**). Micrographs were filtered to 10 Å. BHK-And and Vero cells were infected with the RVFV and particles were purified by ultracentrifugation (see Methods **2.4.3**). A) BHK-And cell produced particles were and disperse on the grid. B) Vero cell produced a greater number of viral particles which were equally distributed on the grid. The white box shows an example of a particle that looks pleomorphic or possibly damaged during the purification process. Most of the particles observed from this grid and the BHK-And grid in *panel A* had a similar deformed shape.

To assess the number and quality of particles produced by each cell type, the virus was inspected by cryo-EM. Virus was concentrated from the cell supernatant by pelleting through a sucrose cushion in an ultracentrifuge. Virus was resuspended in buffer, grids were prepared and imaged in an FEI Tecnai 30 'Polara' electron microscope. As expected, particles ~110 nm in diameter were observed, matching the previously reported structure of the virus [90]. Comparison between particles produced in BKH-And and Vero cells showed that particle density was much greater for Vero cell produced virus (Figure **4.1**). However, virus produced in either cell line did not appear icosahedral. Most particles looked pleomorphic and possibly deformed by the purification process.

4.3.3 Optimising particle quality.

A method to purify the virus without compromising the particle integrity was sought after but because of CL3 restrictions methods involving chromatography or other sophisticated methods were simply not feasible. The method tested which retained particle integrity the best was chemical fixation. The method involved placing supernatant containing virus from Vero cells in a dialysis flask and dialyse against buffer containing 0.2% formaldehyde or glutaraldehyde. The chemically fixed virus was tested in a plaque assay for infectivity and it was found that the fixing agents effectively deactivated the virus. The fixed supernatant was then pelleted through a sucrose cushion, the virus resuspended in buffer and imaged in the FEI Tecnai F30 'Polara' microscope under cryo-conditions (Figure **4.2**). The two chemical fixing agents had different effects on the viral particles. Judging from the micrographs, formaldehyde appeared to fix a particle in cis, i.e. forming cross-links within each particle, while the glutaraldehyde seemed to act more in trans. Glutaraldehyde fixed particles formed very large aggregates and very few single particles were observed.

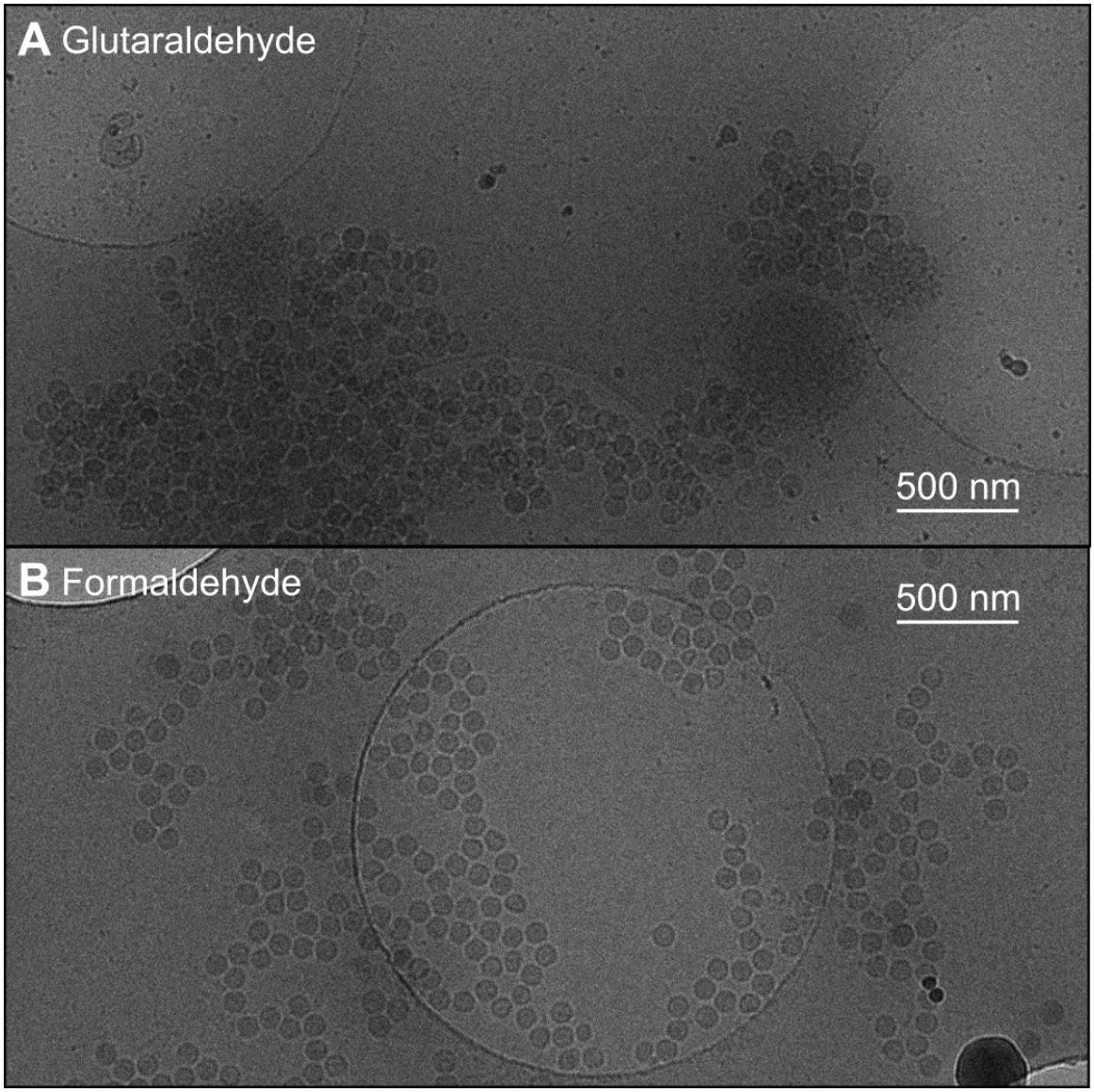


Figure 4.2 The effects of fixing agents on the RVFV particle quality.
Supernatant from Vero cells containing RVFV particles was chemically fixed and the effect on the particle was assessed by cryo-EM. Images were taken as described in Figure **4.1** A) Fixing with glutaraldehyde led to particle aggregation. B) Fixing with formaldehyde resulted in some aggregation but most of the aggregated particles still formed a monolayer which could be imaged. Additionally, more single undispersed particles were observed than in *panel A*.

Pelleting was avoided since it was suspected that subjecting the virus particles to 100,000g and crushing them at the bottom of a tube might be a major reason for particle deformation. A reverse dialysis method was adopted instead. Virus supernatant fixed in formaldehyde was place in dialysis tubing and solid PEG-35 was poured over the membrane to concentrate the virus via osmosis. The concentrated supernatant was loaded

onto a sucrose gradient and spun in an ultracentrifuge overnight to reach equilibrium. A sharp band was observed on the sucrose gradient (Figure **4.3A**) and the gradient was fractionated (Figure **4.3B**). The peak fraction from the gradient was used to make grids for cryo-EM and the virus particles were imaged in the 'Polara' microscope (Figure **4.3C**). /

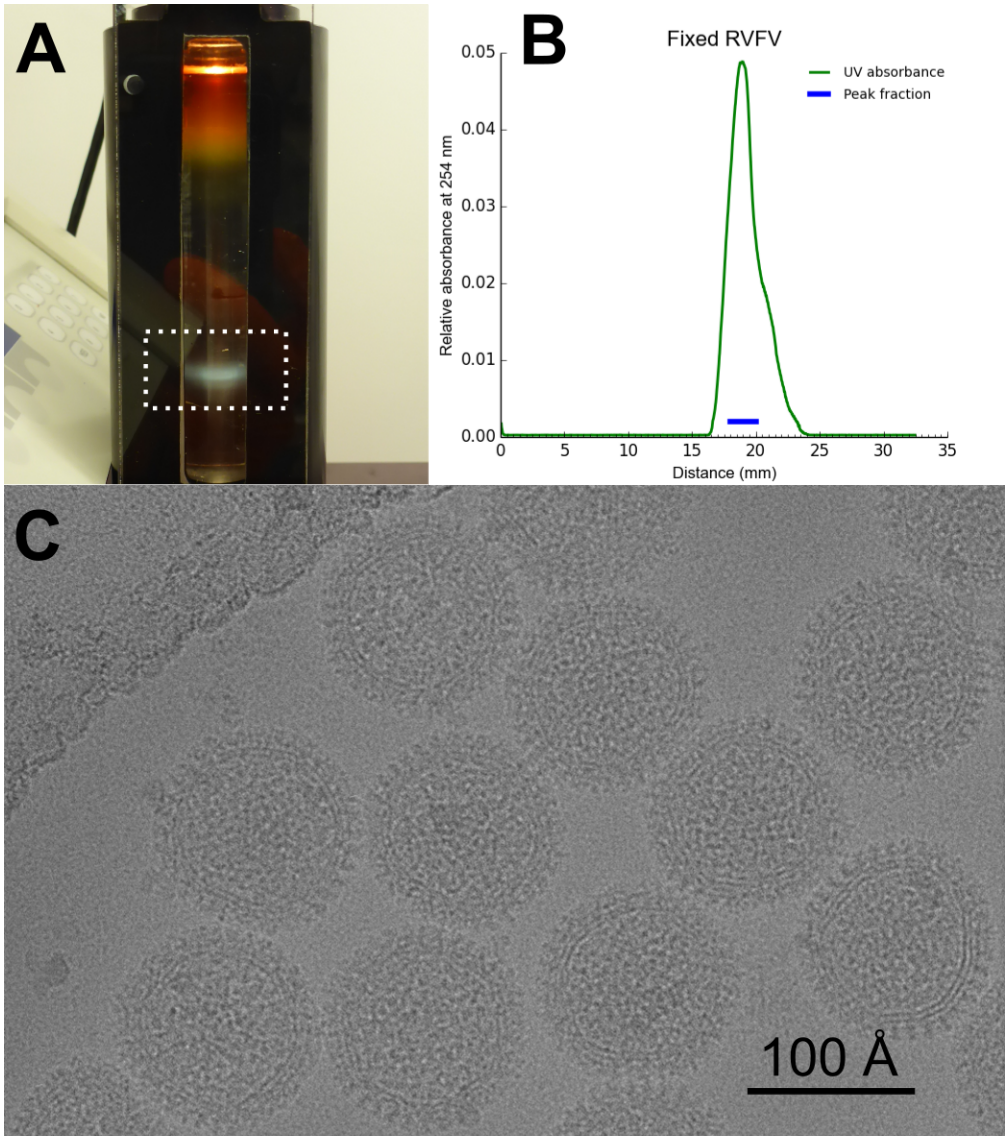


Figure 4.3 Purification of fixed RVFV particles.
RVFV particles were chemically fixed, purified to homogeneity and imaged. A) RVFV particles were run on a sucrose gradient in an equilibrium run. The resulting gradient had a very strong band (highlighted) at ~40% sucrose. B) The gradient was fractionated using a Piston Gradient Fractionator. The resulting peak fraction was collected and used to make cryo-EM grids. C) A micrograph was collected as described in Figure **4.1**. The virus particles prepared using this method looked more uniform and had the regular morphology as expected for an icosahedral particle. Visual inspection of the micrographs shows however that not all the particles are perfect and there is a level of heterogeneity present.

Initial visual inspection revealed that most of the particles were more uniform and homogenous in comparison to unfixed particles (Figure **4.1**). Many of the particles had an overall shape reminiscent of an icosahedron. However, the problem of particle heterogeneity had not been fully solved and some particles appeared distorted.

4.4 Single-particle cryo-EM of the whole RVFV

Despite the slight imperfections observed, a single-particle dataset of 943 micrographs was collected (see Methods **2.5.2**). 4,663 particles were picked manually and data was processed in RELION 1.4. An initial round of 2D classification with narrow angular sampling was done and followed by 3D classification and refinement. A final map was obtained at 13.3 Å resolution from 2,995 particles. Refinement statistics can be found in Table **4.1** and Fourier shell correlation (FSC) curves can be found in Appendix **10.11**.

The virus has a T=12 icosahedral symmetry and contains 4 different kinds of 'capsomers' on the surface which are made of only two glycoproteins Gn and Gc (Figure **4.4**). Pentamers lie on the 5-fold symmetry axis of the particle, type 2 hexamers lie on the 2-fold symmetry axis, type 3 hexamers lie on the 3-fold symmetry axis and the type 1 hexamers do not lie on a symmetry axis. The lipid bilayer of the virus is well resolved and some of the transmembrane regions are even visible in the density. The inside of the virus was masked during the refinement to avoid aligning the interior which is expected to be largely unstructured.

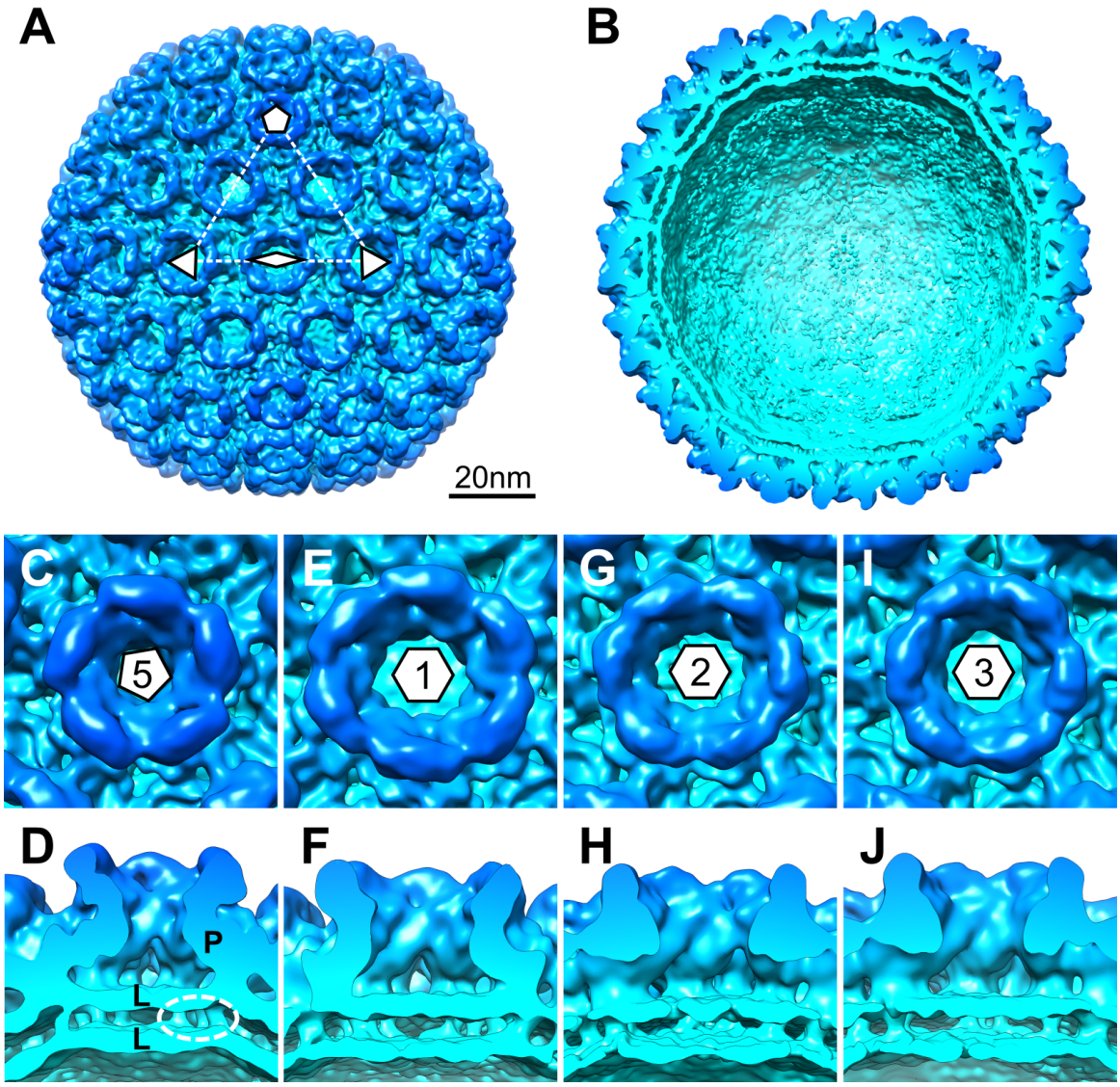


Figure 4.4 Single-particle cryo-EM map of RVFV at 13.3 Å resolution.
A) Reconstructed map of the whole RVFV. The assymetric unit is highlighted with a white dotted line and axes of symmetry are labelled as follows: the pentamer on the 5-fold axis is shown with a penton, the hexamer type 2 on the 2-fold symmetry axis is shown with a diamond, and the hexamer type 3 on the 3-fold symmetry axis is shown with a triangle. B) The interior of a virus segmented in half. Density for the glycoprotein layer and the lipid bilayer is well resolved while the lumenal part of the virus is unresolved. C) A zoom-in on the pentamer showing a top view. D) A side view of the pentamer density cut through the middle. The glycoprotein (P) layer and the lipid (L) bilayer are well defined and even transmembrane helices (dotted circle) are visible between the two leaflets. E) & F) The type 1 hexamer, not on a symmetry axis. G) & H) The type 2 hexamer from the 2-fold symmetry axis. I) & J) The type 3 hexamer from the 3-fold symmetry axis.

The relatively low resolution of 13.3 Å did not provide much more interpretable information than the previously published 20 Å resolution map [90]. The low resolution

may be explained by the observed heterogeneity of the purified particles, despite the steps taken to limit particle distortion. The large particle size of ~110 nm in diameter also complicates data collection and computational processing. Very fine angular sampling is required during refinement which can greatly slow down the refinement and any minor local distortions to the particle may propagate into several Å shifts on the global scale of the particle. The large particle size limits data collection efficiency as exemplified by the fact that the final 2,995 particles came from 934 micrographs which gives an average of only 3.2 particles per micrograph. To overcome this limitation, we turned to a novel method.

4.5 Localized reconstruction overcomes particle heterogeneity

In order to maximise the information present in the micrographs of the RVFV, a localized reconstruction method [127], developed locally within the Huiskonen group, was adopted. The great benefit to this method is that components (sub-particles) of the virion can be defined and refined independently of the whole virus. Deviations from a perfect icosahedron can thus be accounted for by allowing for rotation and translation of the defined sub-particles which otherwise would not be possible when refining an entire virion. Furthermore, sub-particles can be classified and only sub-particles which positively contribute can easily be selected and ill-aligned particles can be discarded.

Each of the four capsomer types was defined as an independent sub-particle using the localized reconstruction method and processed as a single-particle data set in RELION 1.4 (See Methods **2.6.2**). The sub-particles were boxed out and subject to and initial 3D classification. A large part of the sub-particles was weeded out during this step, especially the hexamers where 68-70% of the initial particles were removed (Figure **4.5**). The selected

classes were then refined to give four final maps at resolutions between 8.6 and 7.7 Å (Figure **4.6**). Refinement parameters and statistics can be found in Table **4.1**.

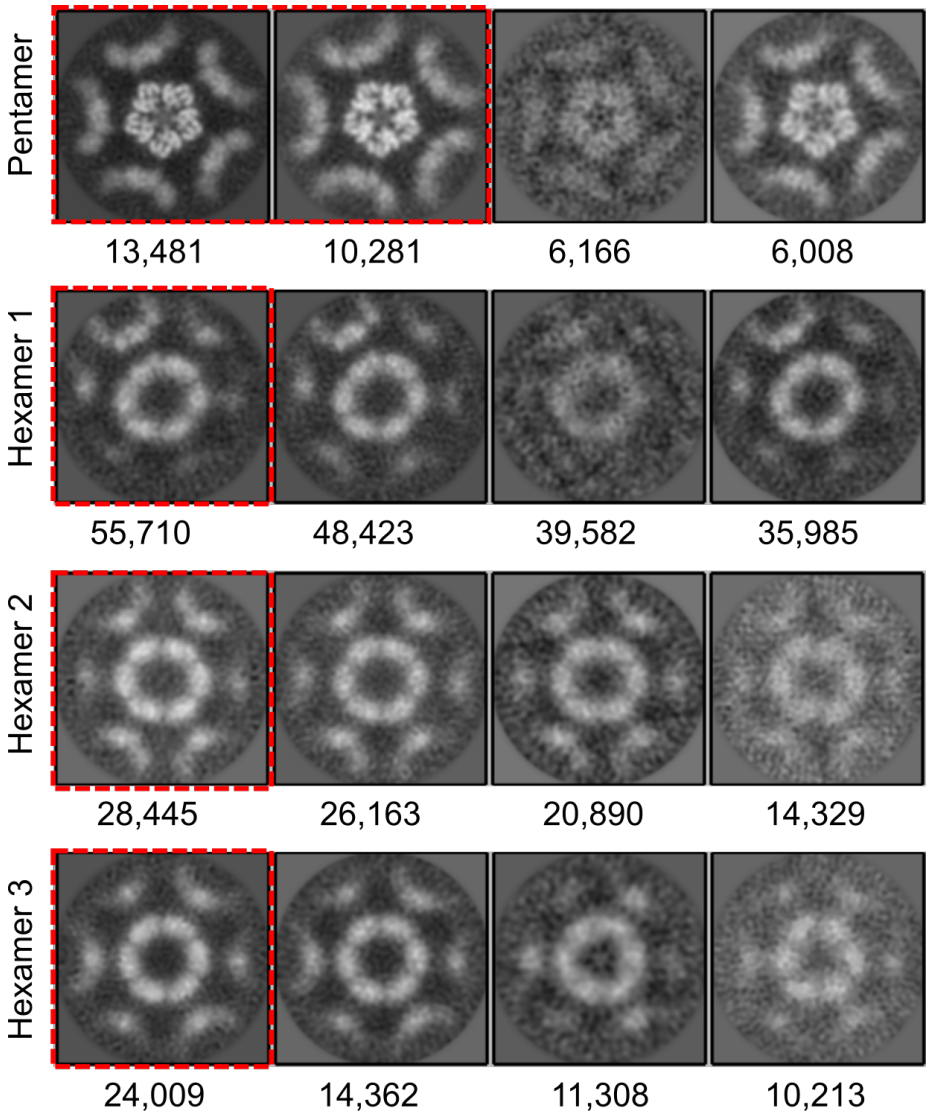


Figure 4.5 3D classification.
The sub-particles were subject to 3D classification in RELION. Each row shows the four classes resulting from the classification of each of the capsomers. Each class is represented as a central section through a 3D density from each class. The numbers bellow each class shows the number of particles in each class. The red dotted line shows which classes were selected for further 3D refinement.

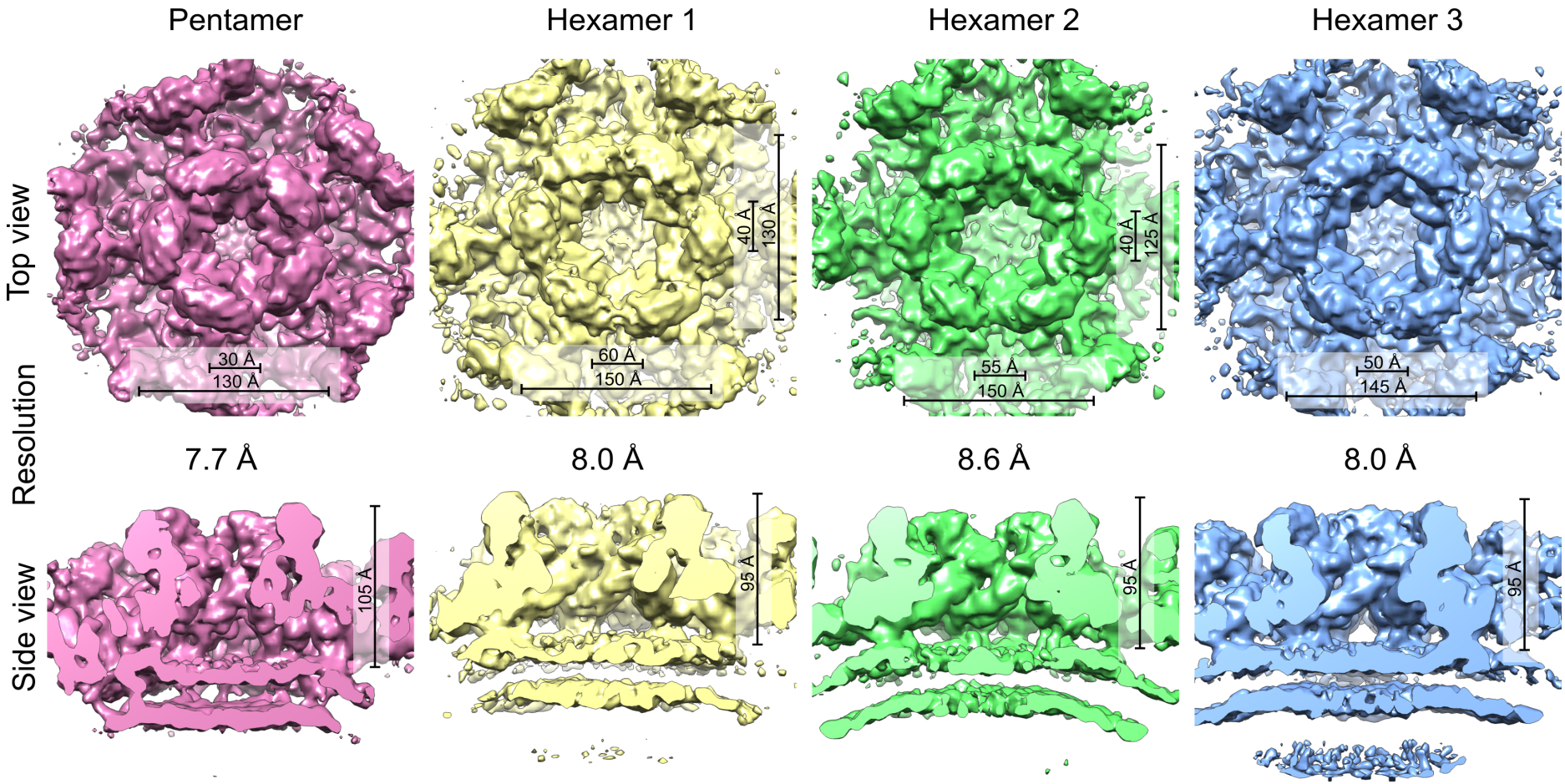


Figure 4.6 Localized reconstructions of individual capsomers.
Individual capsomers were subject to single-particle refinement independent of each other, resulting in four distinct models of the pentamer and the type 1–3 hexamers. Each capsomer is shown in top view and side view. The outer and inner diameter of each capsomer is shown in the top views, for the type 1 and 2 hexamers two values are shown. The height of each capsomer from the outer membrane leaflet is shown in the side view. In contrast to the low-resolution map of the whole virion, transmembrane helices are mostly absent in the density of the individual capsomers, apart from the transmembranes of the pentamer. As discussed in the main text, the resolution estimates may be inaccurate.

The gain in resolution from localized reconstruction is evident by the increased detail of the density maps. The one exception is the absence of transmembrane helices from all hexamer types. Their absence is surprising given that transmembrane helices were largely visible in the density of the whole virus for all the capsomer types. All the hexamers have a similar height of ~95 Å measured from the outer membrane leaflet while the pentamer rises slightly higher at ~105 Å. As expected the pentamer has the smallest outer diameter of ~130 Å, however the hexamers show a more varied morphology. During refinement of the hexamers, a range of different symmetry was applied and it was found that applying the local symmetry operator gave the best results (Table **4.1**), e.g. applying a C2 symmetry on the type 2 hexamer gave the best resulting maps, as assessed by resolution. Effectively, all the hexamers show a deviation from a C6 symmetry and are morphologically distinct from each other.

The FSC curves for the localized reconstructions (see **Appendix 10.11**) indicate that the resolution estimates may not be reliable. The curves do not seem to reach a value of zero, indicating that some parts of the half-maps used to estimate the resolution correlate beyond what is expected if the maps were generated completely independent. Since the refinement and post-processing was done using standard procedures in RELION which did not produce the same artefacts for the reconstruction of the whole virus, it is likely that the source of this model bias comes from the localized reconstruction procedures. A likely source of bias is the partial signal subtraction (all-but-one) where the model of the whole virus is used to directly subtract signal from each micrograph. The signal subtracted sub-particles are then used as input particles for the single-particle reconstruction in RELION. There are other possible sources of model bias, such as improper masking, overfitting during refinment or starting model bias. However these are not likely sources of bias since the starting model was obtain directly from back projections of the micrographs of the sub-

particle, overfitting is largely avoided in RELION due to the Baysian based algorithms and masking with a soft edged mask to eliminate solvent contributions is a standard procedure which did not yield such artefacts for the model of the whole virus. Ultimately, the resolution of the resulting maps is overestimated and is likely to be an Ångstrom or two lower than the FSC curves indicate.

4.6 The tomographic reconstruction of a native virus shows the physiological relevance of chemically fixed virus.

One obvious caveat of using chemically fixed virus is the potential for the fixing agent to distort the structure of the virus. To assess the validity of the maps obtained from the chemically fixed virus, the native active virion was imaged using cryo electron tomography (cryo-ET) approaches. Virus was purified from infected Vero cell supernatant by simply pelleting the virus though a sucrose cushion and resuspending the pellet in buffer. Grids were prepared from the pelleted virus and cryo-ET data was collected on a 'Polara' electron microscope in collaboration with Dr. Sai Li. The pentamer was boxed out from 120 virus particles and refined to 20 Å resolution. The tomographic reconstruction was done in a similar way as the reconstruction of Tula virus [81]. All tomography data processing and model building was done by Dr. Sai Li and the data is presented with Dr. Li's permission for comparison (Figure **4.7**).

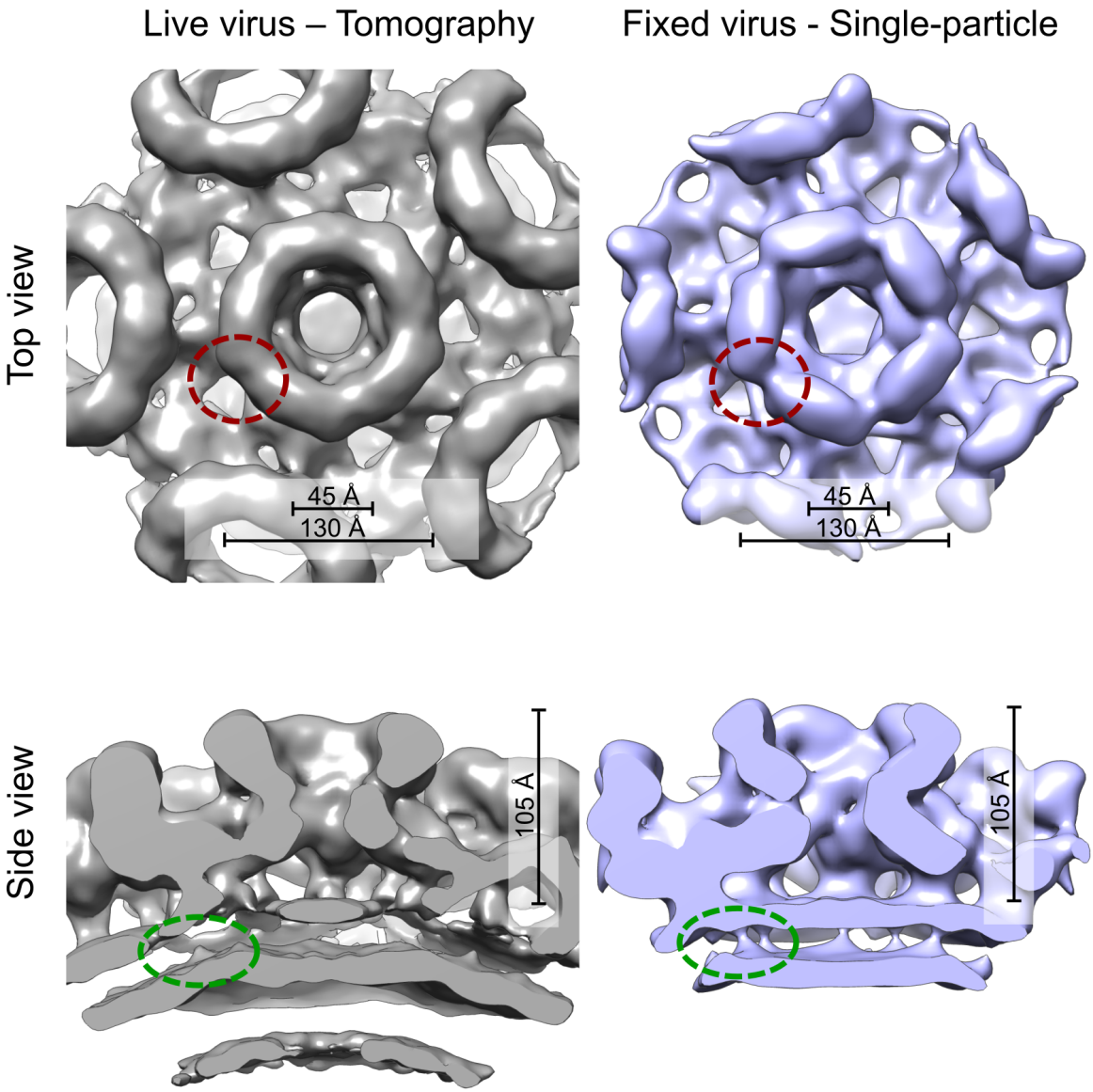


Figure 4.7 Comparing reconstructions from fixed and live virus.
Live, unfixed RVFV was analysed by sub-tomogram averaging and a map of the pentamer was obtained to 20 Å by Dr. Sai Li. The pentamer map obtained from single-particle analysis of the pentamer was filtered to 20 Å resolution to allow comparison between the two maps. The overall morphology of the pentamer is consistent between the two methods. The separation between the subunits of the capsomer is more pronounced in the single-particle structure (red circle). Both maps show density for transmembrane helices in the same location (green circle) although their presence is clearer in the single-particle map.

The pentamer map from localized reconstruction was filtered to the same resolution for comparison. The two pentamer maps superpose well with the 'Fit in Map' function in Chimera [130], giving a cross-correlation of 0.91. The capsomer height and diameter is the same, and even the faint transmembrane density in the tomography reconstruction is also

found in the same location as in the single-particle map. The major difference between the two maps is in the membrane-distal part where separation between the caspomer subunits is more pronounced in the single-particle map.

In conclusion, the overall structure of the pentamer is not altered significantly by the presence of formaldehyde. The caveat here is that the comparison is only at 20 Å resolution. It may be that the effect of the fixing agent is only visible at higher resolution when side chains become visible, but the best model is somewhat away from showing such features, even at the potentially overestimated 7.7 Å resolution. Therefore, the tertiary and quaternary structure of Gn and Gc is assumed to be largely physiological in the maps obtained from localized reconstruction.

4.7 Concluding remarks: Inherent flexibility of the glycoprotein layer may be important for virion assembly.

The single-particle analysis of the RVFV is an important step towards understanding the assembly of phleboviruses. The low resolution of the whole virion map indicates that the particle has some inherent flexibility. Localized reconstruction of the four capsomer types improved the resolution, which may though be overestimated, and revealed that the capsomer types are distinct from each other. Furthermore, the maps obtained from single-particle analysis are physiologically relevant as shown by comparison with a pentamer map from sub-tomogram averaging of a live virus.

The maps obtained from localized reconstruction show a large variability in the capsomer morphology which may be attributed to some inherent flexibility of the glycoprotein layer. In an attempt to locate such areas of flexibility within the different capsomers, maps showing local resolution of the different capsomers (Figure **4.8**) was created with ResMap [143]. In general, the parts of the glycoprotein layer which is

membrane proximal exhibits higher local resolution than the membrane distal parts. Whether this has implications for particle flexibility is difficult to say. There are several experimental and computational factors that may influence the local resolution. Since the viral particles were fixed with formaldehyde it is likely that the solvent-exposed areas are more affected and this may be reflected in lower resolution. The alignment process during image processing may also introduce artefacts. In most single-particle cryo-EM maps the central core of a structure is better resolved than the peripheral parts, so it is perhaps not surprising that the membrane-proximal parts are better resolved than the membrane-distal parts.

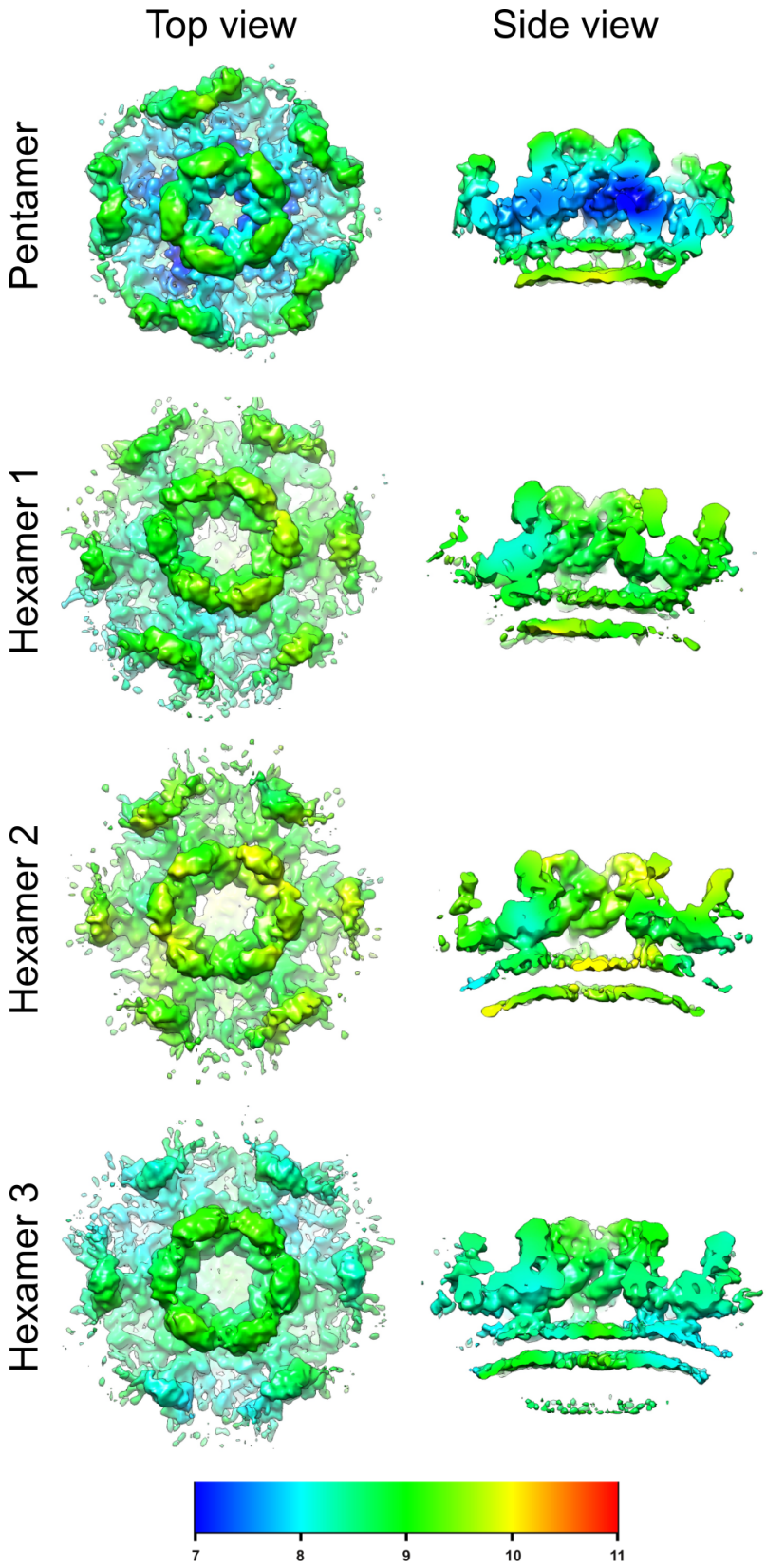


Figure 4.8 Local resolution of capsomers.
The local resolution of the localized reconstructions of the four capsomers was assessed using ResMap [143]. Resolution was assessed in 1 Å steps from 7-11 Å. Each capsomer type is shown from the top and from the side of a segmented density. The best resolved region is generally the membrane-proximal part.

The maps presented here may not be able to tell us exactly how flexibility is required for virus assembly but it is most likely that flexibility does play a very important role in assembly. The subunits which make up the distinct capsomers exist in quasi-equivalent environments, as first discussed by Caspar and Klug [144], and a T=12 number means that the asymmetric unit is composed of twelve quasi-equivalent subunits. The observed flexibility of the RVFV capsomers is thus not surprising. In viral capsids of high T-numbers of T=7 and above, accessory proteins are often required to 'cement' or stabilise the assembly [145] and several bacteriophages have evolved more than one highly discriminatory cement proteins [146, 147]. In a contrasting way, multiple copies of a single capsid protein in HIV assemble into hexamers and pentamers which form an irregular 'fullerene cone' [148]. This mode of heterogenous assembly is allowed by internal flexibility of the capsid proteins itself and evades the problem of having to form a perfect icosahedron. The RVFV virion assembly may lie somewhere in between these two modes where the strict icosahedral rules are slightly 'relaxed' but not so much that the virus becomes pleiomorphic. Somehow, Gn and Gc are all that is required to form a T=12 quasi-equivalent virion [149].

A higher resolution structure of the capsomers would greatly help in understanding exactly how the two proteins assemble into a full virion. This in turn might reveal the extent of the flexibility and the quasi-equivalent nauture of the Gn-Gc assembly. Since the crystal structures of the two ectodomains exist, a pseudo-atomic model of the whole virion can be built.

Table 4.1 RVFV single-particle data collection and refinement parameters

Parameters	RVFV	Hexamer 1	Hexamer 2	Hexamer 3	Pentamer
Data acquisition					
Frames per movie	88	N/A	N/A	N/A	N/A
Exposure per frame (s)	0.2	N/A	N/A	N/A	N/A
Dose rate (e^-/ pix / s)	2.5	N/A	N/A	N/A	N/A
Total dose (e^- / Å^2)	22	N/A	N/A	N/A	N/A
Defocus* (μm)	1.0–3.0	N/A	N/A	N/A	N/A
Data processing					
Micrographs	943	N/A	N/A	N/A	N/A
Particles**	2,995 (4,336)	55,710 (179,700)	28,445 (89,827)	24,009 (59,892)	23,762 (35,936)
Box size (pixels)	512	128	128	128	128
Symmetry	I1	C1	C2	C3	C5
Pixel size (Å)	2.7	2.7	2.7	2.7	2.7
Resolution (Å)***	13.3	8.0	8.6	8.0	7.7
B-factor applied	–800	–200	–200	–200	–200

*** Positive value denotes underfocus.**
**** Number of particles used for the reconstruction (numbers of all extracted particles are parenthesis).**
***** Resolution (Fourier shell correlation = 0.143). See Appendix 10.11 for FSC graphs**

5. The pseudo-atomic model of the RVFV revealed by flexible fitting

5.1 Foreword

One of the main aims of this thesis was to describe the glycoprotein layer of the RVFV in detail. To this end, single-particle cryo-EM and localized reconstruction presented in chapter 4 have pushed the resolution limits to sub-nm resolution. The resulting maps fall a bit short of the resolution needed to accurately define the domain organisation of the Gn and Gc. However, the crystal structure of the Gn ectodomain presented in chapter 3 is the missing piece needed to solve the puzzle of the glycoprotein layer. In this chapter, I use the Gn and Gc crystal structures and the localized reconstructions to build a pseudo-atomic model of the RVFV using the Molecular Dynamics Flexible Fitting (MDFF) tool. As the resolution is limited, most of the analysis focuses on domains and secondary structure and leaves out side chains and atomic details. For consistency, all figures involving the Gn and Gc models are coloured in the same way (unless specified otherwise): domain I in red, domain II in yellow, domain III in blue, fusion loop in orange, domain A in teal, domain B in forest green and the β-strand in purple.

5.2 Summary

The pseudo-atomic model of the RVFV is presented in this chapter. A pseudo-atomic model of each capsid was constructed using MDFF and the quality of the fits was assessed by the agreement between map and model, predicted glycosylation sites and location of C-termini. The capsomers are made of Gn-Gc heterodimers where membrane-distal Gns form a ring which caps the membrane proximal Gcs. A structural analysis shows that the

capsomer assemble in a heterogenous manner from the stable yet flexible Gn-Gc heterodimer building block. The model suggests that the Gn functions as a fusion chaperone much like the E2 from CHIV. Furthermore, based on the pseudo-atomic model of the full virion, a mechanism of assembly and fusion activation is presented.

5.3 Generating the starting model – fitting of Gc

5.3.1 Finding the Gc density

One of the initial difficulties in fitting the crystal structures was finding the part of the density map that corresponded to either the Gc or the Gn. The highest resolution map was that of the pentamer at 7.7 Å and at this resolution one would expect alpha helices to be discernible while β-strands separation only start to become visible at ~5Å. Since both glycoprotein structures are mainly β-based, secondary structural elements are not particularly helpful tools at this resolution and only the overall morphology of the glycoproteins could be used as a guide.

An initial assumption was made that the two glycoproteins form a heterodimer which forms the building block for each of the capsomers, i.e. a pentamer is assumed to be made up of five Gn-Gc heterodimers and a hexamer is made up of six heterodimers [90]. The assumption is based on two lines of evidence: Firstly, during biosynthesis the glycoproteins are expressed in a 1:1 ratio and must form a heterodimer to be trafficked correctly to the Golgi [53–55]. Secondly, heterodimers have been isolated from UUKV and RVFV infected cells [34, 150].

Domains I and II of the Gc form a ~100 Å long rod when filtered to 8 Å resolution (Figure **5.1A**) and should constitute roughly 35% of the mass of the ectodomains of a heterodimer and was used as a starting model. The pentamer density was used for the initial

Gc fit since it was the highest resolution map. When the density of the pentamer is viewed from an intra-viral perspective, one can see that the membrane-proximal part of the glycoprotein layer is made up of densities that could accommodate five rod-shaped Gc domains I and II (Figure **5.1B&C**). A protomer of the Gc in the pre-fusion conformation [47] was manually fit into the corresponding density using Chimera [130] and subsequently a rigid body fitting (fit-in-map function in Chimera) was used to generate the first fitting. Suffice to say, the first fit of Gc did not completely explain the density which the Gc was suspected to occupy and a correlation score of only 0.71 did not evoke confidence in the fitting (Figure **5.2A–C**). Various attempts with different starting orientation did not yield better fits as assessed by the cross-correlation value accompanying the rigid body fitting and it was concluded that rigid body fitting was not the optimal tool.

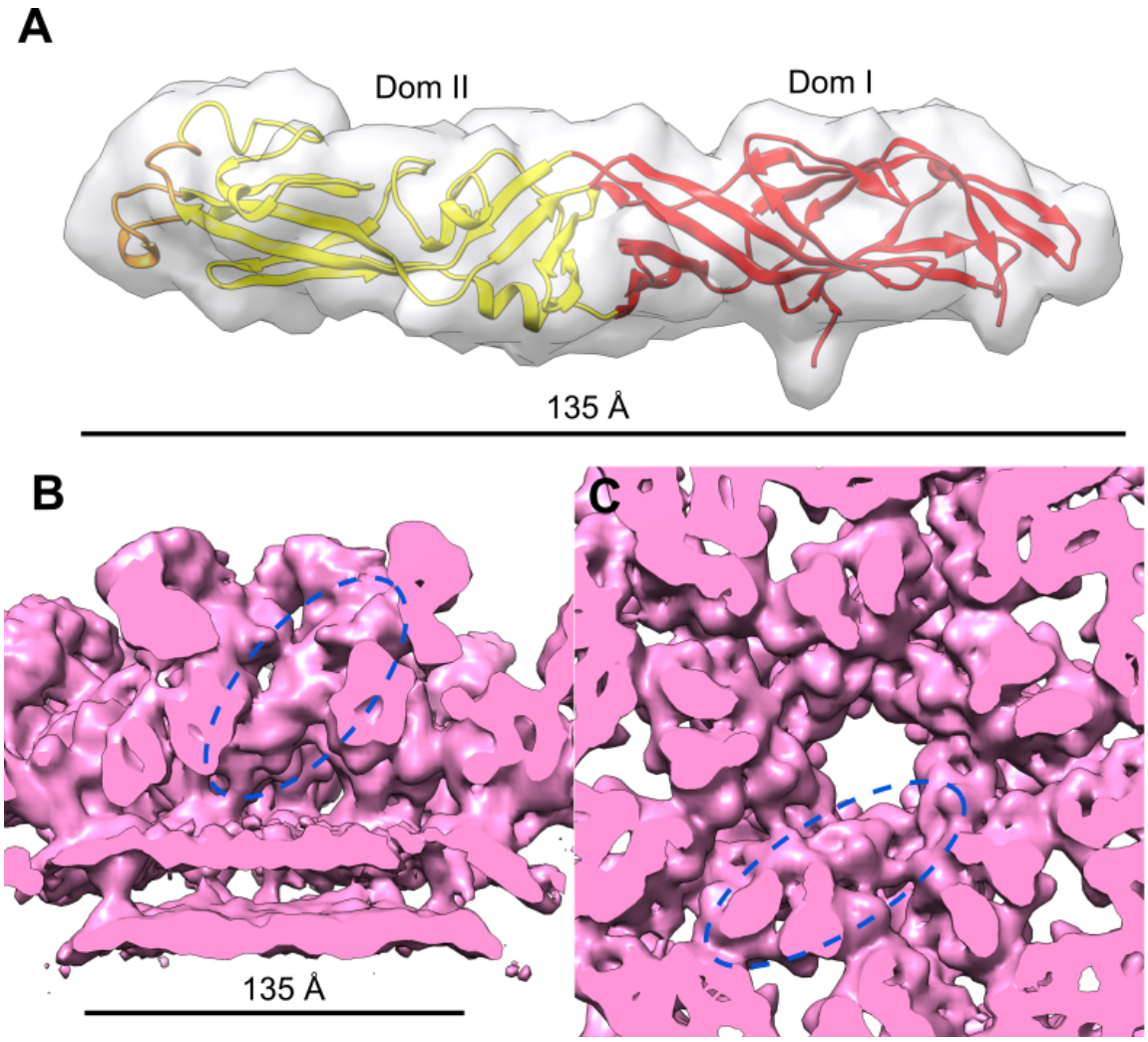


Figure 5.1 Searching for the Gc density.
A) Domain I and II of the RVFV Gc crystal structure form a continuous 135 Å long rod when filtered to 8 Å resolution. B) A side view of a sectioned density of the pentamer reconstruction from cryo-EM. A rod-like density seen in the reconstruction is highlighted with a blue dashed circle. C) The pentamer reconstruction viewed from the lumenal (intra-viral) perspective showing the same highlighted density.

5.3.2 Flexible fitting of a single Gc protomer.

After little success with rigid body fitting, the next step was to explore flexible fitting. Currently, one of the best tools available, which the Huiskonen group had some prior expertise in using, was MDFF [151]. The program essentially runs a molecular dynamics simulation on an atomic structure and uses an EM density as an additional term in the force

field to attract atoms. The strength of this term can be adjusted and a variety of additional restraints can be added, such as secondary structure, symmetry and domain restraints.

MDFF was used on a single Gc protomer in the pre-fusion state with secondary structure restraints and the rigid body fit from Chimera mentioned above was used as a starting position (see Methods **2.6.3**) (Figure **5.2A–C**). The EM density was segmented to speed up the simulation and excluded the viral membrane and roughly two thirds of the pentamer density (Figure **5.2A**). The Gc quickly slotted into a density which explained well each domain (Figure **5.2D&E**). However, the conformational changes which accompanied the shift were much more dramatic than expected and the straight rod which domain I and II form together was bent at the interface between the two domains at a ~90° angle (Figure **5.2F**). The simulation was repeated with a weaker force field but the same results emerged. The resulting bent Gc molecule did not invoke a great deal of confidence in the fitting since such a dramatic shift is unprecedented. A 'greasy hinge' is present in an equivalent position in other class II fusion proteins [70] but the bending between domain I and II sampled by X-ray crystallography is much more modest. Despite these reservations, the coordinates for the fitted Gc were used for further MDFF of the entire pentamer and the quality of the initial Gc fit was assessed in the context of the whole capsomer model. As the resulting pseudo-atomic model shows, the fitting proved to describe the EM density well.

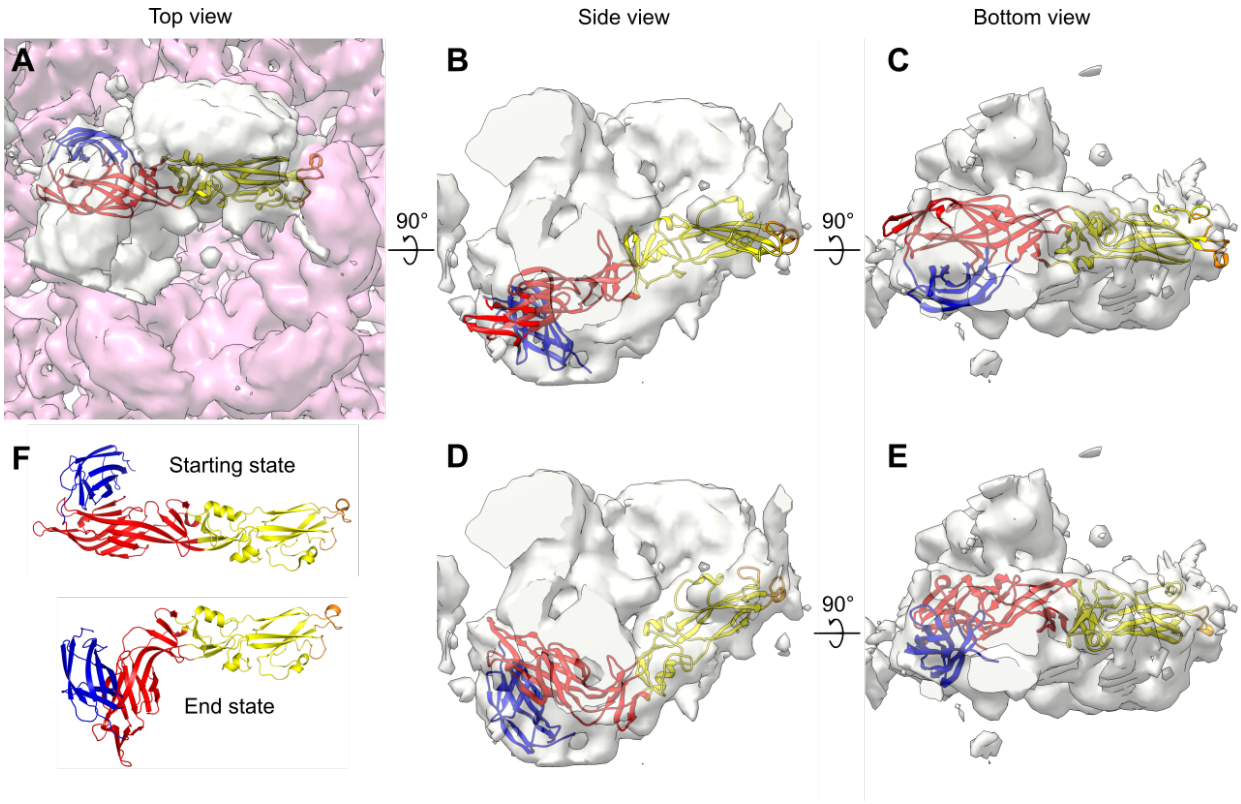


Figure 5.2 Flexible fitting of Gc.
A) A top view of the pentamer reconstruction (pink), the segmented density used for the MDFF simulation (grey) and the rigid body fitted Gc crystal structure (starting state). B) and C) show the segmented density and the Gc starting state from a side view and a bottom view. D) and E) show the Gc structure after the fitting (end state) in the segmented density from a side view and a bottom view. F) Atomic models of the Gc show the conformational change associated with the shift from the starting state to the end state, which results in a ~90º angle shift between domains I and II.

5.4 Flexible fitting of entire capsomers

5.4.1 Performing the simulations

The initial MDFF fitted Gc was placed into a map of the pentamer and five-fold rotational symmetry (C5) was applied to place four other Gc protomers into the density (Figure **5.3**). This left vacant densities at the membrane-distal part of the pentamer where the Gn crystal structure was manually placed in Chimera, followed by a rigid body fit and symmetrisation (Figure **5.4**). To avoid any overlapping molecules, the fitted crystal structures were manually moved laterally a few Å away from the centre of the pentamer (Figure **5.4**). Additionally, the map was segmented to save computational time. A 5 ns simulation was

run in MDFF to generate a fitting for the entire pentamer (see Materials **2.6.3**). The same kind of fitting was performed for all the other hexamers.

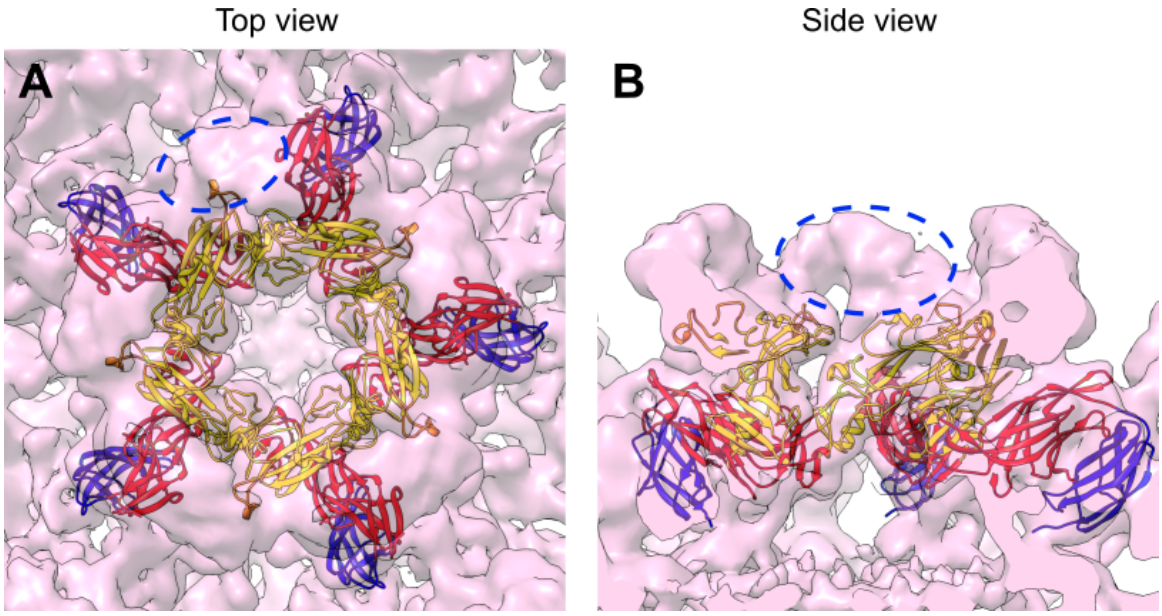


Figure 5.3 Finding a density for the Gn.
The initial fit of the Gc was placed in the pentamer reconstruction and a five-fold symmetry was applied to fit five Gc molecules. A) A top view of the pentamer and B) a side view of the pentamer show a vacant membrane-distal density highlighted with a blue dotted circle.

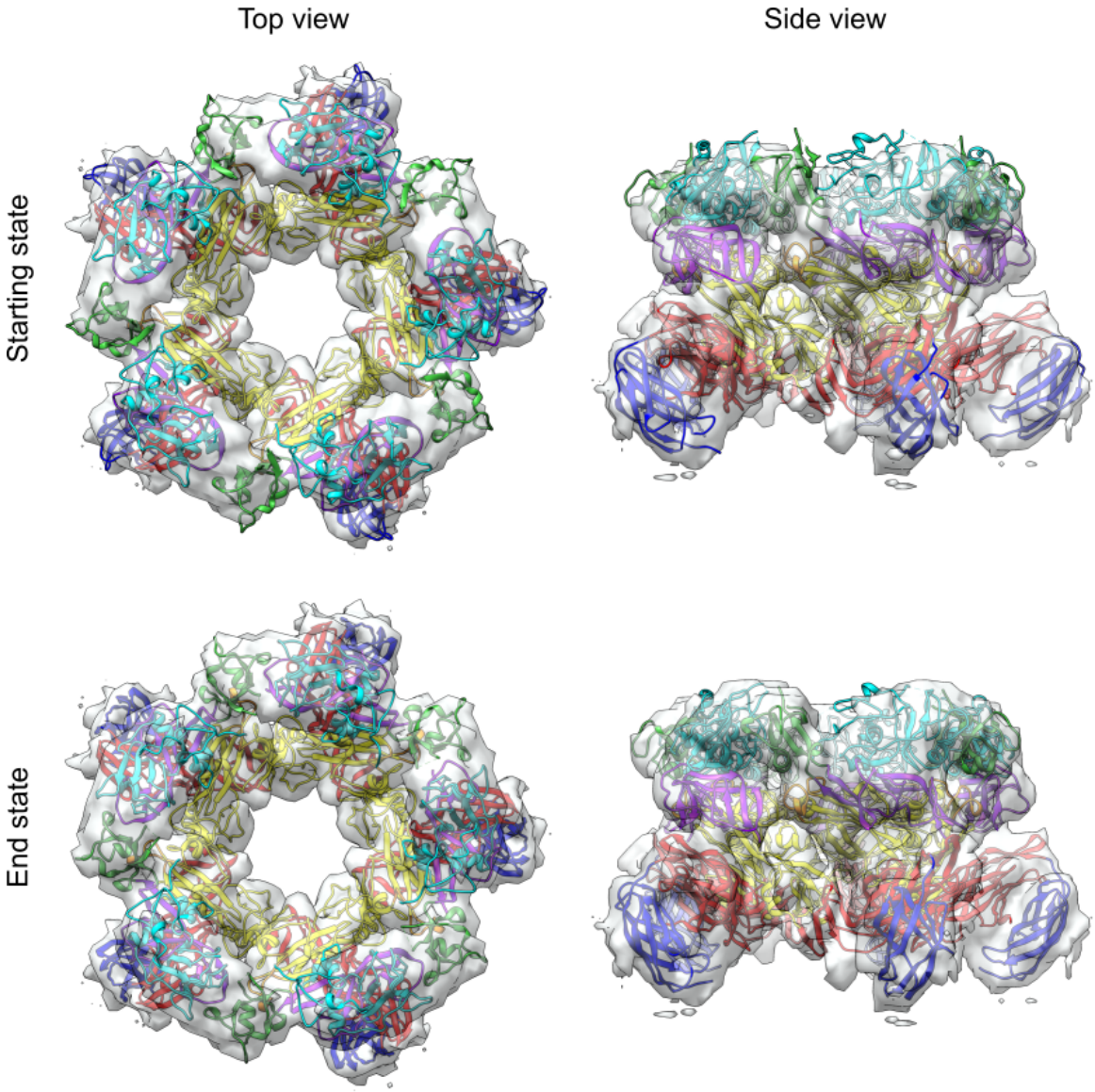


Figure 5.4 Flexible fitting of the pentamer.
The Gc and Gn were fitted into a segmented pentamer density (show in grey). The top two panels show the starting state of the entire pentamer in the segmented density and the bottom two panels show the pentamer after the fitting.

5.4.2 Overall view of the capsomers

The Gns assemble into a ring-like structure which caps the Gcs (Figure **5.5**). Neither the Gn nor the Gc changed much during the simulations apart from slotting back into the density which they had been manually moved out of to avoid clashes. The overall fits explained the densities of the capsomers very well and almost all density was occupied, apart from a small globular density at a quasi-three-fold axis at the interface of three

neighbouring capsomers (Figure **5.7B**). The cross correlation between the maps and the models was monitored during the simulation which showed that the models had largely reached equilibrium (Appendix **10.12**).

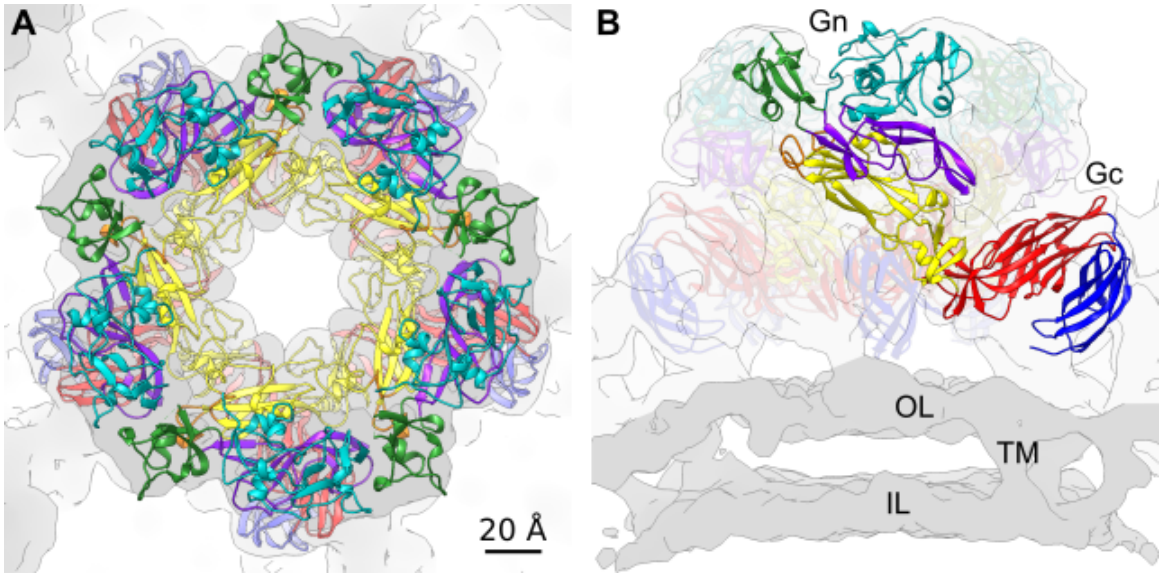


Figure 5.5 Overview of the capsomer model
A) The map of the pentamer (in grey) shown from the top with the pseudo-atomic model fitted inside. B) A side view of the same map. A single Gn-Gc heterodimer is highlighted. The two membrane leaflets (OL = outer leaflet, IL = inner leaflet) and an example of a transmembrane helix (TM) are labelled.

5.4.3 Assessing the quality of the fits

As mentioned above, the membrane-distal density, largely occupied by the Gn, was not as well resolved as the membrane-proximal density occupied by the Gc. Despite the low local resolution of the density for the Gn, the fit explains well the density when viewed up close (Figure **5.6**). Viewing each of the domain of Gc in the context of the map reveals a good match and the secondary structure elements of the Gc are particularly well explained in the density, where density for β-sheets can be seen (Figure **5.6**). However, the small number of α-helical elements are however not well resolved. The final cross-correlation scores

between maps and the models are between 0.87 and 0.9, indicating a good agreement between experimental and simulated data (see Appendix **10.12**).

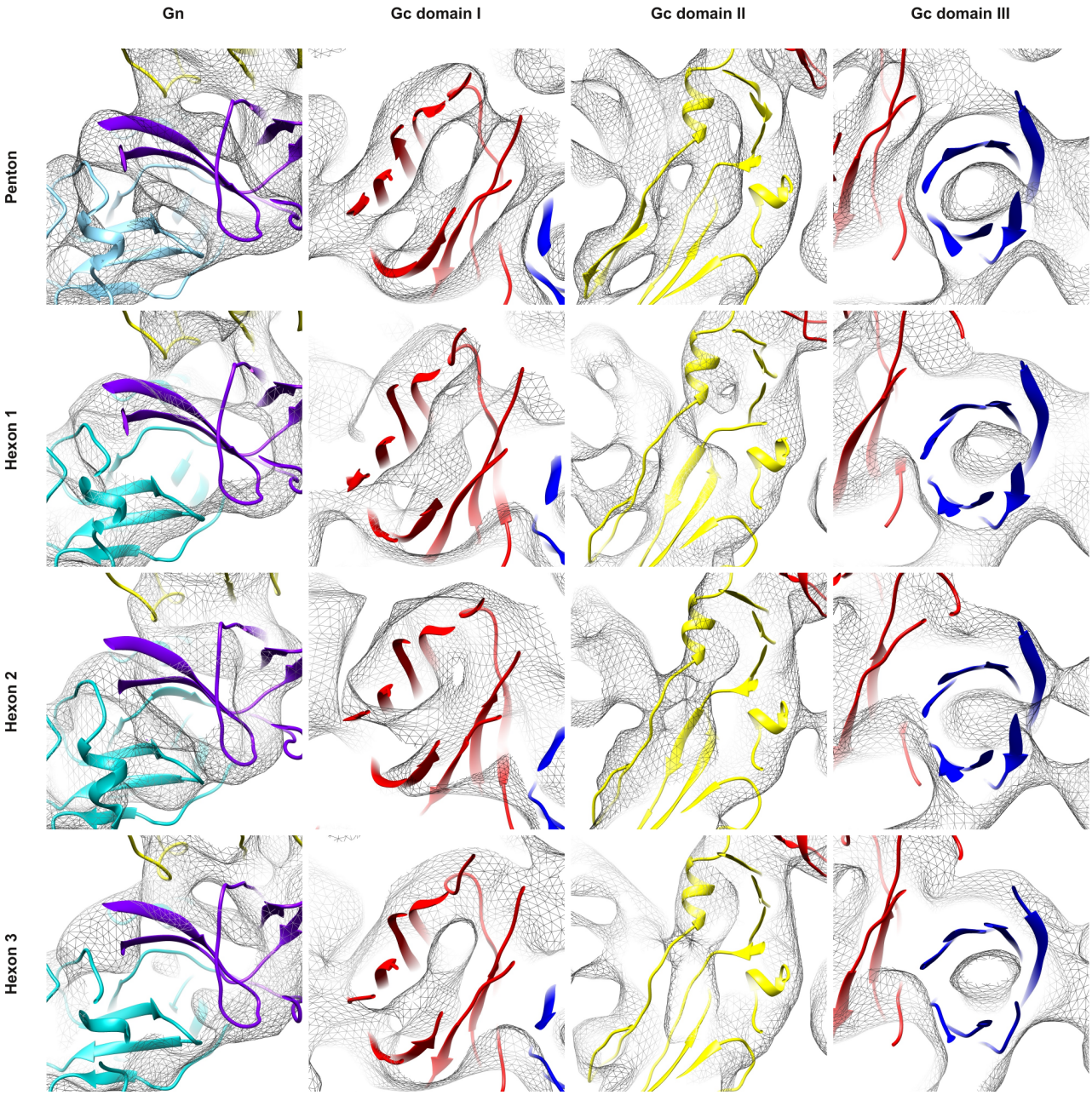


Figure 5.6 Assessing the quality of the flexible fits.
The fits of the capsomers were assessed visually by looking for agreement between the models and the density. The atomic models are shown in cartoon and the density of the capsomers are shown with a surface mesh. Evidence for β-sheet elements of the Gc are clearly seen in the densities of all the capsomers.

To further add confidence to the fit, the C-termini of the Gc and the Gn was analysed. The distance from the C-termini of the glycoprotein models to transmembrane helix bundles should help determine the accuracy of the fits. The Gc ectodomain crystal structure is missing ~40 amino acids which connects the ectodomain to the transmembrane region. The C-terminus of the Gc is right next to a transmembrane helix bundle which should easily accommodate 40 amino acids, even if they were folded into an α-helix (Figure **5.7A**). The C-terminus of Gn is a bit trickier to assess since more than 110 amino acids are missing, part of which might fold into a small globular domain as mentioned in Chapter **3.6**. The C-terminus of the Gn points towards the viral membrane and a tubular density is visible from the C-terminus which connects the Gn model to the unoccupied density mentioned above (Figure **5.7B&C**). Furthermore, the unoccupied density is directly above a transmembrane helix bundle. Even though this is speculative, the unoccupied density may be occupied by an equivalent of domain C in CHIV E2 protein. A rigid body fit of domain C from E2 was fitted in the density for demonstrations only to show that indeed a β-sandwich domain of that size can explain the density (Figure **5.7D**).

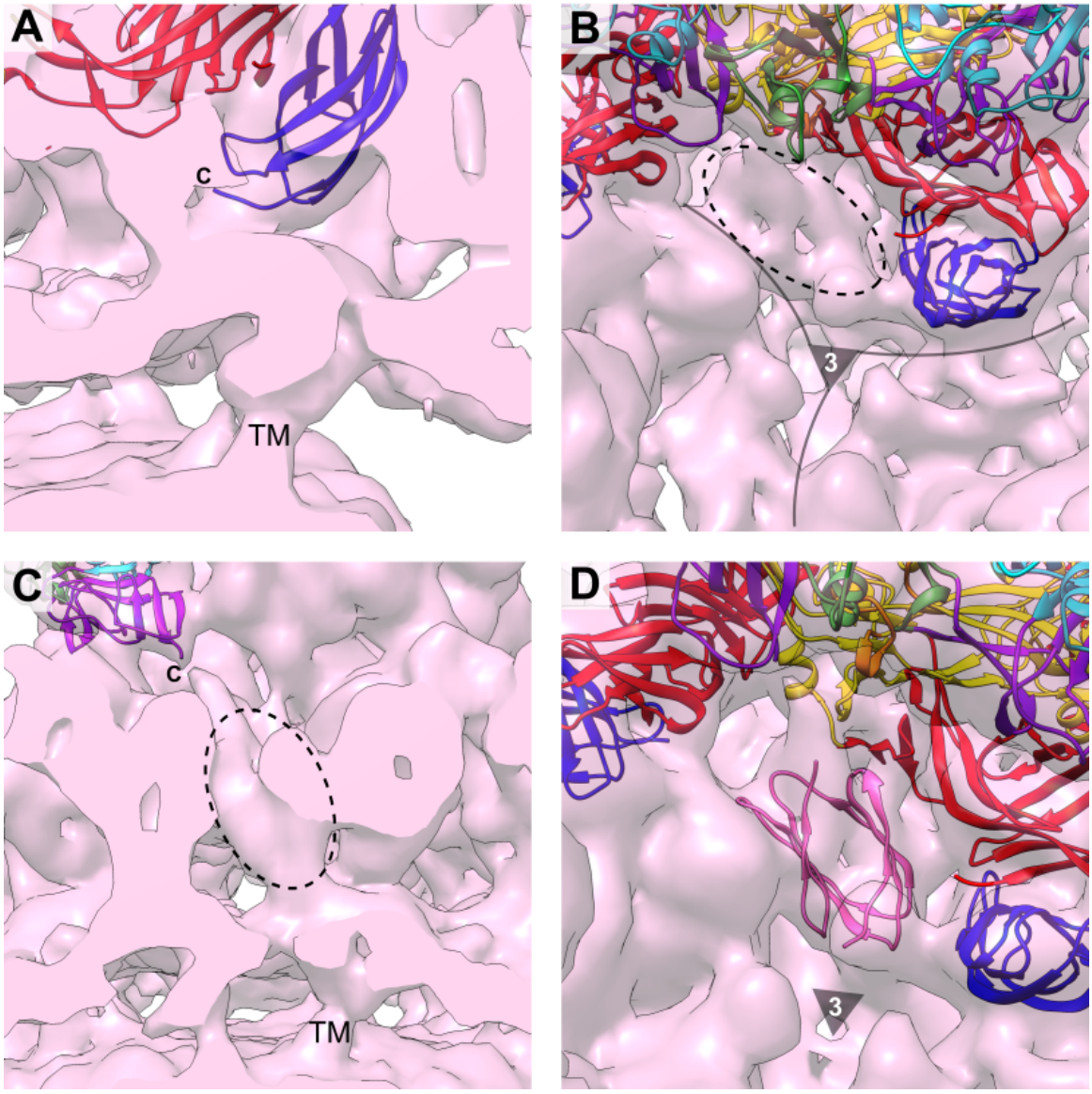


Figure 5.7 C-termini of Gc and Gn explored in the fitted map.
The validity of the fitted structures was assessed by looking at the C-termini of the Gn and Gc in the context of transmembrane helixes and vacant densities. The pentamer density is shown as an example. The C-termini are marked with C and transmembrane helixes are marked with TM A) A side view of the pentamer showing a single fitted Gc. The C-terminus of the Gc appears near the outer membrane leaflet of the viral envelope and a transmembrane helix bundle. B) A top view of the quasi-three-fold axis highlighted with a triangle and the number 3 and the interface between neighbouring capsomers is shown with transparent lines radiating from the three-fold axis. The full atomic model of the pentamer is shown, which reveals a vacant density at the quasi-three-fold axis, highlighted with a dotted circle. C) A side view near the quasi-three-fold axis of a sectioned pentamer map with a single fitted Gn. The C-terminus if the Gn points into the vacant density highlighted in panel *B*. The vacant density is directly above a transmembrane helix bundle. D) The structure of domain C of E2 from CHIV was fitted into the vacant density by rigid body fitting.

Lastly, the location of glycosylation sites was examined. There are four N-linked glycosylation sites found in our model (N438, N794, N1035, N1077, Figure **5.8A**). If the model is correct, the sites are expected to be either solvent exposed or close to a density that may encompass a glycan. The single N-linked glycosylation site predicted for the Gn, N438, is located on domain B and according to the model the glycan should point into the centre of the capsomer (Figure **5.8B**&**C**). The Gc contains one glycosylation site on domain II and two sites on domain III. N749 on domain II is surface exposed and points into the same central cavity as the Gn glycan (Figure **5.8B**). The glycans of domain III however are found in the inter-capsomer space near the quasi-three-fold axis (Figure **5.8D**&**E**). N1035 appears to be surface exposed, but N1077 is located deeper in the glycoprotein layer and a small globular density may account for it.

Overall, the pseudo-atomic models of the capsomers are in a good agreement with the EM maps. Despite the large conformational changes of the Gc, the secondary structural elements agree particularly well with the Gc fit. Furthermore, cross-correlation values, unoccupied densities, the location of the C-termini and glycosylation sites all favour the placement of the two glycoproteins.

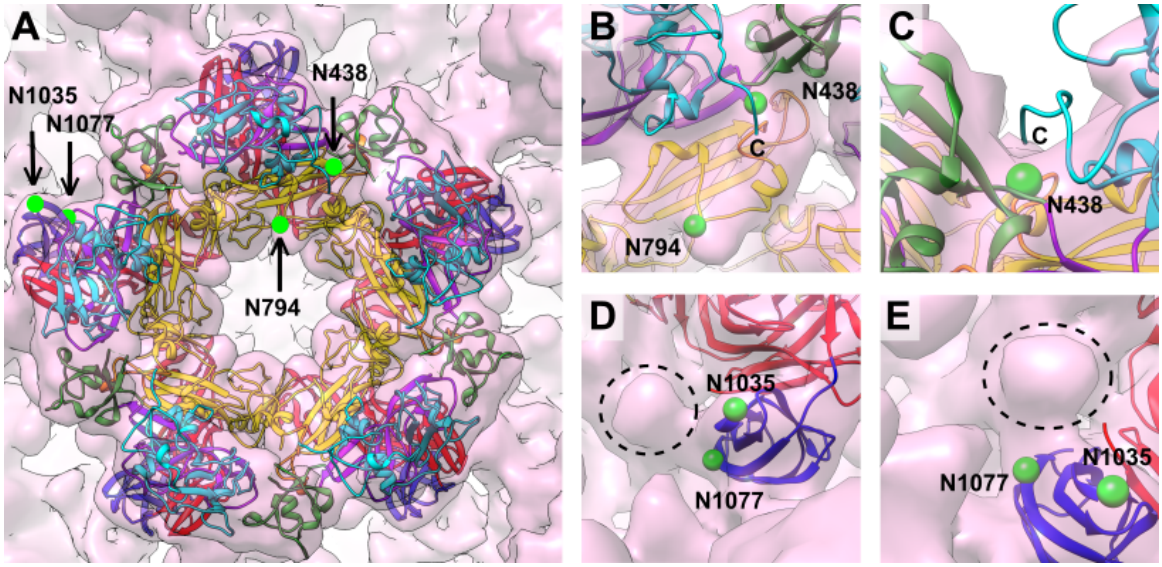


Figure 5.8 Predicted N-linked glycosylation sites on the capsomer assembly.
A) An overview of the pentamer model within the localized reconstruction density (shown in pink). Glycosylation sites are represented by a green sphere. B) The N438 and N794 glycosylation sites face into the central cavity of the capsomers. N794 is surface exposed and a small density is visible adjacent to the site facing into the cavity. A glycan at N438 of Gn may overlap with the Gn C-terminus. C) From another perspective, the density at the C-terminus is not resolved which may indicate flexibility. D) and E) N1035 and N1077 are found in the inter-capsomer region. N1035 is solvent exposed while N1077 is buried in the density. A small globular density is found close to the N1077 which may be occupied by the glycan, but structural details at this site is lacking.

5.5 Structural analysis of the capsomer pseudo-atomic models

5.5.1 The capsomers and intra-capsomer interactions

Overall, all the distinct capsomers assemble in a similar mode. The Gn-Gc heterodimers engage in a left-handed fashion to form the ring like capsomer structures (Figure **5.9**). Angles between heterodimers of the same capsomers are symmetrical and follow the local symmetry operator, i.e. two adjacent angles of the type 3 hexamer are repeated with a C3 symmetry (Figure **5.9**). The largest differences between any two angles are 54º (within the type 1 hexamer) and 72º (within the pentamer), a real testament to the flexible nature of the glycoprotein assembly. The angles at which the glycoproteins rise from the membrane is also distinct between distinct capsomers (Figure **5.9**). Most of the Gc's within a capsomer

rise at a similar angle (±1º). Unsurprisingly, the Gc's of the pentamer rise the highest from the membrane at an angle of 43º, while the Gc's from the type 3 hexamer only rise at an angle of 37º.

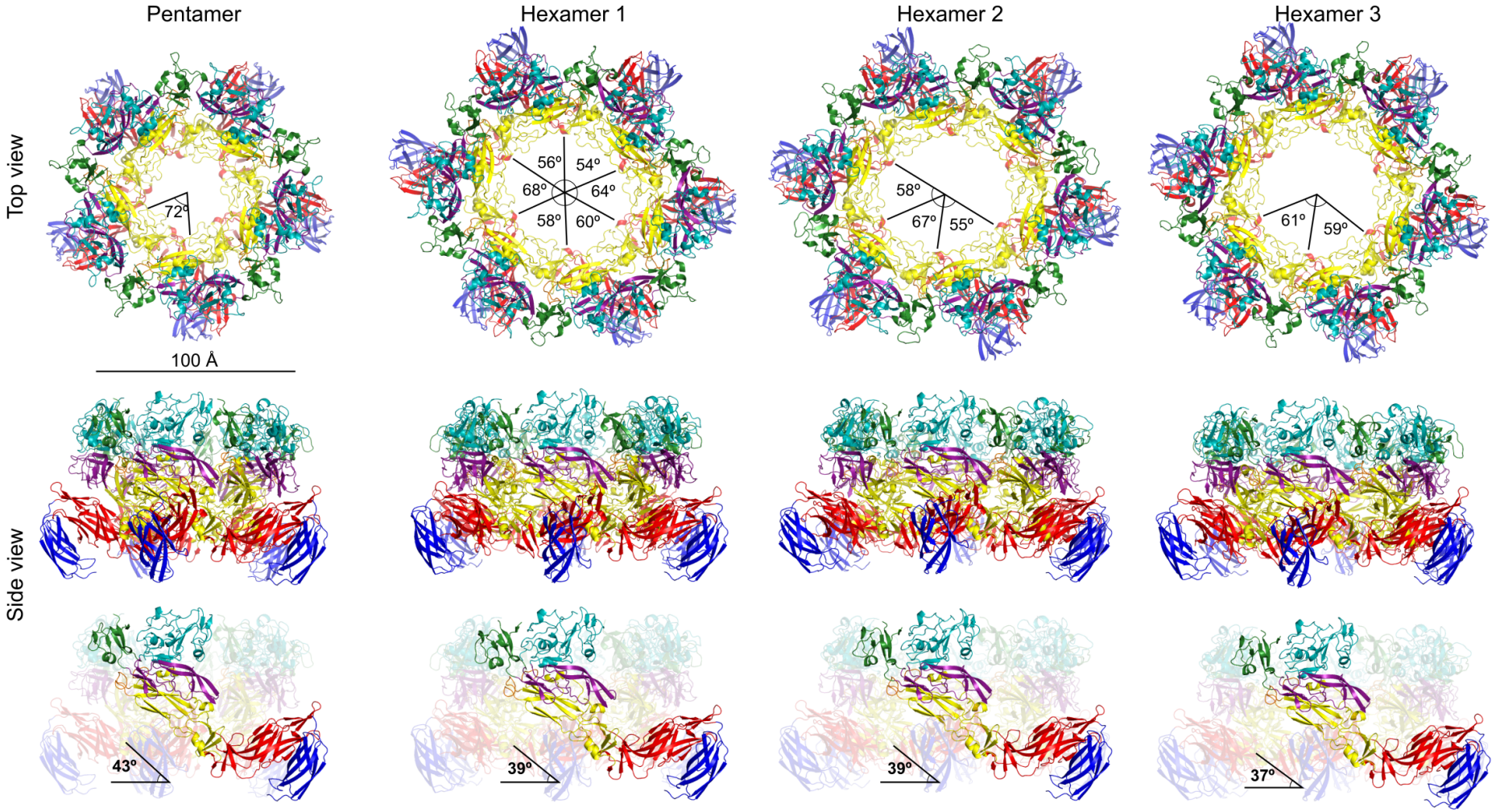


Figure 5.9 Overview of all the pseudo-atomic models of the distinct capsomers
The four distinct pseudo-atomic models of each capsomer type are shown in cartoon from a top view or a side view. Angles between neighbouring heterodimers are shown in the top panel. The bottom panel highlights a single heterodimer and shows the angle at which domain II of Gc rises from the viral membrane.

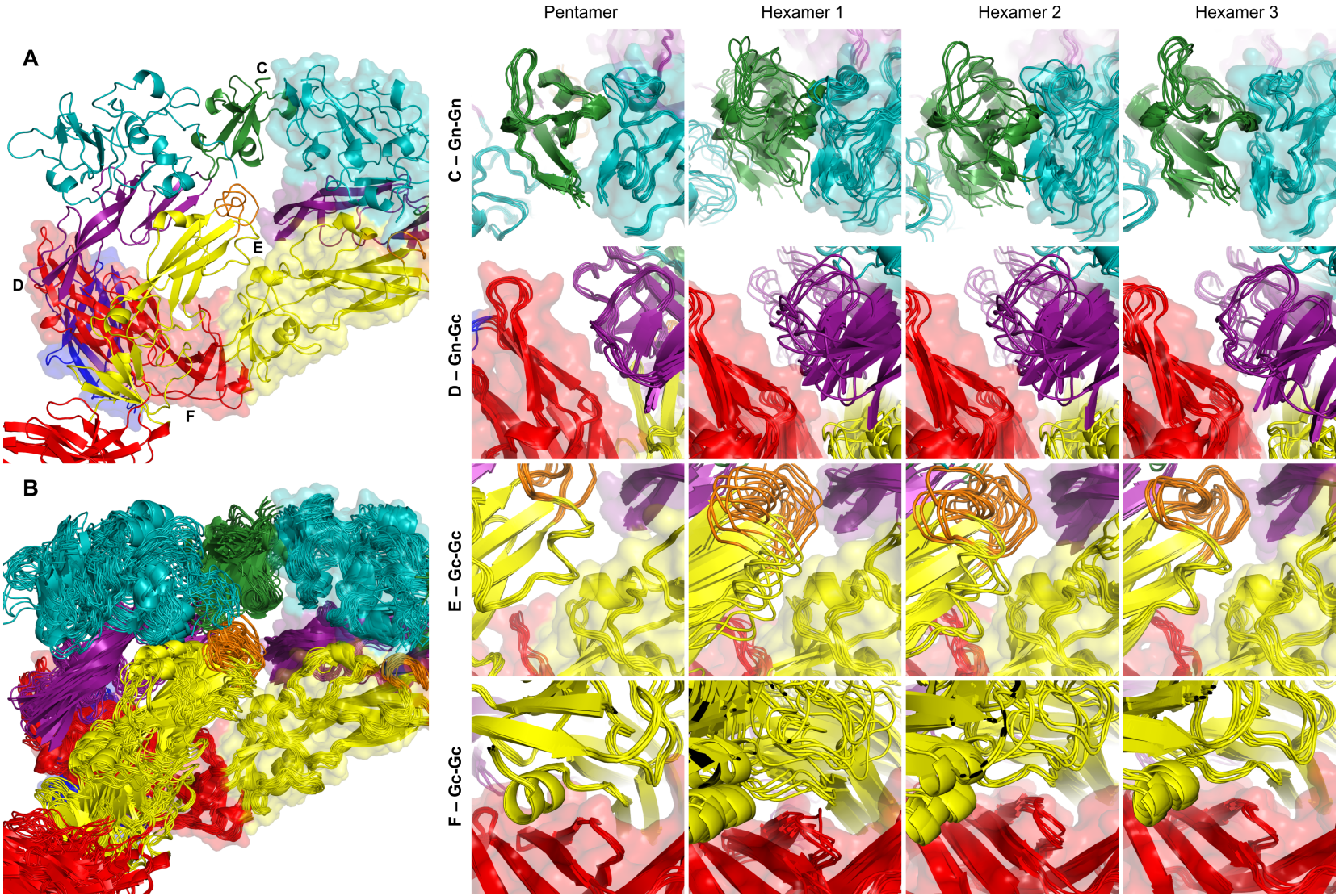


Figure 5.10 Inter heterodimer contacts.
The contacts made between heterodimers vary depending on the capsomer. A) An overview of the four areas of contacts, highlighted with a dotted circle. B) All pairs of adjacent heterodimers aligned to one heterodimer shows the breadth of variability. C-F) Zoom-in panels of the four areas of contacts from all capsomer types.

Inter-heterodimer contacts are made at four sites which involves all the domains of Gn and all but domain III of the Gc (Figure **5.10A**). Domain B and A of neighbouring Gn's contact via a small 3-10 helix on the Gn. The indented corner which Gc forms is an area of major contacts with domain II of a neighbouring Gc. Additionally, the β-ribbon of Gn makes contacts with domain I of a neighbouring Gc. Interestingly, the contacts vary considerably between the distinct capsomers. When all pairs of adjacent heterodimers are aligned together, such that only one heterodimer in each pair is aligned, the breadth of the interaction is revealed (Figure **5.10B-F**). The contacts within the pentamer and the type 3 hexamer, respectively, are uniform (despite the C3 symmetry of the type 3 hexamer) while the type 1 and 2 hexamer contacts are highly variable. This highly variable mode of engagement contributes substantially to the structural variability of the capsomers.

5.5.2 The heterodimer is a stable but flexible unit

To see if the heterodimer itself is the source of variability or stability, an all-against-all alignment was carried out on all heterodimers and an RMSD matrix was produced (see Apendix **10.13**). The RMSD of all heterodimers gives an average value of 1.53 Å (Table **5.1**). The same analysis shows that the variation between heterodimers within the same capsomer type is much less for the pentamer, while the variation within each hexamer is greater. When the Gc and the Gn are analysed separately, a similar pattern arises. With this analysis, areas of flexibility within each glycoprotein were identified. Domain B of the Gn appears highly mobile and can move 7.5 Å (Figure **5.11**). Similar hinges are observed between all the domains of the Gc, particularly the angle between domain I and II has a range of 89–95º (Figure **5.11**). Interestingly, domain B and A are consistently splayed away from each other in the pentamer but the angle between domain I and II is the narrowest in the pentamer.

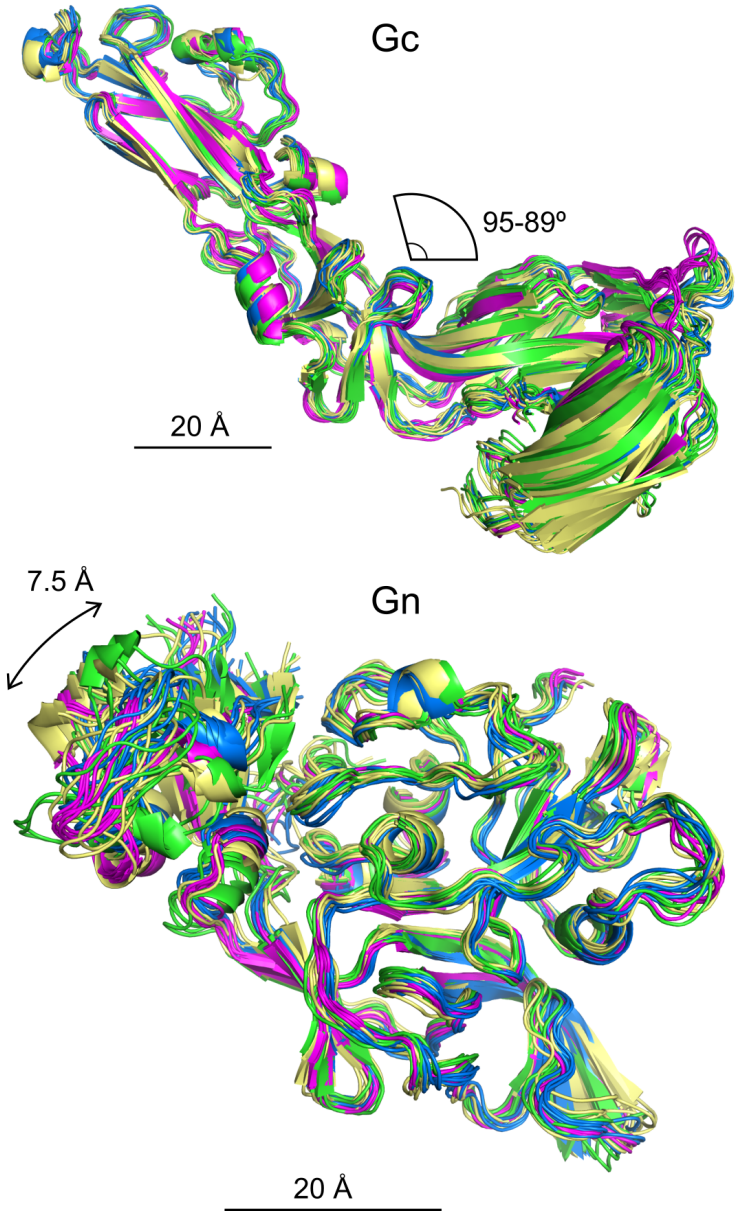


Figure 5.11 Flexible regions within the glycoproteins
An all-against-all alignment was performed in PyMOL on all copies of Gn and Gc from each capsomer. Each glycoprotein was coloured according to the capsomer type of origin: cyan for the pentamer, yellow for the type 1 hexamer, green for the type 2 hexamer and blue for the type 3 hexamer. The two panels show the aligned glycoproteins overlaid. In the top panel, the range of angles observed at the hinge between domain I and II is shown. In the bottom panel, the displacement of domain B is shown

On the scale of the capsomers, RMSD differences of 1-2 Å between heterodimers do not appear so large considering the differences in the inter-heterodimer contacts (Figure **5.10**). One might suspect that these differences arise from random noise in the MDFF runs

or errors in initial model building. To address this uncertainty, a clustering algorithm was used to classify the heterodimers. The all-against-all alignment produced a pairwise alignment RMSD matrix which was fed into a hierarchical clustering algorithm. The resulting heatmap shows that the heterodimers form distinct clusters (Figure **5.12**), most of which can be explained by their capsomer type. All the pentamer heterodimers cluster tightly while most of the hexamer heterodimers cluster into a separate group. The type 3 hexamer heterodimers cluster into two distinct sub-clusters which can be explained by the local symmetry of the type 3 hexamer. There are two outliers, Hex1_D and Hex2_D, which somehow seem to be similar to almost all other heterodimers. Overall, the clustering of the heterodimers shows that indeed the difference between heterodimers is not random but has an organisation which can in part be rationalised by the constraints imposed by the local symmetry of each heterodimer. The Gn-Gc heterodimer can thus be viewed as a flexible yet stable unit of the RVFV glycoprotein layer.

Table 5.1 Average root-mean-square deviation (RMSD*)
An all-against-all alignment in Pymol of the Gn-Gc heterodimer.

	Heterodimer		**Gn**		**Gc**	
	RMSD	**STDEV****	**RMSD**	**STDEV**	**RMSD**	**STDEV**
All capsomers	1.59	0.43	1.12	0.34	1.33	0.55
Pentamer	0.60	0.03	0.53	0.28	0.43	0.22
Hexamer 1	1.47	0.29	1.03	0.56	1.08	0.53
Hexamer 2	1.60	0.59	0.91	0.17	1.00	0.53
Hexamer 3	1.19	0.36	0.94	0.51	0.81	0.43

*RMSD is measured in Å
**STDEV = standard deviation

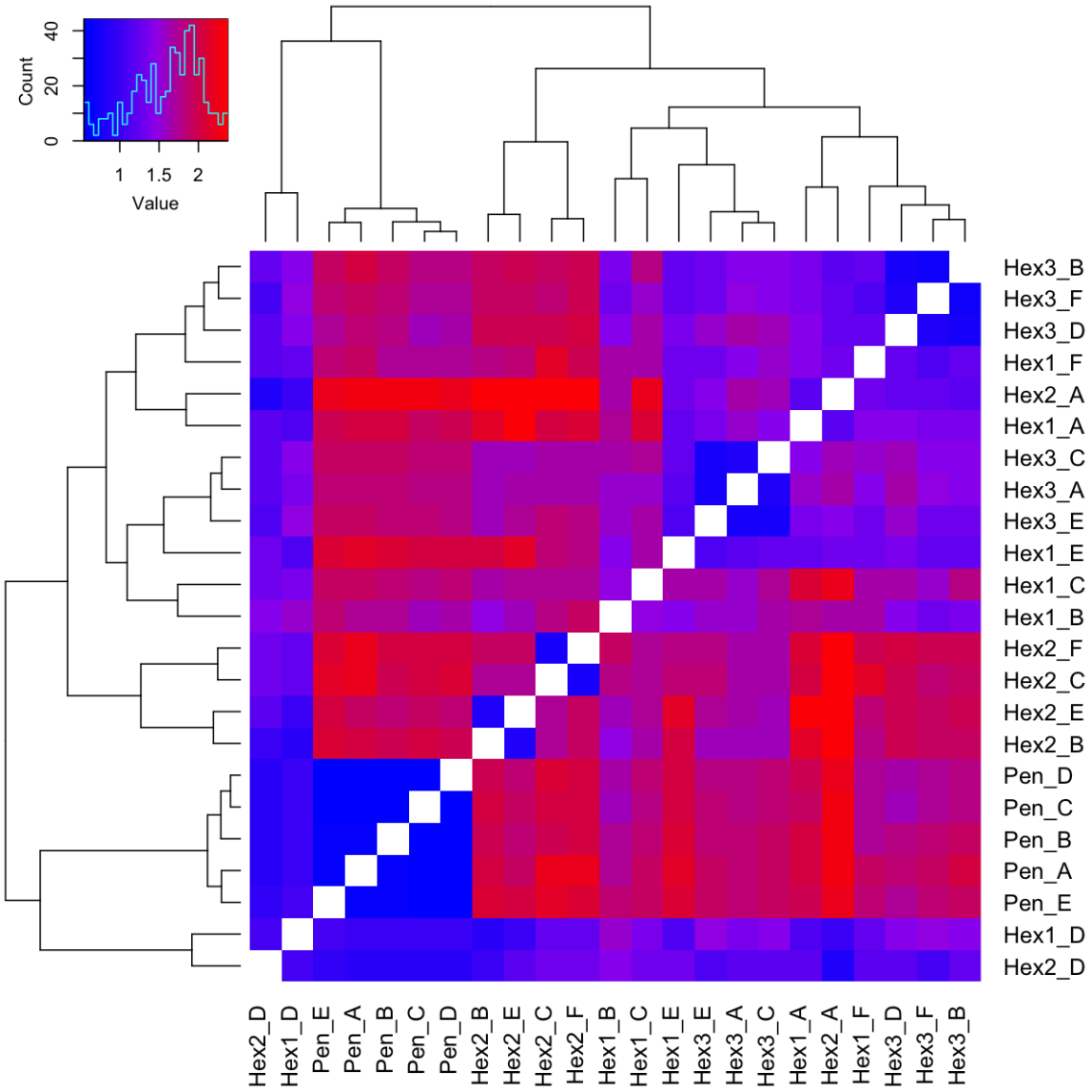


Figure 5.12 Clustering of heterodimers based on pair-wise alignment.
The all-against-all alignment matrix for the heterodimers was fed into hclust in the statistical package R [152] to generate hierarchical clustering and a heatmap. The small panel on the top right shows the colour key and a histogram of values (in RMSD). Each heterodimer is labelled according to originating capsomer and relative position, e.g. Hex2_A is a heterodimer from the type 2 hexamer and is the first structure in the PDB file. The clustering algorithm produces a tree of relatedness which is shown both above and left of the heatmap.

5.5.3 Shielding of the fusion loop

The main interaction between the Gn and the Gc is focused on the β-ribbon of the Gn and domain II of the Gc which has a buried surface area of ~950 Å^2. The fitting suggests that

the interface is characterized by a continuous six-stranded β-sheet formed by three β-strands from Gn and three from Gc, in addition to several short interacting α-helices from both proteins (Figure **5.13A&B**). The interface also includes the fusion loop of the Gc which interacts with the domain A and B interface of the Gn (Figure **5.13C**). The result is effective shielding of the fusion loop from solvent. Interestingly, this is reminiscent of the shielding of the CHIV E1 fusion loop which also involves the interface between domain A and B of E2 (Figure **5.13D**). These findings are in line with the secondary structural similarities discussed in Chapter **3.4** which indicated that the RVFV Gn and the CHIV E2 originate from a common ancestor protein. This suggests that the role of the Gn is indeed shielding and chaperoning the Gc to ensure fusion fidelity, much like the role suggested for the E2 [78].

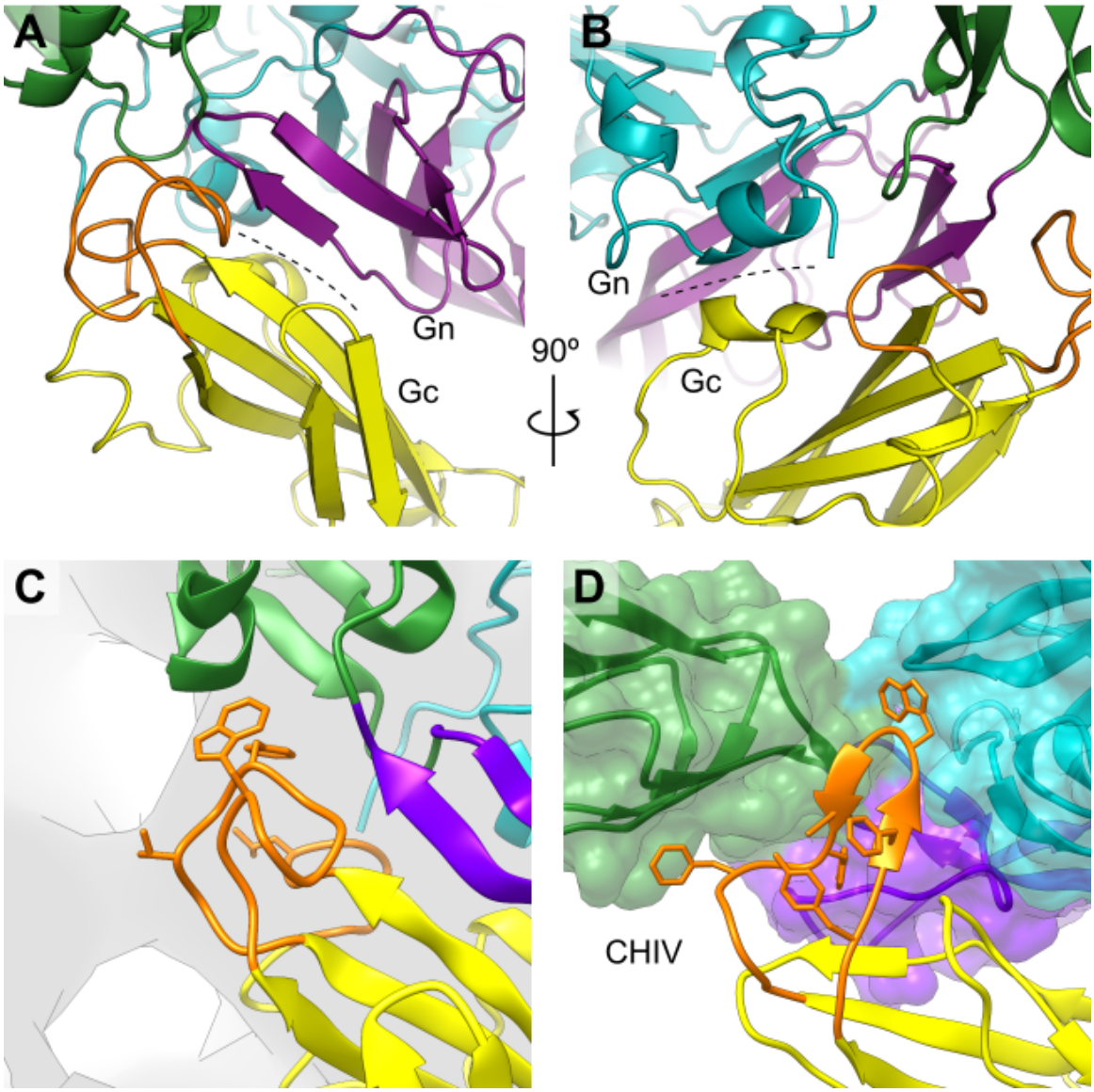


Figure 5.13 The heterodimer interface and fusion loop shielding.
A) The main site of interaction between the Gn and Gc involves the β-ribbon of the Gn and the tip of domain II. A three stranded β-sheet from each protein interact to form a single six stranded β-sheet. The interface is shown with a dotted line. B) The interaction is additionally stabilised by a helix-helix interaction adjacent to the β-sheet. Here domain A of the Gn contributes to the interface. C) The fusion loop of the Gc is shielded by the Gn at the interface between domain B and the β-ribbon. A map of the pentamer is overlaid and shown in grey. D) The fusion loop of the E1 is similarly shielded by the Gn at the interface between domain B and the β-ribbon and additionally domain A [78].

5.6 The pseudo-atomic model of the entire RVFV

5.6.1 Creating the assymetric unit

To probe the inter-capsomer interactions a model of the entire RVFV virion was created. The first step in creating the model was defining the asymmetric unit. Each capsomer map obtained from localized reconstruction was fitted into a corresponding density from the whole virion in Chimera, as well as the pseudo-atomic model associated with the capsomer. One heterodimer from each quasi-equivalent group was included in the asymmetric unit, i.e. one heterodimer from the pentamer, two from the type-3 hexamer, three from the type-2 hexamer and all six from the type-1 hexamer (Figure **5.14A**). An icosahedral symmetry was then applied to the asymmetric unit to create a model of the full virion (Figure **5.14B**).

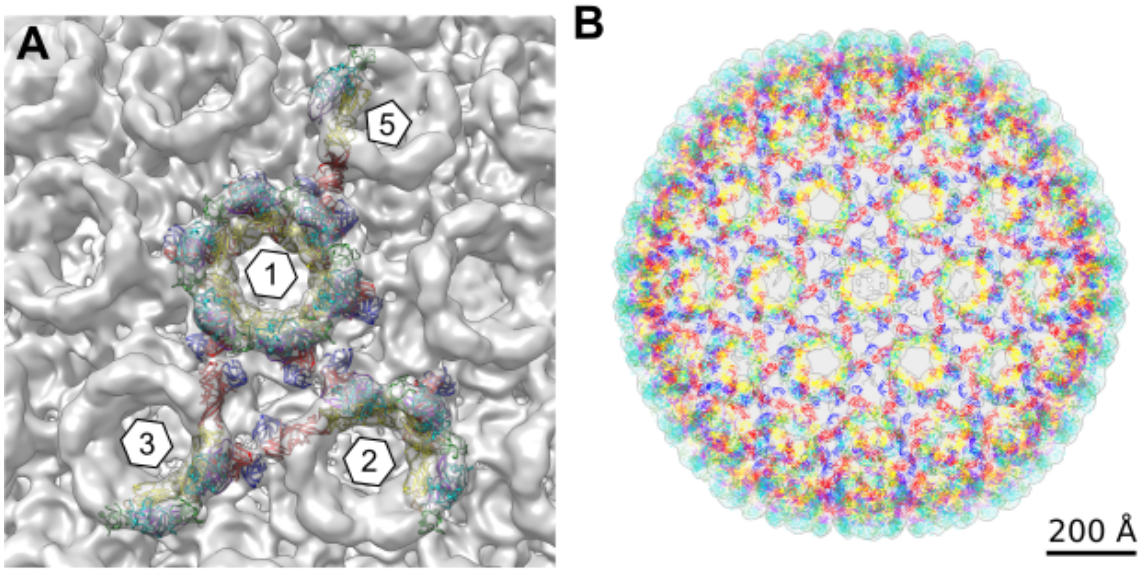


Figure 5.14 Pseudo-atomic model of the RVFV
A) The asymmetric unit of the RVFV contains heterodimers from all four capsomer types based on local symmetry: One from the pentamer, all six from the type 1 hexamer, three from the type 2 hexamer and two from the type 3 hexamer. The asymmetric unit is shown in the context of the full virion map shown in grey. B) The full pseudo-atomic model of the RVFV fitted into the full virion map.

5.6.2 Inter-capsomer interactions

There are two modes of interaction between all capsomers: The quasi-two-fold axis and the quasi-three-fold axis. Since these contacts were not probed directly by MDFF and they are based on the 13.3 Å map, their nature is not well defined. At both interfaces the putative domain C of the Gn and domain III of the Gc seem to be key players in forming inter-capsomer contacts. At the quasi-two-fold axis, domain I of the Gc is additionally featured (Figure **5.15A&B**). The interactions at the quasi-three-fold axis involves alternating domain C and domain III contacts (Figure **5.15C&D**). These are the same contacts as seen on the quasi-two-fold axis (which is adjacent to the quasi-three-fold axis) but they are in fact organised in a three-fold symmetrical way. What is interesting about the quasi-three-fold axis is the concentration of transmembrane helix bundles. It is difficult to say anything concrete about the quasi-three-fold axis as we lack a structure of the full ectodomain of the Gn, but as discussed below, one can speculate that the quasi-three-fold axis plays a role in the assembly of the capsomer.

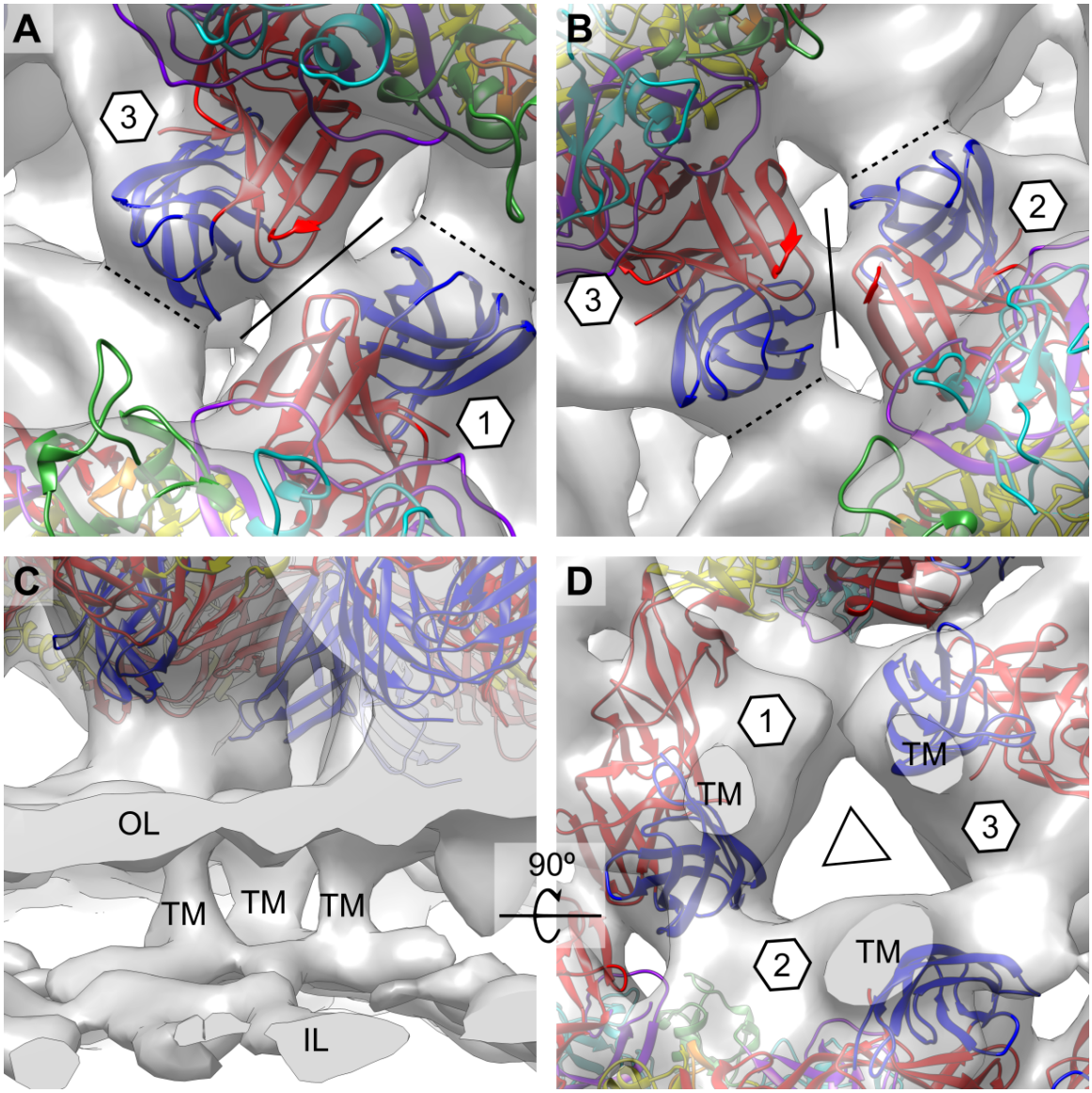


Figure 5.15 Inter-capsomer interactions.
Details from the pseudo-atomic model of the entire virion. The map shown in gray is the 13.3Å icosahedral reconstruction. A) An example of the quasi-two-fold axis between the type 1 and type 3 hexamer. The interface involves domain I and III of the Gc and is shown with a solid line. Additionally, the density occupied by the putative domain C of the Gn seems to interface with domain III of the Gn. The interface is shown with a dotted line. B) The same interface as in panel *A* but between the type 2 and type 3 hexamers. C) The quasi-three-fold axis shown from a side view. Three transmembrane helix bundles (TM, each containing three helices) are visible between the inner and outer leaflet (IL and OL) of the viral envelope. D) The quasi-three-fold axis viewed from inside the virion. The envelope has been segmented away and cut through the three transmembrane helix bundles. The three-fold axis is marked with a triangle. The putative domain C of Gn (not specifically shown but occupies the vacant density where the hexamer labels have been placed) is expected to connect the three domain IIIs of the Gc's with a quasi-three-fold rotational symmetry.

5.7 Concluding remarks – flexibility, simplicity and complexity

5.7.1 Limitations of our current model

In this chapter, the pseudo-atomic model of the RVFV was presented. The model was based on MDFF simulations of the crystal structures of the Gn and the Gc, fitted into each of the four capsomer densities. The quality of the fits was assessed by secondary structure location, C-termini and glycosylation sites. The resulting models were characterised by a flexible but stable heterodimer and heterogenous inter-heterodimer contacts. Additionally, the model indicates that the Gn functions as a fusion chaperone, much like the E2 from CHIV, by shielding the Gc fusion loop.

Despite the analysis presented in this chapter, there are some important limitations to consider. Firstly, only parts of the ectodomains were modelled, roughly 65 % of the total mass of protein present on the viral envelope is modelled. Secondly, our resolution limits us to just under 10 Å so anything beyond the placement of bulk secondary structural elements is uncertain. Thirdly, the MDFF was run with secondary structure restraints and the secondary structure of Gn and Gc on the virion surface may be dynamic or different from the secondary structure captured by crystallography. However, this is the most informative model of any phlebovirus to date and despite the shortcomings we can certainly learn something about assembly and entry mechanism of RVFV.

5.7.2 Why all the flexibility?

As mentioned earlier in this chapter, the virion assembly is highly varied. The heterodimer itself shows a level of variability, but as the clustering revealed this internal flexibility is not random and can be related to the local symmetry. However, the major source of variability within the capsomers is the inter-heterodimer interaction. Of course, our simulation may not be able to capture exactly what happens at these interfaces but it is

striking to see how much variation there exists. A more thorough analysis of the inter-heterodimer interactions may be the key to understanding how the virion assembles and how flexibility is accommodated.

So why all this flexibility of assembly? Other icosahedral viruses which contain a class II fusion protein seem much less flexible, such as CHIV and DENV. One possible explanation relates to the virus T-number. An analysis of viral capsid T-numbers suggested that with increasing number of distinct hexamers (hexamer complexity) comes less stability to the viral capsid [153]. The authors constructed a periodic table of viral capsids which suggests that some T-numbers are less stable than other because they require the capsid proteins to adopt a large range of conformations to allow for the hexamer complexity imposed by the T-number. A T=12 capsid is considered particularly unstable and has a high hexamer complexity number and such capsids are rarely seen in nature. The RVFV does not have a capsid, but rather an envelope with glycoproteins arranged in a T=12 icosahedron. It may be that the presence of an anchoring membrane allows the glycoprotein layer some flexibility. If a particular interaction is unstable, the membrane is still going to be present and will keep the whole virus intact despite some local deviations from the assembly. As the analysis in this chapter has shown, the RVFV capsomers are distinct and the assembly is flexible. It may simply be that a T=12 icosahedron is constrained in such a way that in the absence of accessory proteins, assembly is unstable. It is quite remarkable that a single heterodimer unit is even capable of forming a T=12 icosahedron but maybe such a constrained assembly is seen here in an enveloped virus because of the anchoring membrane and the inherent flexibility of the glycoproteins.

5.7.3 Comparing RVFV assembly models

Two attempts have been made at placing crystal structures from RVFV into a map of the entire virion [47, 101]. Each of them involved only one of the glycoproteins and used the 22 Å map [90]. What these two models and the model presented in this thesis have in common, is that the Gn is assumed to occupy the top layer of the capsomer ring and the Gc the membrane-proximal layer. Unfortunately, neither of these models are available in the PDB for direct comparison.

The first model used the pre-fusion head-to-tail dimer of the Gc [47]. The fitting suggested that the inter-capsomer contacts are mediated by domain II and the core of the capsomer ring is made up of domains I and III. This is in stark contrast to the model presented in this thesis where domain III is found on the capsomer interfaces and domains I and II form the base layer of the capsomer ring.

The second model used the crystal structure of the Gn only, despite the crystal structure of Gc being available [101]. The glycoprotein layer of phleboviruses was suggested to be stabilised by homodimeric interactions of the Gn [101]. The model suggests that at a quasi-two-fold axis between two capsomers, the Gn dimerises through a series of disulphide bonds at the stem region of the Gn, i.e. the putative domain C. In the model presented here, the putative domain C of Gn is present at the quasi-two-fold axis, but the domains do not form a homodimeric contact (Figure **5.15A&B**). The cysteine residues involved in the dimeric contacts are localised near the transmembrane region which is directly below the quasi-three-fold axis, as mentioned above, and is a place of major interactions between domain III of Gc and the putative domain C of the Gn (Figure **5.15C&D**). Since structural data of this region is limited this is purely speculative, but one can theorise that a Gn trimerisation model is just as likely as a dimerisation model given these observations of the three-fold axis. The dimerisation model is based purely on data

from recombinant proteins but not the whole virion. It is easy to imagine how a trimeric interaction involving four disulphide bonds on a virion surface can be artificially shifted to a dimeric interaction *in vitro* (Figure **5.16**). In a Gn trimeric model, upon acidification, the Gn-Gc heterodimer would dissociate and the Gns would move towards the three-fold axis as the Gcs are released (Figure **5.17**). A similar theory of Gn-Gn interactions has been postulated for a hantaviral Gn crystal structure which was captured in a low-pH state [154]. Here the low-pH was speculated to favour oligomerisation or cluster of the Gn which might be a mechanism of freeing the Gc fusion proteins on the virion surface. In our model, as the Gn proteins move towards the quasi-three folds, the Gcs have ample space to rise vertically from the viral membrane and insert the hydrophobic fusion loop into a target membrane [48]. In this model, at any hexamer, there would be six Gc molecules in close proximity to each other and according to fusion experiments in flaviviruses, at least two fusion protein trimers are needed for the merger of viral and liposomal membranes [85, 86].

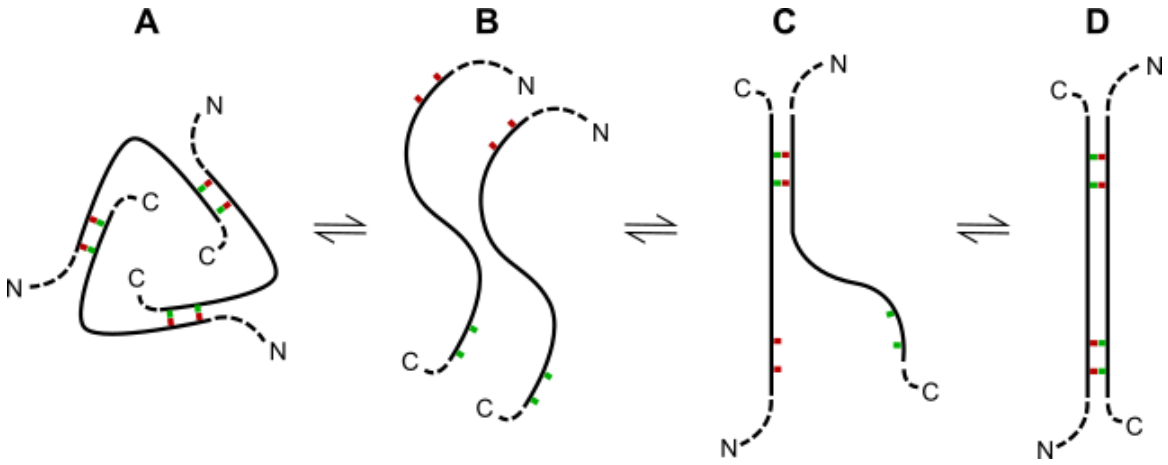


Figure 5.16 Transitioning from a trimer to a dimer.
Let us imagine a trimeric interaction involving a flexible part of a protein with four cysteines per protein involved in disulphide bonds (panel *A*). The three parts contact each other in a C3 symmetrical way which involves two disulphide bonds at each interface. The letters C and N represent the C- and N-termini, respectively. This form of interaction may be favourable due to steric constraints imposed by the presence of transmembrane regions, contacts made by N-terminal or C-terminal portions of the protein or other proteins present. If this region would be taken out of context, e.g. expressed as a soluble fragment, we can imagine the monomeric form in solution (panel *B*). When two fragments come together they may start to form disulphide bonds at one of the two possible sites (panel *C*). Given the proximity and the reduced degrees of freedom of the second site on each fragment, it is much more favourable that the dimeric contacts are perpetuated to the second site rather than a third fragment coming into play. This would result in the formation of a homodimers which would be the predominant species in solution (panel *D*).

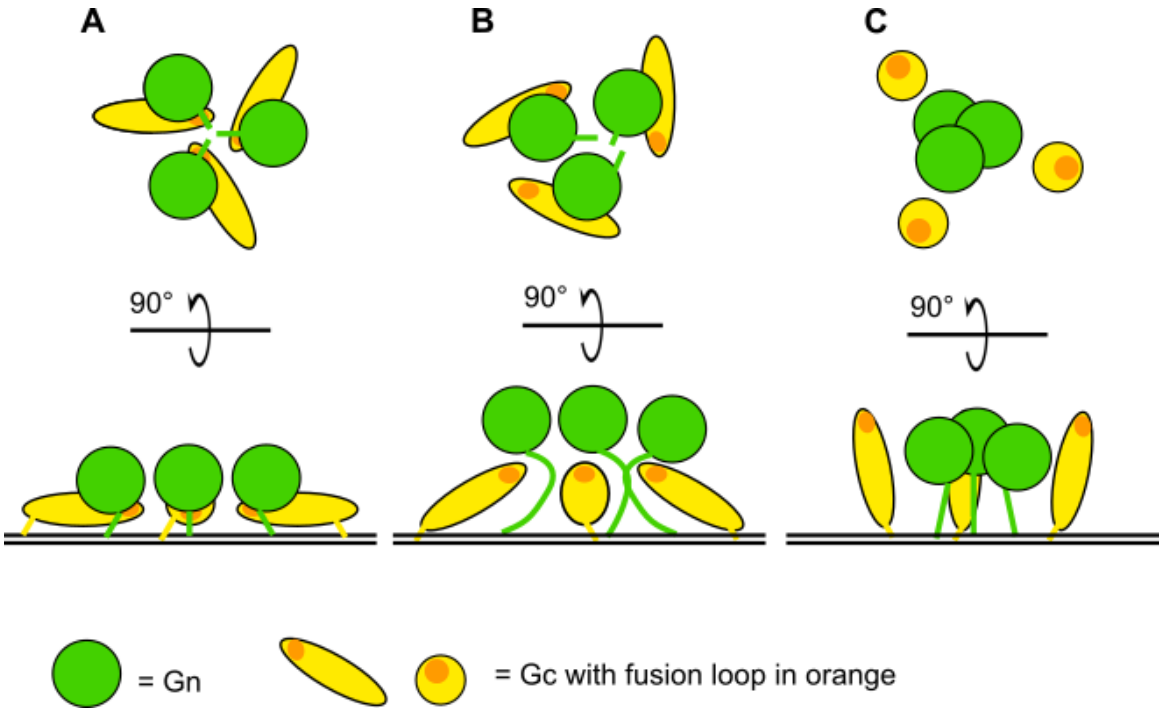


Figure 5.17 Fusion activation model.
Let us imagine the quasi-three fold axis between three capsomers (panel *A*). Each panel has a top view in the upper half and a side view in the lower half where the viral membrane is depicted with a double line. At neutral pH the Gn shields the fusion loop from the extracellular environment. Upon acidification, the heteordimers dissociate from one another and the Gn-Gc contacts are disrupted (panel *B*). If we assume that the Gns form covalent trimers, as the Gns are pushed away from the Gcs, the Gns would cluster together at the quasi-three-fold axes. The released Gcs are then free to rise vertical from the viral membrane and interact with a target membrane via their fusion loops (panel *C*).

A model based on trimerisation of Gn could also explain virion assembly. Since the Gn and Gc are transported as heterdimers from the endoplasmic reticulum to the Golgi where assembly takes place, it is likely that the trimerisation of Gn is in fact a trimerisation of heterodimer. Trimers or triangles can be thought of as the basic building blocks of any icosahedron so it does not take too much imagination to see how trimeric heterodimers of Gn and Gc could form an icosahedral virion. And certainly, a flexible mode of interaction between heterodimers could arise from the trimer being the stable unit, since the flexible interactions would in fact be inter-trimer interactions.

Currently, this model it not backed by much data and is speculative. No trimers of heterodimer have been isolated from virions or cells infected with the virus, as far as the

author is aware. It remains to be seen if any data comes up in the future which could support or refute this hypothesis. A higher resolution map of the viron assembly could tell us more, but this remains to be addressed in further studies.

6. The crystal structure of Gc from SFTSV in the post fusion conformation reveals a conserved class II membrane fusion mechanism

6.1 Foreword

Much of the content presented in this chapter has been published [155] and several of the figures bellow are taken from that publication. The work presented here sheds a light on the entry mechanism of the phleboviruses which is a common theme throughout this thesis. This chapter is presented as the last results chapter since it deals with the late stages of phleboviral fusion in contrast to the pre-fusion shielded state presented in the previous chapter.

6.2 Summary

To further our understanding of the fusion mechanism of pheboviruses, we sought to solve the crystal structure of the ectodomain of Gc from SFTSV. The protein was transiently expressed in mammalian cells, purified, and crystallised. The structure of SFTSV Gc was solved to 2.45-Å resolution by SAD phasing. The protein crystallized in a three domain (I–III), trimeric post-fusion configuration. Comparison of the SFTSV Gc structure to the pre-fusion structure of the Gc homolog from RVFV shows that the fusogenic rearrangements of the phleboviral Gc appear to be analogous to those observed for the envelope fusion glycoproteins of alpha- (family *Togaviridae*) and flaviviruses (family *Flaviviridae*). These results indicate a conserved mechanism of membrane fusion between these otherwise non-related groups of viruses. Two putative fusion loops were identified, which are likely inserted into the host membrane during

host cell entry. Site-directed mutagenesis on the fusion loop of Gc was performed on the soluble Gc construct. The same mutants were generated by reverse genetics and rescue by collaborators and when taken together, these mutational analyses reveal that these residues are stringently required for the virus lifecycle. Histidines in domains I and III of the Gc were also identified as important residues for the virus life cycle with the same mutation analysis and, by analogy to RVFV, likely contribute to the pH-induced conformational rearrangements of the molecule. Combined, these data provide structural and functional evidence for a unified mechanism of membrane fusion between phlebo-, flavi-, and alphaviruses.

6.3 Expression and purification of the soluble SFTSV Gc ectodomain

6.3.1 Construct design and small scale expression trials

The previously reported structure of Gc from RVFV [47], also a phlebovirus, was used as a template to design a crystallisable construct of the ectodomain of SFTSV Gc. The two Gc glycoproteins have a sequence identity of ~25% and were therefore expected to adopt the same class-II fusion glycoprotein fold [156]. An initial construct of SFTSV Gc, termed V996, was designed with the same boundaries as the crystallized RVFV Gc (Figure **6.1A**) and cloned into the pHLsec vector for transient mammalian expression. For comparison, a construct of RVFV Gc mimicking the crystallised construct was cloned into the same vector. A small-scale expression trial was conducted in HEK293T cells in a 6-well format and the cell supernatant was probed for expression by Western blot analysis (Figure **6.1B**). Comparison of SFTSV Gc V996 with the RVFV Gc positive control revealed that SFTSV Gc V996 was not a highly expressing construct. A better expressing construct was sought and several SFTSV Gc constructs were designed, which had varying C-terminal lengths (Figure **6.1A**). These constructs were also assessed by Western blot analysis of the supernatant resulting from a

small-scale expression trial (Figure **6.1C**). The expression of the SFTSV Gc appears to be boosted in a step-wise manner as the length of the C-terminus is increased. The two longest constructs, SFTSV Gc R1035 and SFTSV Gc D1023, express to higher levels with respect to the RVFV Gc construct and the expression of SFTSV Gc V996 in comparison seems negligible on a Western blot.

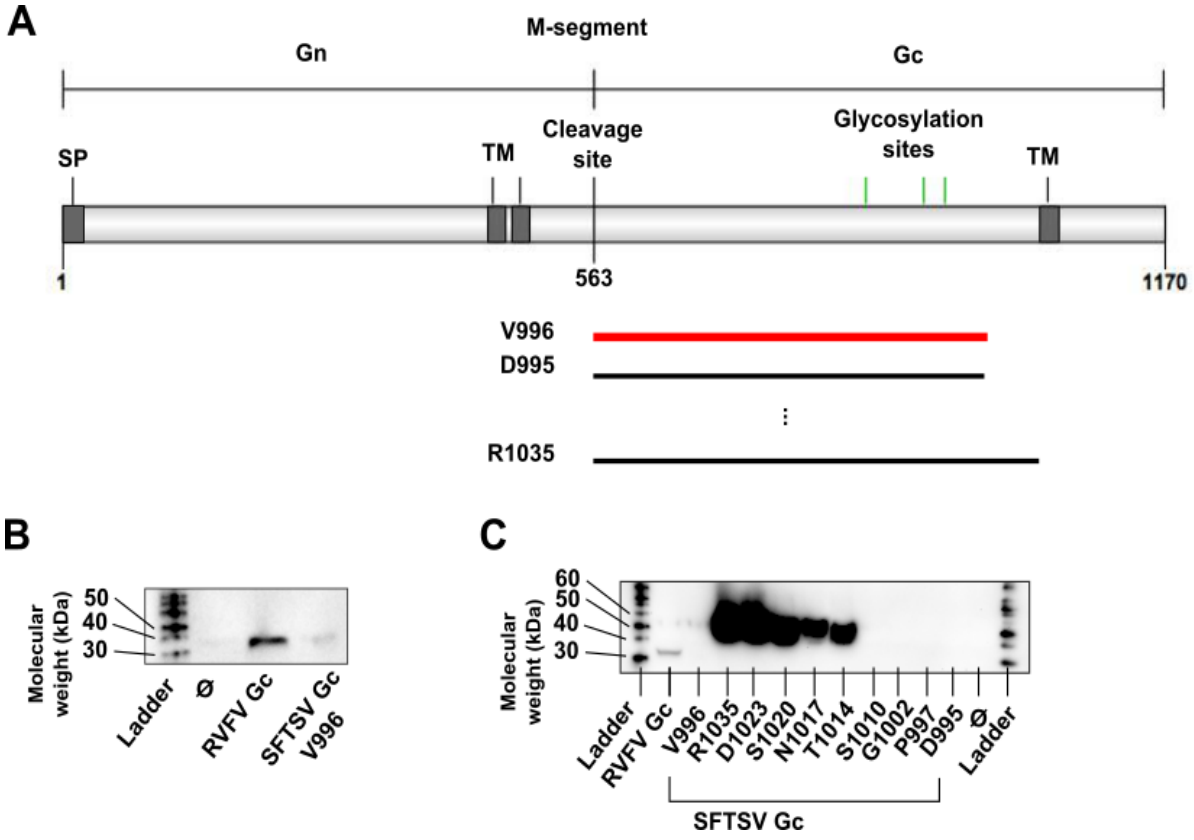


Figure 6.1 Construct design and expression tests of SFTSV Gc.
A) A domain diagram of the M-segment of SFTSV shows the arrangements of Gn and Gc within the genome. Glycosylation sites, transmembrane regions (TM) and the signal peptide (SP) are highlighted. Gc constructs are labelled according to the C-terminal amino acid and its position in the genome and are represented by a line. Several constructs of varying C-terminal length were designed. B) An initial small-scale expression trial of SFTSV Gc V996 and the equivalent RVFV Gc construct. Compared to RVFV Gc, the SFTSV Gc V996 is a lower expressing construct. Ø is a negative control. C) A small-scale expression trial of a panel of SFTSV Gc constructs with varying C-terminal lengths. Expression of the Gc appears to be boosted in a step-wise manner by lengthening the C-terminus.

6.3.2 Large-scale purification

Despite the low expression levels observed in the small-scale expression trials, the SFTSV Gc V996 construct was tested for large scale expression. This particular construct was chosen since it matched the previously crystallised RVFV Gc construct and the C-terminus of class II fusion proteins are largely flexible and often unstructured in the crystal [70], such a flexible region may prove refractory to crystallisation. SFTSV Gc V996 was transiently expressed in HEK 293T cells in a roller bottle format. After 5-6 days following transfection, the protein was purified from the supernatant by diafiltration and immobilized nickel affinity purification. The protein was then partially deglycosylated with endoglycosidase F1 and subjected to a final size-exclusion chromatography purification (Figure **6.4A**). The V996 Gc eluted as a single peak on SEC, corresponding to a putative monomer of ~47 kDa (Figure **6.4B**). Total protein yield after SEC was estimated ~0.5 mg/L of tissue culture media

6.4 Crystallisation and structure determination

Tetragonally-shaped crystals of SFTSV Gc V996 became visible after 4 days in a condition from the pentaerythriotol screen [157] containing 45% (vol/vol) pentaerythriotol 426 and 0.1 M sodium acetate at pH 4.6. Crystals were allowed to grow for a further 15 days. The largest crystals grew to ~100 μM long in the longest dimension and ~30 μM in the other two dimensions (Figure **6.2A**). Crystals were flash-frozen in liquid nitrogen straight from the precipitant solution since the pentaerythriotol precipitant is also a cryo-protectant. X-ray data was collected at beamline I03 at Diamond Light Source. Data from two crystals was combined and processed to 2.45 Å resolution. Data collection and refinement statistics are found in Table **6.1**.

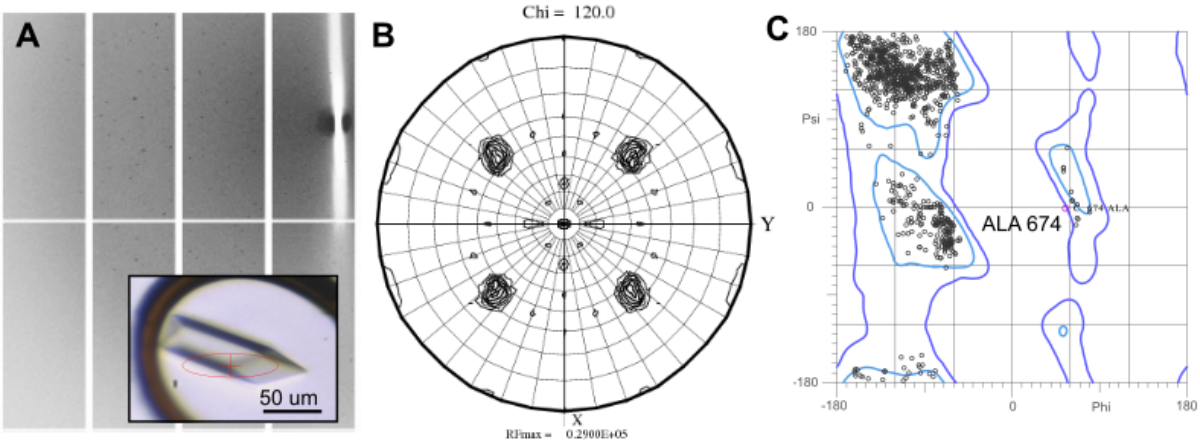


Figure 6.2 X-ray diffraction data.
A) An example of diffraction data collected for SFTSV Gc. The inset shows the crystal on the beamline and the aperture used. B) A self-rotation function showing peaks at 120°, indicating a three-fold rotational symmetry present in the crystal C) A Ramachandran plot of the final refined model generated by MolProbity [114]. The structure includes only a single outlier while 99.9% of all residues are either in allowed or most favoured regions.

Analysis of the Mathews coefficient (61% solvent content for three molecules) and self-rotation function were suggestive that the crystal contained three molecules in the asymmetric unit, as one would expect for the glycoprotein forming a 'post-fusion' conformation. Molecular replacement with Phaser [110] using individual protomers of RVFV Gc as well as known post fusion structures (DENV2 E PDB code 1OK8, SFV E1 PDB code 1RER, TBEV E PDB code 1URZ), as a search models, however, did not yield a solution despite exhaustive efforts. As a result, an experimental phasing approach was then adopted and crystals were soaked for 90 min in the precipitant solution saturated with K_2PtCl_6. X-ray data from the soaked crystal was collected on beamline I04 at Diamond Light Source at a wavelength of 1.072 Å, which was determined by a fluorescence scan of the crystal to correspond to the peak of the anomalous signal. Data from three crystals was combined to give high redundancy to maximise the anomalous signal. Initial phases and model were obtained from autoSHARP [108] in a SAD experiment. Initial refinement was performed using REFMAC5 [111] from the CCP4 suite, final refinement was performed using PHENIX refine [112] and manual model building was done with Coot [113].

Table 6.1 Crystallographic data collection and refinement of SFTSV Gc V996.

Parameters	K_2PtCl_6 SAD data	Native data
Data collection		
Beamline	I04, DLS	I03, DLS
Resolution range (Å)	108.62-2.89 (2.97-2.89)	108.51-2.45 (2.51-2.45)
Space group	$I2_12_12_1$	$I2_12_12_1$
Cell dimensions		
a, b, c (Å)	147.2, 152.3, 160.9	147.4, 152.3, 160.4,
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
No. of crystals	3	2
Wavelength (Å)	1.072	1.008
Unique reflections	40,765 (2,964)	66,293 (4,838)
Completeness (%)	99.9 (98.9)	99.8 (99.3)
R_{merge} (%)[a]	16.4 (63.7)	10.2 (108.4)
I/σI	22.1 (2.6)	15.7 (1.8)
Avg. redundancy	34.1 (6.0)	8.1 (7.8)
Refinement		
Resolution range		108.51-2.45 (2.51-2.45)
Number of reflections		66,125 (2,792)
R_{work} (%)[b]		19.0
R_{free} (%)[c]		23.2
RMSD		
Bonds (Å)		0.002
Angles (°)		0.500
Molecules per a.s.u.		3
Atoms per a.s.u (protein/carbohydrate/ water)		9,565/42/104
Average *B*-factors (Å^2) (protein/carbohydrate/ water)		81.3/103.6/62.5
Ramachandran plot (%)		
Most favoured region		96.9
Allowed region		3.0
Outliers		0.1

Numbers in parentheses refer to the relevant outer resolution shell.
r.m.s.d.: root mean square deviation from ideal geometry.
[a] $R_{merge} = \Sigma_{hkl} \Sigma_i |I(hkl;i) - \langle I(hkl) \rangle| / \Sigma_{hkl} \Sigma_i I(hkl;i)$, where $I(hkl;i)$ is the intensity of an individual measurement and $\langle I(hkl) \rangle$ is the average intensity from multiple observations.
[b] $R_{factor} = \Sigma_{hkl} ||F_{obs}| - k|F_{calc}|| / \Sigma_{hkl} |F_{obs}|$
[c] R_{free} equals the R_{factor} as calculated above but using against 5% of the data removed prior to refinement.

6.5 Structural analysis

6.5.1 Structural overview

A single 120 Å-long trimer of SFTSV Gc was observed in the asymmetric unit (Figure **6.3A**). The trimeric state is similar to the well-characterised post-fusion structures of class-II fusion proteins such as the E protein trimers from DENV [68] and E1 from Semliki Forest virus [158] (Figure **6.3C&D**). The similarity between the SFTSV Gc structure and other post-fusion structures suggests that the Gc trimer is in the post-fusion conformation.

Like the related RVFV Gc and other class II fusion proteins, each protomer of the trimer is composed of three domains: I, II, and III (Figure **6.3B**). Domain I consists of an elongated thirteen-stranded β-sandwich at the central core of the structure, domain II consists of a five-stranded β-sandwich and a six-stranded β-sheet, and domain III consists of a seven-stranded β-barrel-like module and forms extensive protein-protein contacts (1,321 $Å^2$) with domain I. Overlay analysis between different protomers of SFTSV Gc reveals little deviation in structure between the symmetry-related molecules. The average root-mean-square-deviation (RMSD) was 0.64 Å over 428 Cα residues, with the greatest differences detected at regions responsible for forming crystallographic contacts.

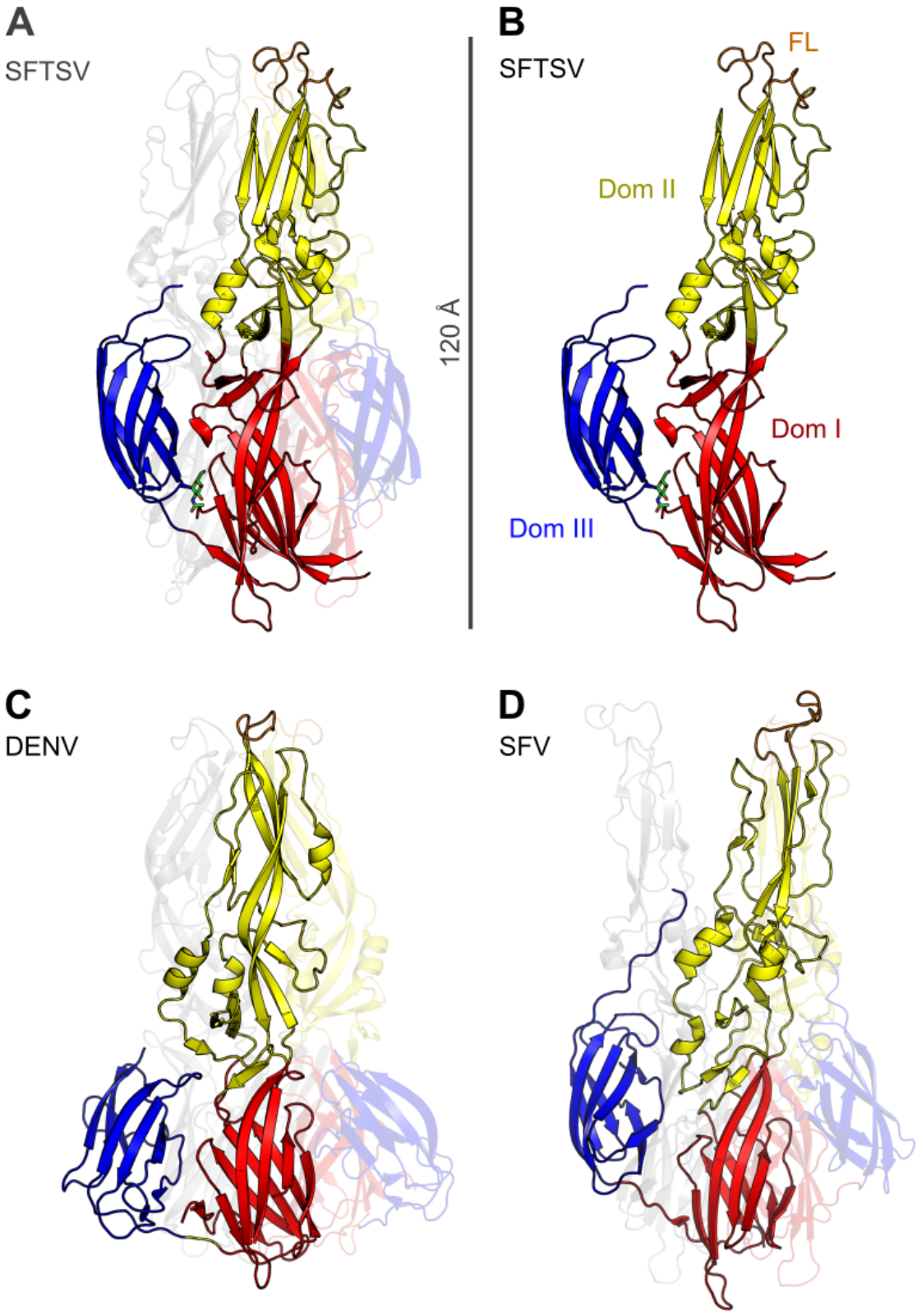


Figure 6.3 SFTSV Gc crystallised in the post-fusion trimeric state.
A) The SFTSV Gc trimer structure is a 120 Å long trimer. B) Each protomer of the trimer is composed of three domains: I, II and III. The hydrophobic fusion loops (FL) are located at the tip of domain II. C) and D) show the well-characterised DENV E protein (PDB code 1OK8) [68] and SFV E1 (PDB code 1RER) [158], respectively, in the post-fusion conformation.

6.5.2 The effect of pH on the oligomeric state of Gc

The appearance of a trimer in the crystal from a low-pH condition posed the question whether the protein forms trimers in solution under similar low pH conditions. To assess this, fully-glycosylated SFTSV Gc V996 was purified by SEC (Figure **6.4A**). The purified protein was split into two fractions. One fraction was buffer exchanged to a buffer containing pH 5.0 citric acid and the second fraction was kept in the original TRIS-based pH 8.0 purification buffer. Both fractions were subject to a further SEC analysis and compared to a protein size standard also run on the same SEC column (Figure **6.4B**). Both the pH 8.0 and pH 5.0 samples have a peak (ii) which corresponds to a monomeric SFTSV Gc (43 kDa three ~ 2 kDa N-linked glycosylation sites) when compared to the size standards. At pH 5.0, an additional peak (i) of a larger mass appears. Western blot analysis of the resulting fractions confirmed that both peaks (i) and (ii) are indeed SFTSV Gc, suggesting that peak (i) is an alternative oligomeric state. Since the crystal form is trimeric, we suggest that this form is also a trimer induced by low pH. These data indicate that a low pH environment alone can influences the trimerisation of SFTSV Gc.

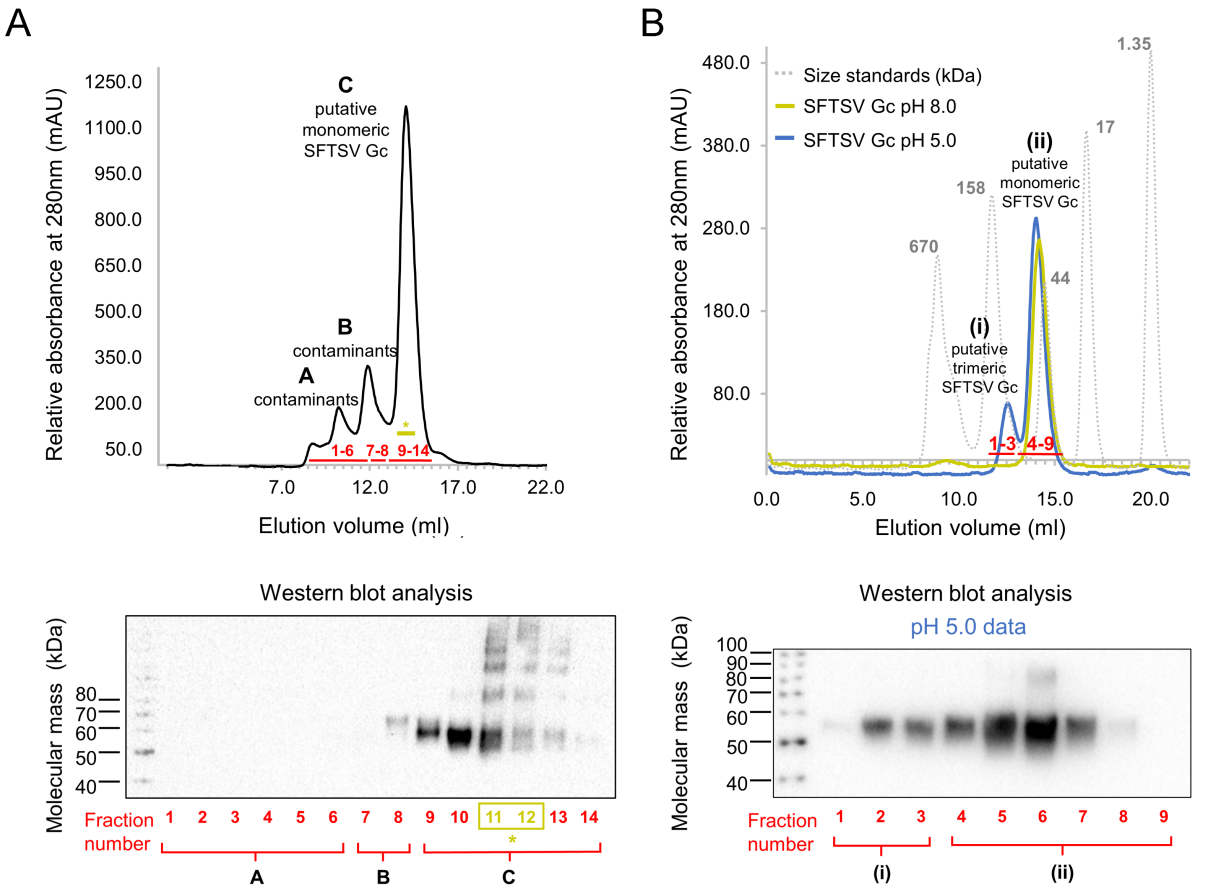


Figure 6.4 Effect of pH 5.0 on SFTSV Gc in solution.
SEC and Western blot analysis of purified SFTSV Gc V996. A) SEC of SFTSV Gc run at pH 8.0 following immobilized nickel affinity purification revealed one major peak (peak C) and two minor peaks (peaks A and B). Western blot analysis of fractions confirmed that peaks A and B contained contaminants from recombinant protein expression and peak C contained hexa-histidine tagged SFTSV Gc. Fractions coloured in olive green (*) were pooled and used for further SEC analysis in panel *B*. B) Peak fractions 11 and 12 from SEC performed in panel *A* were split in two and run at pH 8.0 and pH 5.0. A gel filtration standard (Bio-Rad, grey) was run on the same column for comparison. At pH 8.0 (green), a single peak (ii) corresponding to putative monomeric SFTSV Gc was observed. At pH 5.0 (blue) a second small peak (i), corresponding to a putative trimer, was observed in addition to peak (ii). Western blot analysis of pH 5.0 SEC fractions revealed that both the monomeric (ii) and putative trimeric (i) peaks were composed of SFTSV Gc, indicating that exposure to acidic environments may trigger SFTSV Gc trimerisation.

6.5.3 Structural re-arrangements between pre-fusion and post-fusion states is homologous to changes seen in flavi- and alphaviruses.

The RVFV Gc and SFTSV Gc have a sequence identity of ~25% and therefore the crystal structure of RVFV Gc in the pre-fusion conformation constitutes the closest known structural

relative to SFTSV Gc [47]. An overlay of individual domains from the two proteins reveals that the primarily β-stranded domain III is very similar with an RMSD of 1.16 Å over 81 C α residues (Figure **6.5A**). The other two domain show larger structural differences, with an RMSD of 2.13 Å over 117 Cα residues for domain II and an RMSD of 3.41-Å over 109 Cα residues for domain I (Figure **6.5B** and **C**). Some of the difference in domain II may be accounted for by differences in loops and short helical elements decorating the central β-sheet. However, the differences in domain I are due to rearrangements of secondary structure elements focused at the C-terminus, which connects domain I to domain III (Figure **6.6**). In the pre-fusion form, the C-terminal strand '13' hydrogen bonds to strand '2'. In the post-fusion form, strand '13' displaces strand '0' and strand '13' forms hydrogen bonds with strand '3'. The displaced strand '0' becomes continuous with strand '1' and the loop between strands '1' and '2' becomes disordered. These rearrangements are propagated through to domain III via the C-terminal strand '13', resulting in a 24-Å shift in the position of domain III between pre- and post-fusion conformations. Analogous conformational changes have been described for DENV E protein, further consolidating the conservation of the class II fusion protein fold [68].

Analysis of protein-protein contacts between the pre- and the post-fusion states reveals that domain III makes more than two times greater protein contacts in the post-fusion state (1,321 $Å^2$) with other parts of the trimer than domain III makes in the pre-fusion state (570 $Å^2$). This is analogous to the changes in protein-protein contacts observed in DENV E (from 826 to 1,315 $Å^2$) [68, 91] and SFV E1 (from 572 to 1,530 $Å^2$) [158, 159]. The formation of such extensive contacts likely stabilizes the post-fusion conformation, as has been suggested for SFV E1 [158].

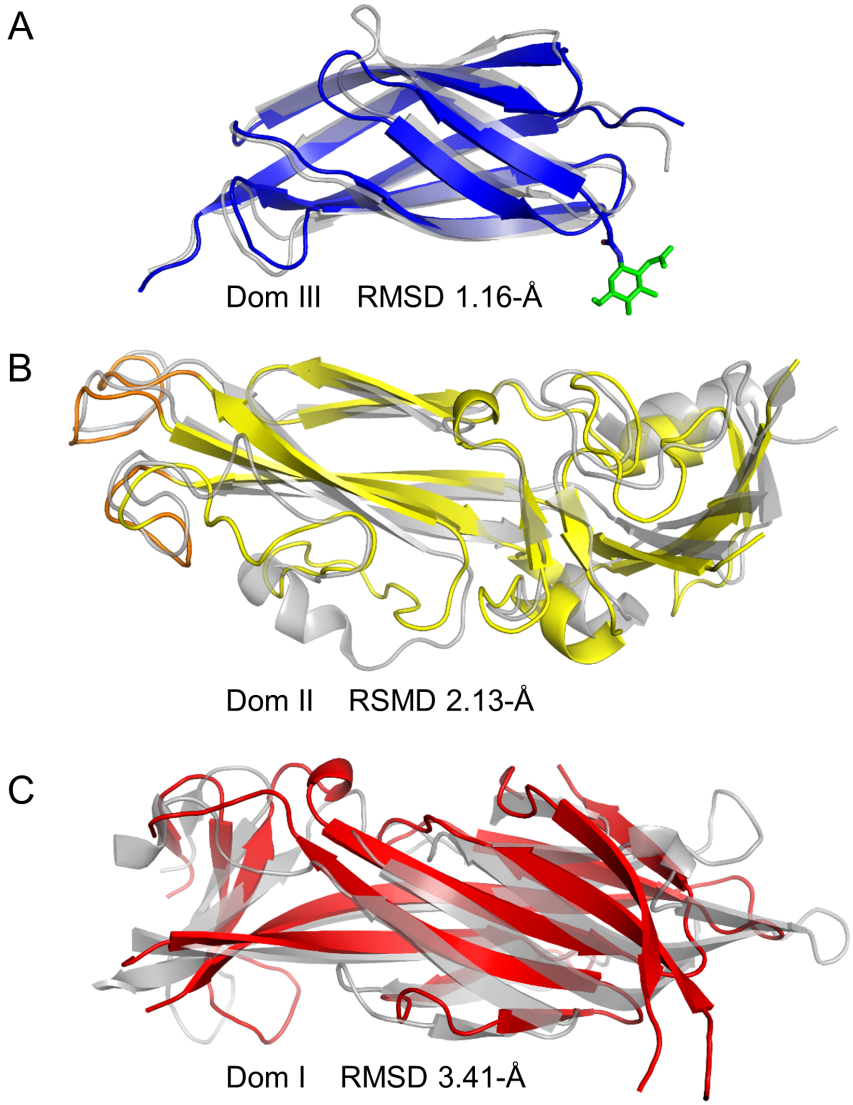


Figure 6.5 Overlay analysis of individual domains of SFTSV Gc and RVFV Gc.
Structures are shown in cartoon representation where RVFV Gc is transparent and coloured gray and SFTSV Gc is colored according to domains as in **Figure 6.3**. A root-mean square deviation (RMSD) was calculated between each pair of domains and shown below each domain. (A) Overlay of domain III (1.16-Å RMSD over 81 Cα residues), (B) domain II (2.13-Å RMSD of over 117 Cα residues), and (C) domain I (3.41-Å RMSD over 109 Cα residues).

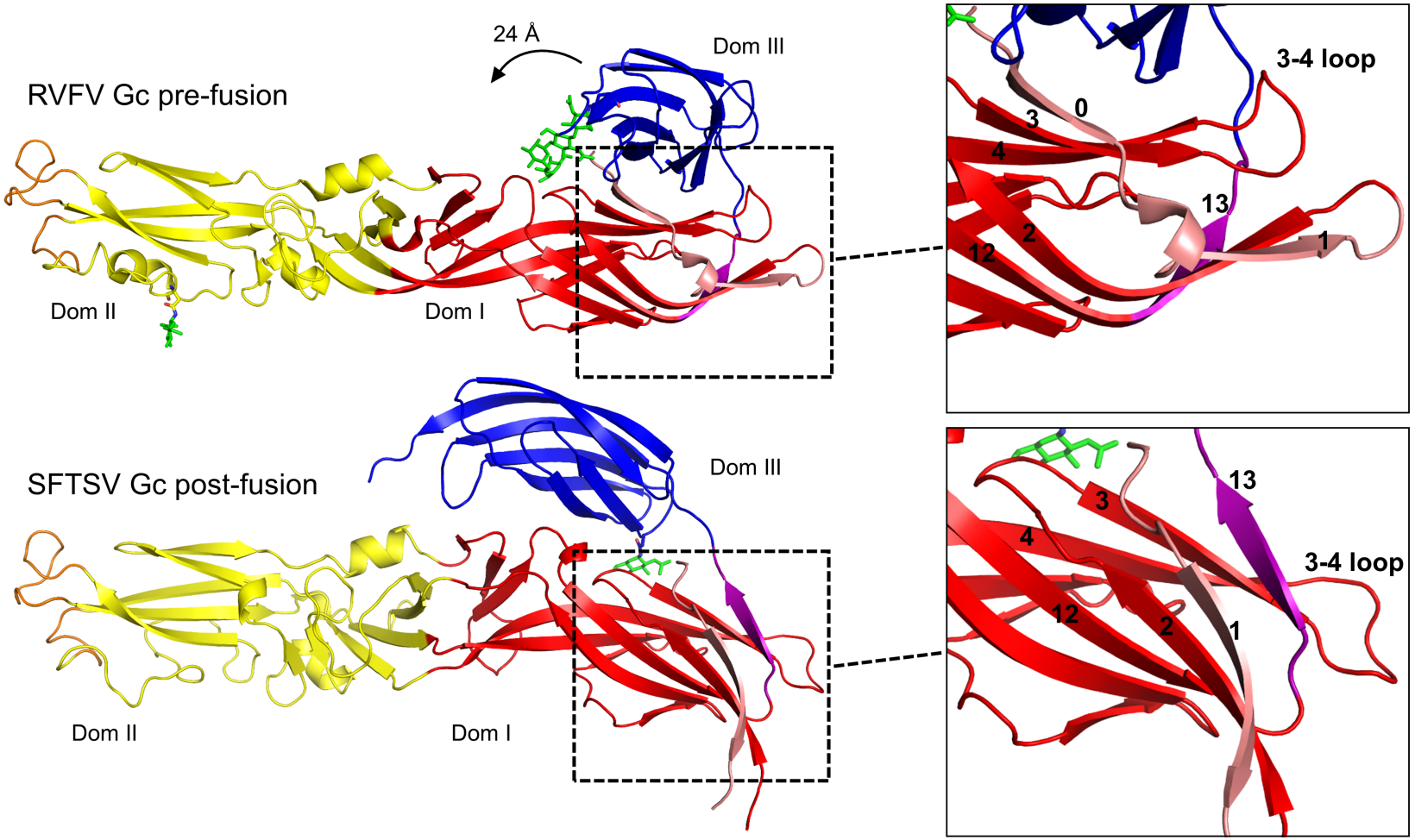


Figure 6.6 Structural rearrangements of the phleboviral Gc from prefusion to postfusion conformations.
A single protomer of RVFV Gc [47] and SFTSV Gc are shown in cartoon representation and domains are coloured as in **Figure 6.3**. Glycans are shown as green sticks. Zoom-in panels of domain I are shown on the right side and highlight the strand swap occurring between pre- and post-fusion states. In the postfusion conformation, strand no. 13 (purple) reorientates around the 3-4 loop, forming a β-sheet with strands no. 3 and 4, and strand no. 0 (pink) becomes continuous with strand no. 1 (pink). This results in a 24 Å shift in domain III.

6.5.4 The phleboviral fusion loops are conformationally rigid

Essential features of the class-II fusion glycoprotein architecture are the hydrophobic fusion loops located at the apex of domain II, which are present in all known structures of viral class II fusion proteins. In flaviviruses and alphaviruses, this is a single loop of aromatic residues, which is inserted into the host membrane and draws the viral and host membranes together upon fusogenic rearrangements of the molecule [68]. In contrast to a single fusion loop, which performs this function in alpha- and flaviviruses, the divergent Rubella virus E1 has two putative fusion loops [71]. Similarly, bunyaviral Gc glycoproteins contain two (RVFV, SFTSV, HNTV) or even three (PUUV) putative fusion loops. In SFTSV, these loops (Cys650–Cys656, loop 1; Cys691–Cys705, loop 2) contain the hydrophobic amino acids Ala694, Ala695, Ala701, Trp652, and Phe699 and are fully solvent exposed in the crystal structure. The hydrophobic residues of the fusion loops are not well conserved between SFTSV and RVFV, where Phe699 from SFTSV and Phe826 from RVFV are the only homologous residues. Additionally, the fusion loops of RVFV Gc are concealed within oligomeric protein–protein contacts of the crystallographically observed head-to-tail homodimer. Despite these differences, the loops from both viruses form a strikingly similar conformation. Superposition reveals little difference in structure with an RMSD of 0.75 Å over 22 Cα residues (Figure **6.7A**). The conserved conformation of the two loops is consistent with that observed in the DENV E protein fusion loop [68] (Figure **6.7B**), but contrasts the conformational differences observed in the fusion loop of SFV E1 protein [158] and the HNTV Gc [95] (Figure **6.7C**&**D**). The conformational differences observed in SFV E1 between the pre- and post-fusion states may be influenced by crystallographic packing of the post-fusion state, where the fusion loops of several trimers pack together to sequester the hydrophobic fusion loops from solvent.

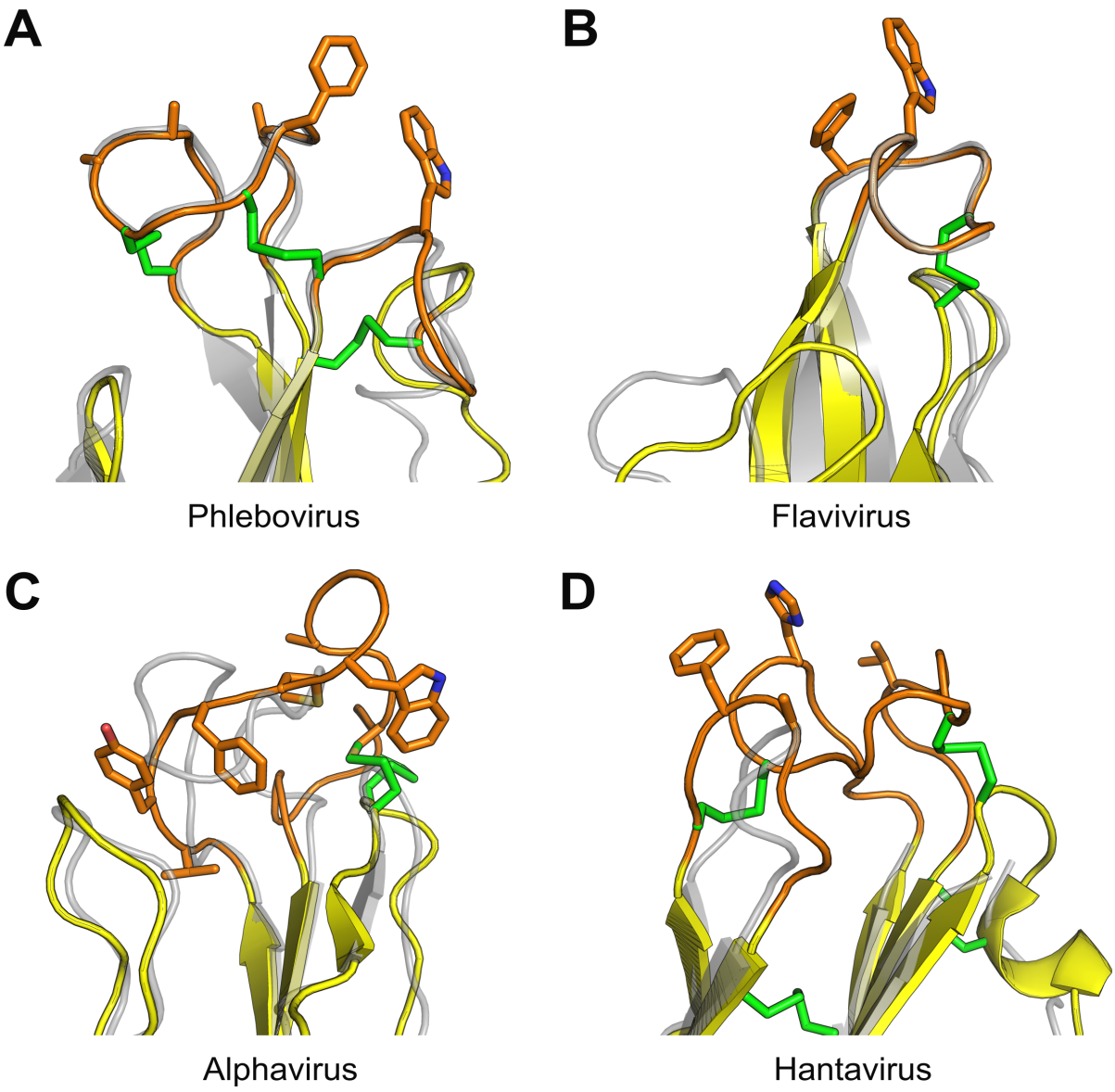


Figure 6.7 Conformations of class II fusion loops.
Overlay analysis between the pre- and post-fusion states of several class II fusion proteins. The post-fusion state in each figure is presented in colour and cartoon, with hydrophobic residues and cysteines shown in stick representation. The fusion loops are coloured orange and cysteine bonds in green. The pre-fusion state is presented as a transparent gray cartoon. A) Comparison between the pre-fusion RVFV (PDB code 4HJ1) [47] and post-fusion SFTSV (PDB code 5G47) shows that the fusion loop backbone is conformationally rigid. B) The DENV-2 fusion loops are also rigid between the pre- [91] and post-fusion [68] conformations (PDB codes 1OAN and 1OK8). C) Extensive changes are seen in the large SFV fusion loop where the backbone seems to refold substantially [158, 159]. These changes may be influenced by crystallographic packaging as several SFV trimers pack together at the site of the fusion loop (PDB codes 1RER and 2ALA). D) The fusion loops of the pre-fusion HNTV are largely disordered in the crystal structure, which may indicates mobility and a lack of rigidity (PDB codes 5LJZ and 5LJY) [95].

6.6 Site-directed mutagenesis and reverse genetics reveal functionally important sites on the SFTSV Gc.

A series of mutations were designed based on the crystal structure of SFTSV Gc to shed light into the function of specific sites on the glycoprotein surface. These mutations were first introduced into a recombinant expression system, to assess the effect of these mutations upon protein folding and secretion (Figure **6.8**) and then in a live-virus context using a previously reported reverse genetics system [160]. All reverse genetics experiments were conducted in biocontainment level 3 by Dr Benjamin Brennan in the group of the late Prof Richard M. Elliott at the MRC-University of Glasgow Centre for Virus Research.

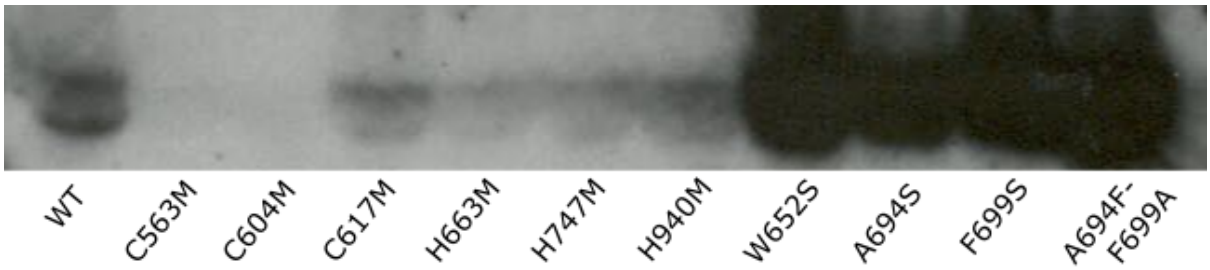


Figure 6.8 Western blot analysis of secreted SFTSV Gc mutants.
Constructs of soluble SFTSV Gc containing various mutations were tested for expression and secretion in a mammalian expression system (see Methods **2.2**). The crystallised construct, SFTSV Gc V996, was used as the wild-type (WT) control. Two mutants, C563M and C604M, were not secreted and folded. All other mutants were expressed at levels equivalent to or greater levels than WT.

6.6.1 The SFTSV Gc fusion loop is highly sensitive to mutations

SFTSV Gc encodes two hydrophobic loops at the tip of domain II, which are most likely important for the fusion activity of the Gc. The hydrophobic landscape of these putative fusion loops is dominated by two bulky residues, Trp652 and Phe699, which extend out towards the solvent (Figure **6.9A**). Phe699 is conserved among phleboviruses while Trp652 is not as conserved, but the position in the protein is occupied by a hydrophobic amino acid

in most other phleboviruses (Figure **6.9B**). To assess the importance of these residues, rescue of live SFTSV with single mutations W621S, F699S and A694S was performed. Additionally, to assess the effect of exchanging residues, a double mutant A694F/F699A was also tested. In comparison to the wild-type virus, none of these mutants could be rescued (Figure **6.9C**) despite proper folding and secretion of the soluble versions of the mutant proteins (Figure **6.8**). These results underscore the sensitivity of this loop region to sequence variation, whereby only limited changes in sequence can preserve functionality.

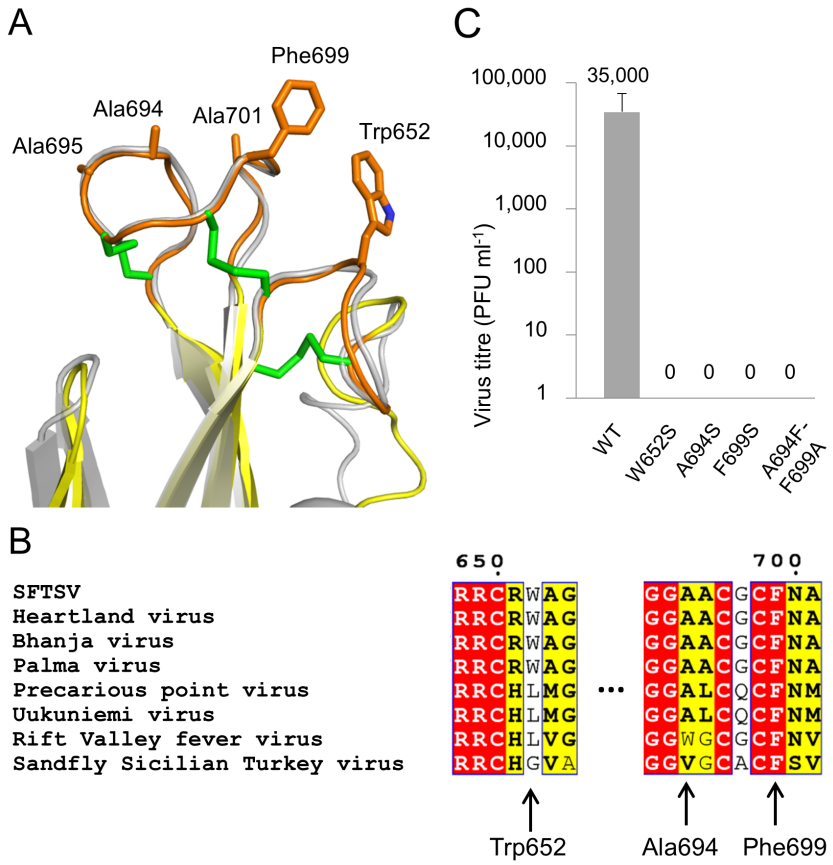


Figure 6.9 The putative fusion loops of SFTSV Gc contain essential residues.
A) An overlay of the fusion loops of SFTSV Gc and RVFV Gc (as in Figure **6.7A**) highlighting hydrophobic amino acids. B) Sequence alignment of fusion loops across selected phleboviruses (Appendix **10.9**). Residues highlighted red are fully conserved and yellow are partially conserved. Residues tested by site-directed mutagenesis are highlighted by arrows. Phe699 is fully conserved while Trp652 is more varied amongst phleboviral sequences. C) Reverse genetics analysis of SFTSV encoding single and double site-directed mutations at the putative fusion loops. Recombinant SFTSV were titred in plaque forming units (PFU) and compared with that of wild type (WT) SFTSV.

6.6.2 pH-sensing histidines on the virion surface.

Functionally active and surface-exposed histidine residues are a canonical feature of the class-II fusion machinery [161, 162]. As phleboviral virions are trafficked through the endocytic pathway, the pH gradually drops to ~5.0 in the late-endosomes. Histidine side chains (with a pK_a ~6.0) are thought to trigger conformational changes in the Gc fusion protein, resulting in viral and cell membrane merger [45, 163]. In alpha- and flaviviruses, many such functionally important histidines are localised at the interface between domains I and III [164, 165], where they seem to stabilise the energetically favourable post-fusion conformation by forming hydrogen bonds and hydrophobic interactions between the domains.

Similar histidine residues have been found on the RVFV Gc, where His778, His857, and His1087 are required for infectivity [163]. However, in contrast to the localized patch of histidines in alpha- and flaviviruses, His778, His857, and His1087 are spread throughout domains I, II, and III of RVFV Gc, respectively. His1087 is localised near the interface of domains I and III and may be important for stabilising the post-fusion state. His857 has no proposed functional role and cannot be fully explained in the context of the RVFV Gc structure, but His778 has been suggested to stabilise the fusion peptide during host membrane interaction. Although none of these residues are conserved in SFTSV, His663, His747, and His940, were identified at nearby sites in the SFTSV Gc structure (Figure **6.10A**). These residues were probed in the reverse genetics system to assess their importance to the virus life cycle. Three single mutants were designed, H663M, H747M and H940M, none of which effected protein folding (Figure **6.8**). H663M was rescued at wild-type levels while H474 was rescued to low titres and H940M was not rescued (Figure **6.10B**).

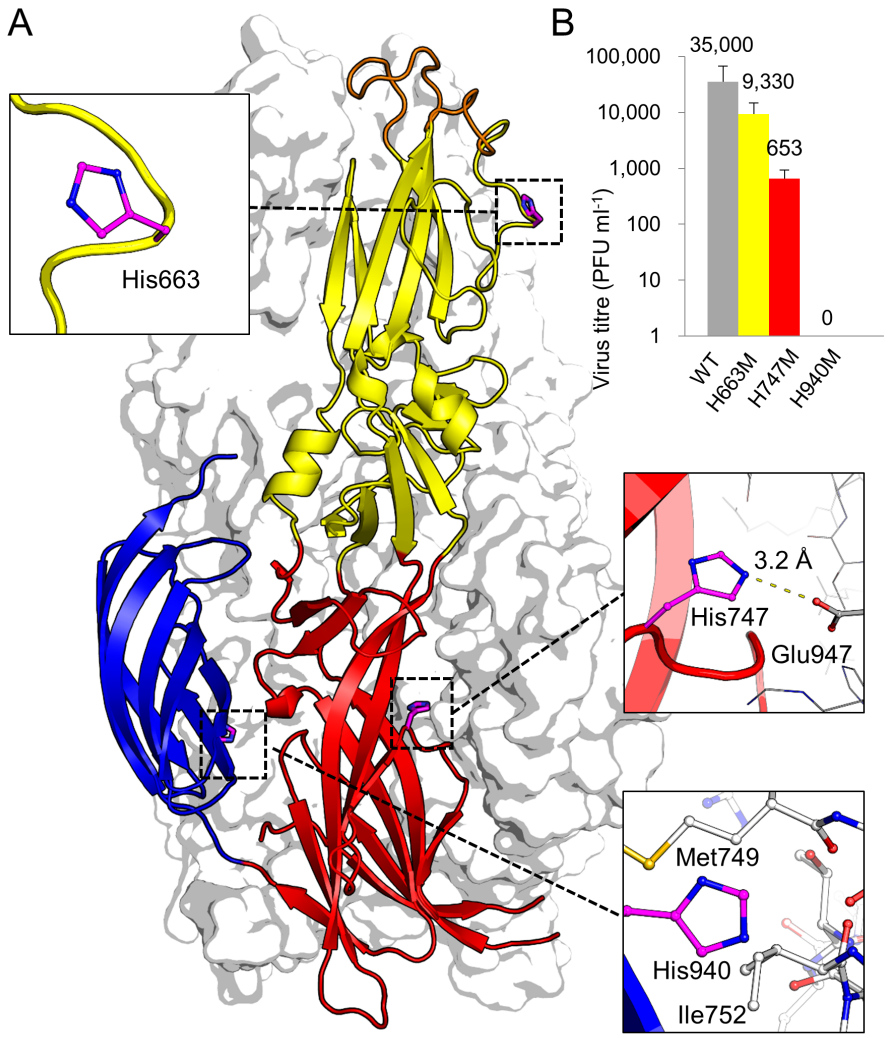


Figure 6.10 Surface exposed histidines play an important role in the virus life cycle. A) A single SFTSV Gc protomer is shown in cartoon representation with the remainder of the trimer shown as a white van der Waals surface. Surface exposed histidines are shown as purple sticks. Zoom-in panels highlight the location of these residues within each of the three domains and residues surrounding His747 and His940 from adjacent protomers are shown as white sticks B) SFTSV encoding single mutations of His663, His747 and His940 were derived by reverse genetics by Dr Benjamin Brennan and the titers of these recombinant viruses, measured in PFU, were compared with that of WT SFTSV.

Although these experiments do not specifically probe the mechanism of fusion, these results suggest that the SFTSV life cycle relies on the functionality of Gc-resident histidines. The importance of histidine residues has also been observed in alpha- and flaviviruses [164, 166]. His747 is involved in a hydrogen bond at the interface between two protomers in the post-fusion state which may be important in stabilising that state, much

like what is seen in DENV type 1 E protein where several histidines come together to form a hydrogen bonding network [165]. His940 is also present on the interface between two protomers but it does not seem to be involved in hydrogen bonding, but rather in hydrophobic interactions and perhaps some weak electrostatic forces (**Figure 6.11**) which may also be important for stabilising the post-fusion state. Mutating His663 did not have an effect on the virus life cycle. Despite the lack of histidine conservation among the phleboviruses, these mutational analyses reveal that surface-exposed histidines play a key role in the phleboviral life cycle.

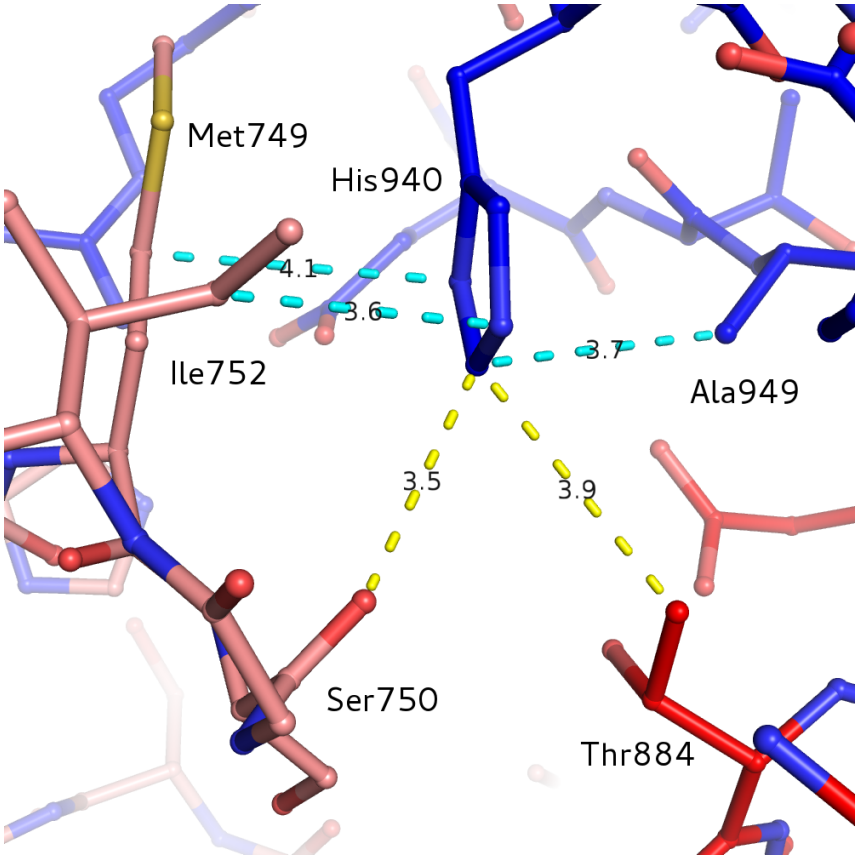


Figure 6.11 The His940 chemical environment.
His940 and residues surrounding it are shown in stick figures. Chains are coloured as in **Figure 6.3**, with the expection that a neighbouring protomer domain I is coloured in salmon. Non-carbon atoms are coloured as follows: oxygens are red, nitrogens are blue and sulfurs are yellow. At a pH of 4.5, the imidazole ring of His940 is expected to be protonated and as such expected to be involved in some hydrogen bonding or electrostatic interactions. Two candidates (Ser750 and Thr884, distance coloured in yellow) are present but the distance from the imidazole nitrogen is too great for a strong hydrogen bond. However, weak electrostatic forces may be involved. Three other residues have atoms within a similar distance from the imidazonle ring (between 3.5 and 4.1 Å, distances coloured in cyan) which are all hydrophobic. The His940 environment in the post-fusion structure is thus a mixture of weak electrostatic forces, which may stabilize the extra proton, and hydrophobic interactions.

6.6.3 A curious free-cysteine is not required for infection.

During the structural analysis of SFTSV Gc, a free cysteine, Cys617, was observed on the surface of the glycoprotein nearby a disulphide bond between Cys563 and Cys604 (Figure **6.12A**). Interestingly, despite being in close proximity to Cys563 and Cys604, the thiol side chain of Cys617 appears to be fully reduced and does not appear to interact with these

residues. To assess the functional importance of this lone cysteine, if any, three single mutations (C563M, C604M and C617M) were designed and tested in the reverse genetics system (Figure **6.12B**). Unsurprisingly, mutations of the residues contributing to the disulphide bond resulted in no rescued viruses. Similarly, recombinant protein expression of soluble mutants of the disulphide bond showed no expression (Figure **6.8**). The C617M mutant on the other hand was rescued to wild-type levels and was properly folded and secreted. Sequence alignment of representative phleboviral sequences shows that Cys563 and Cys604 are conserved among all the sequences, while the Cys617 is only found in the related Heartland virus. It is thus likely that this lone cysteine lacks a distinct functional role in the virus lifecycle.

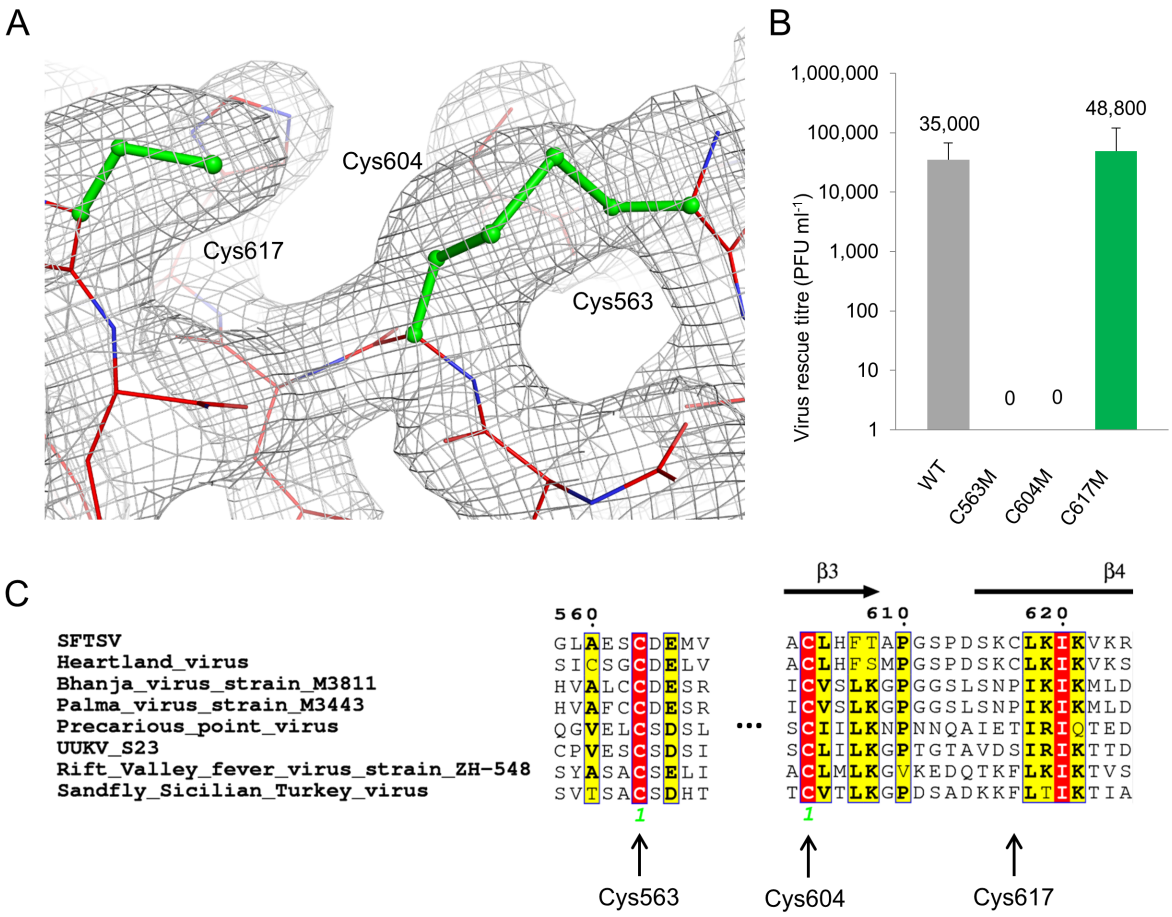


Figure 6.12 The free cysteine, Cys617, is not required for the SFTSV life cycle.
A) Electron density (2*Fo*-*Fc* contoured at 1.0 σ, grey) around cysteines Cys563, Cys604 and Cys617 in SFTSV Gc. B) SFTSV encoding single cysteine mutations (C563M,

C604M, and C617M) were derived by reverse genetics and the titers of these recombinant viruses, measured in plaque forming units (PFU), were compared with that of wild type (WT) SFTSV. Only the C617M mutant was rescued and titrated to levels comparable with the wild-type (WT) virus. C563M and C604M failed to rescue. C) Sequence alignment of selected phleboviruses (Appendix **10.9**) reveals that Cys563 and Cys604 are conserved amongst all selected phleboviruses but Cys617 is only found in the closely related Heartland virus.

6.7 N-linked glycosylation on the SFTSV Gc

Glycosylation are predicted to be presented at the three N-linked glycosylation sequons: Asn853, Asn914, and Asn936. The expressed SFTSV Gc V996 construct was deglycosylated to single N-acetylglucosamine (GlcNAc) moieties, prior to crystallization. In the crystal, only one of these moieties (Asn936-GlcNAc) exhibited strong enough electron density to be fully built and none was visible at Asn853 or Asn914 (Figure **6.13**). The absence of any density for glycans at Asn853 and Asn914 may be due to inherent flexibility of the glycan in the crystal, but most likely these sites may only be partially glycosylated or completely devoid of glycosylation during recombinant expression.

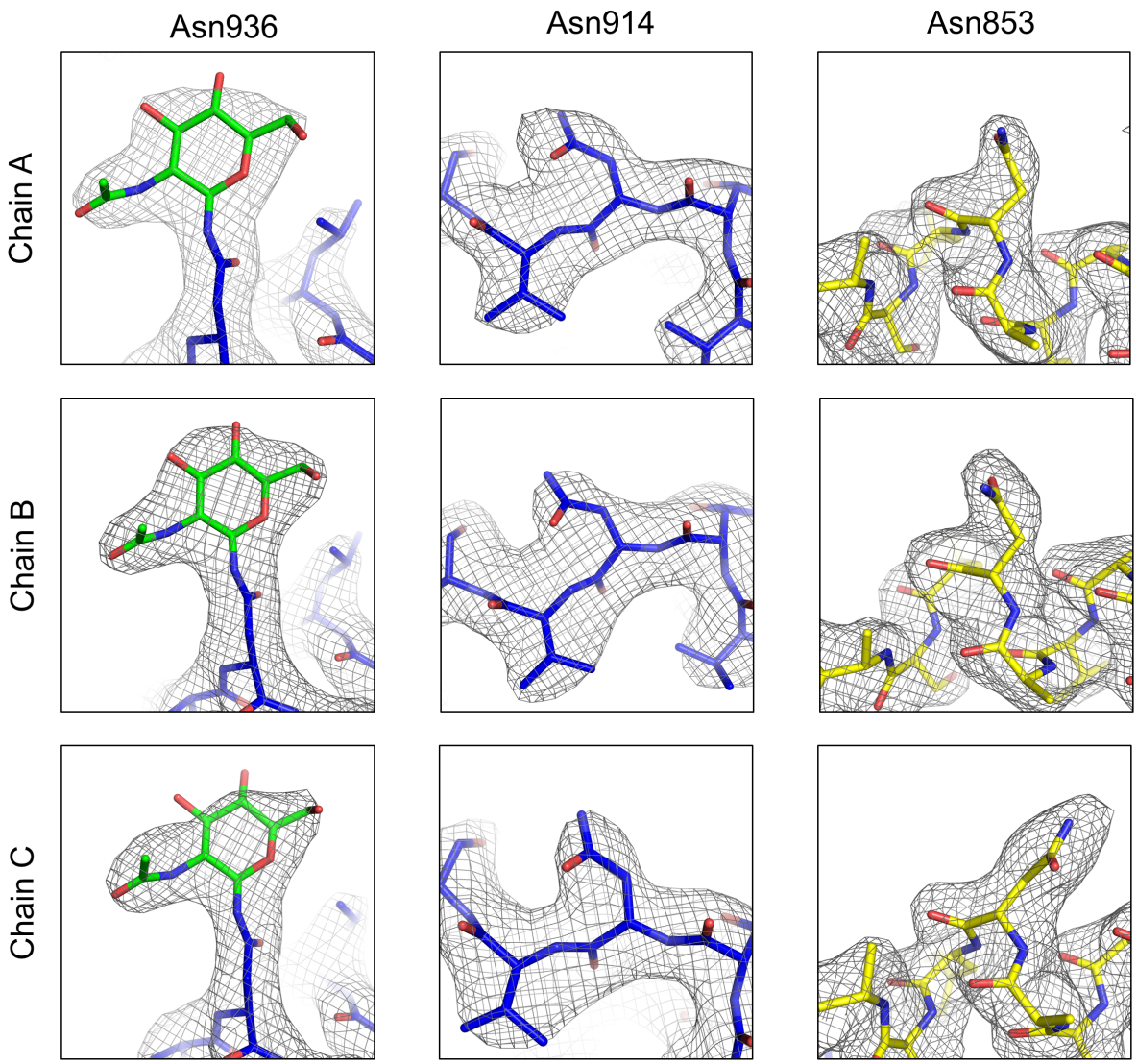


Figure 6.13 Electron density of SFTSV Gc glycans.
Three N-linked glycosylation sites are predicted for SFTSV Gc: Asn936, Asn914 and Asn853. The gray mesh is a *2Fo-Fc* map at 1.0 σ while the structure is shown in stick. Only the glycan of Asn936 could be built with confidence.

SFTSV cell entry is mediated by DC-SIGN [36], a cell surface C-type lectin with a high affinity (nM) for oligomannose-type glycans [167] and it is likely that glycosylation of Gn or Gc of SFTSV are important for lectin-mediated host cell entry. To assess whether the Gc glycans are processed to oligomannose-type glycans in our recombinant expression system, HILIC-UPLC (hydrophilic interaction chromatography-ultra performance liquid chromatography) was performed on SFTSV Gc V996 (Figure **6.14**) in collaboration, by Dr.

Anna-Janina Behrens in the group of Prof. Max Crispin at the Oxford Glycobiology Institute. Unlike the glycoprotein preparations for crystallography, no mannosidase inhibitors were included during expression and endoglycosidase F1 (also known as endoglycosidase H) was used to cleave the glycans for analysis.

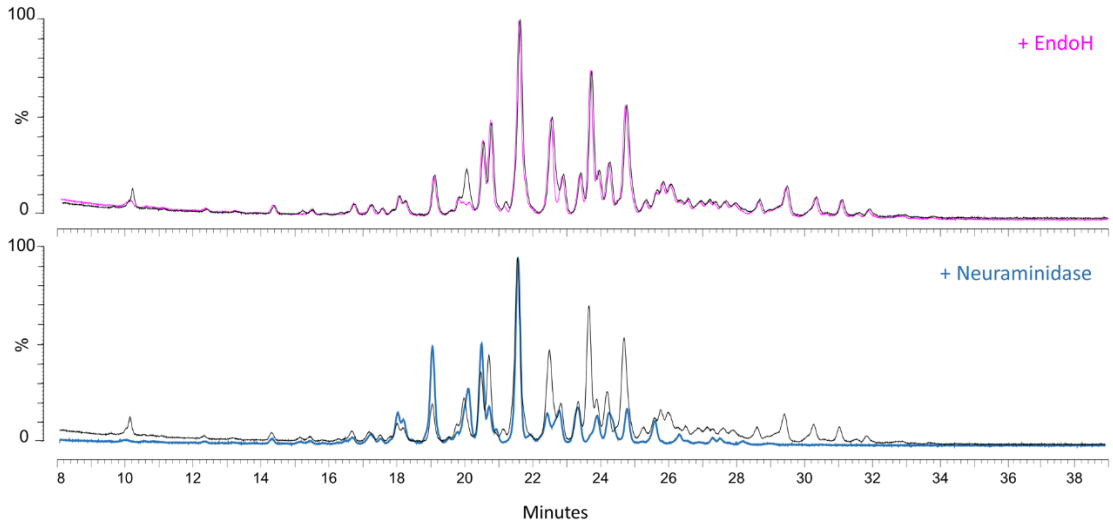


Figure 6.14 Glycan content analysis of SFTSV Gc by chromatography.
Recombinantly expressed SFTSV Gc, produced without the glycosylation inhibitor, kifunensine [105], was incubated with endoglycosidase H (endoglycosidase F1) to release glycans. Glycans were fluorescently labelled and subject to HILIC-UPLC (hydrophilic interaction chromatography-ultra performance liquid chromatography) analysis. The upper panel displays the overall glycan profile before (black) and after (magenta) endoglycosidase H digestion. The spectrum is dominated by highly processed complex-type glycans while only negligible levels of oligomannose-type glycans were present. The lower panel displays the glycan profile before (black) and after (blue) disialylation by neuraminidase, showing the abundance of sialic acid-containing glycans present on SFTSV Gc.

Unlike the glycan composition known to exist on the virion surface of the related UUKV [41], negligible levels of hybrid- and oligomannose-type populations were observed on recombinant SFTSV Gc and the spectrum was dominated by highly processed complex-type glycans. These data are suggestive that glycans on isolated Gc are processed differently than in the context of the whole virus. The HIV-1 trimeric envelope glycoprotein

has been shown to exhibit site-specific differences in glycan processing influenced by the local steoreochemistry [168]. In a related manner, it may be that steric limitation on the surface of SFTSV leads to the presence of oligomannose-type glycans. However, it may be that the glycans on the Gn may also contribute to entry equally as the glycans of the Gc, as suggested for RVFV [42]. Across the *Phlebovirus* genus, the number and position of glycosylation sites is not well conserved (Figure **6.15**). Although entry via DC-SIGN is likely a common feature of phleboviruses [37, 100], it may be that the glycosylation sites utilized for receptor binding may not be conserved across the genus.

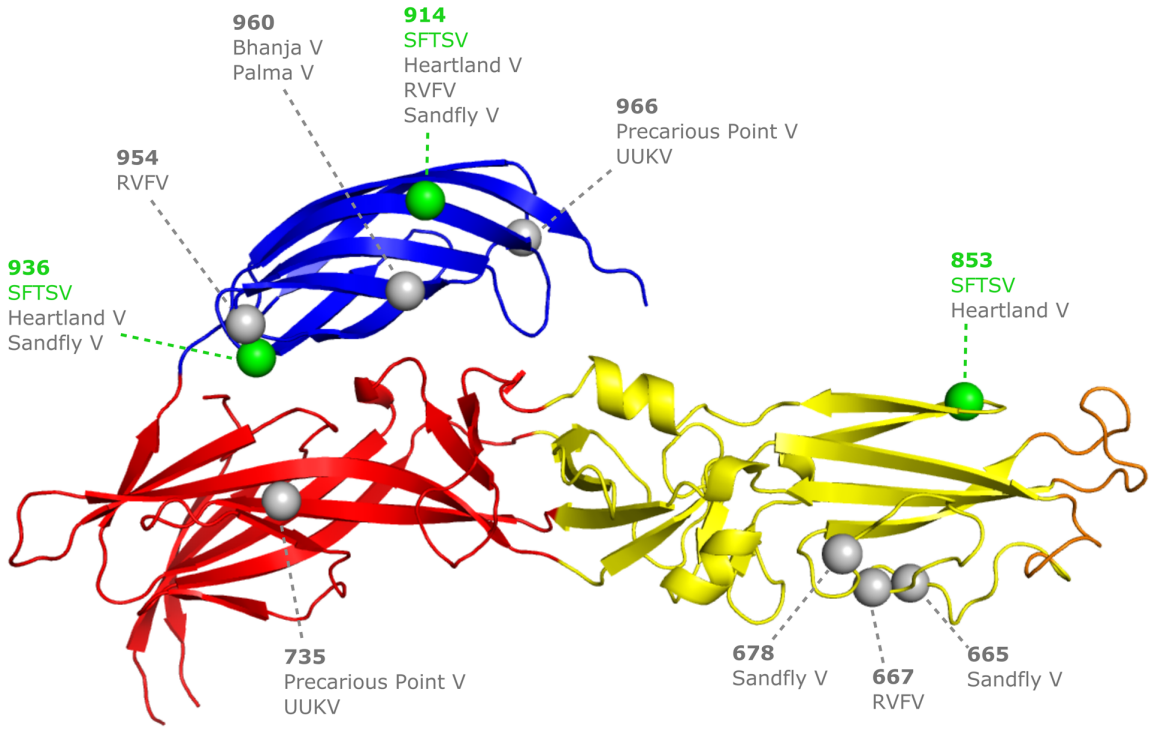


Figure 6.15 Diverse positions of N-linked glycans on the phleboviral Gc.
Predicted N-linked glycosylation sites of representative phleboviral sequences (Appendix **10.9**) were mapped onto the structure of SFTSV Gc with numbers based on the M-segment of SFTSV. Green spheres show sites on SFTSV, while sites in grey are found on other phleboviruses. Glycan sites appear to be clustered on domain II and domain III while the precise location of the sites varies greatly.

6.8 Concluding remarks: Evolution of the class II fusion fold

The work presented herein shows that the phleboviral Gc is structurally and functionally homologous to class II fusion proteins from alpha- and flaviviruses. Upon acidification, SFTSV Gc shifts from a monomeric to a putative trimeric state in solution, indicating that the trimer observed by crystallography likely constitutes a physiologically relevant post-fusion state. Interestingly, the structural transitions inferred by comparison of this post-fusion structure with the neutral-pH structure of RVFV Gc are homologous to those observed from flavivirus structures. Furthermore, reverse genetics and mutagenesis analysis showed that hydrophobic residues on the putative fusion loops and several histidines are integral for the virus lifecycle. Both features are hallmarks of viral class II fusion proteins, which further consolidates a model of a conserved mechanism of fusion between distantly related phleb-, alpha- and flaviviruses.

Another feature that relates phlebo-, alpha- and flaviviruses is their ability to use DC-SIGN as a receptor for host cell entry [40, 169]. Three investigations have studied the glycan content of phleboviruses in the context of the whole virus [41, 42, 170], revealing that N-linked glycans from the Gc of UUKV and RVFV both contain oligo-mannose type glycans, regardless of which cell line the virus originates from. The glycan composition of the Gn however, varies between the two viruses and the cell types. Interestingly, the Gn from RVFV produced in mammalian cells and Gn from UUKV produced in tick cells contains oligo-mannose type glycans. Gn from UUKV produced in mammian cells however, predominantly displays complex-type glycans, which are incompatible with DC-SIGN-mediated entry. It is thus evident that the specificity that underlies DC-SIGN recognition is likely to vary between different phleboviruses and the cell types used to

produce the virus. Direct observation of DC-SIGN binding to the virion might help us address these questions, until then we can only speculate.

Structures of the Gc glycoprotein from two hantaviruses were recently reported, which revealed that the hantavirus Gc is also a class II fusion protein [73, 95]. Phlebo- and hantaviruses used to be classified under the diverse *Bunyaviridae* family, but recently the International Committee of Virus Taxonomy (ICTV) has re-classified the family into the order, *Bunyavirales*. The creation of the broader classification of *Bunyavirales* reflects the broad diversity of bunyaviruses. The wealth of class-II fusion structures now available creates an interesting opportunity to investigate the relationship between the viruses harbouring these glycoproteins using structure-based phylogenetic methods. Accompanying the report of one of the hantaviral Gc glycoprotein structure [73] was an attempt at defining relatedness of class II fusion proteins using the structural classification tool MUSTANG [118]. This analysis included class II fusion proteins identified from viruses as well as eukaryotic organisms. The structural analysis here attempted to perform a similar investigation using several currently available structural phylogeny tools: MUSTANG [118], SHP [116], and the Dali server [120]. Only structures of proteins in the post-fusion sate were included to avoid conformational-induced errors, which Willensky et al. did not take into consideration. While the resulting cladograms may reveal some interesting structural properties of the fusion glycoprotein fold, the results from these several analyses were varied and the inferred relationships between the fusion proteins was inconsistent (Figure **6.16**). Such varied results all point to the limitations of structure based analysis. Future analysis of proteins which exhibit very low sequence similarity but have same overall fold, such as class II fusion proteins, may need tools to account for flexibility in structure and sequence divergence to overcome some of these limitations.

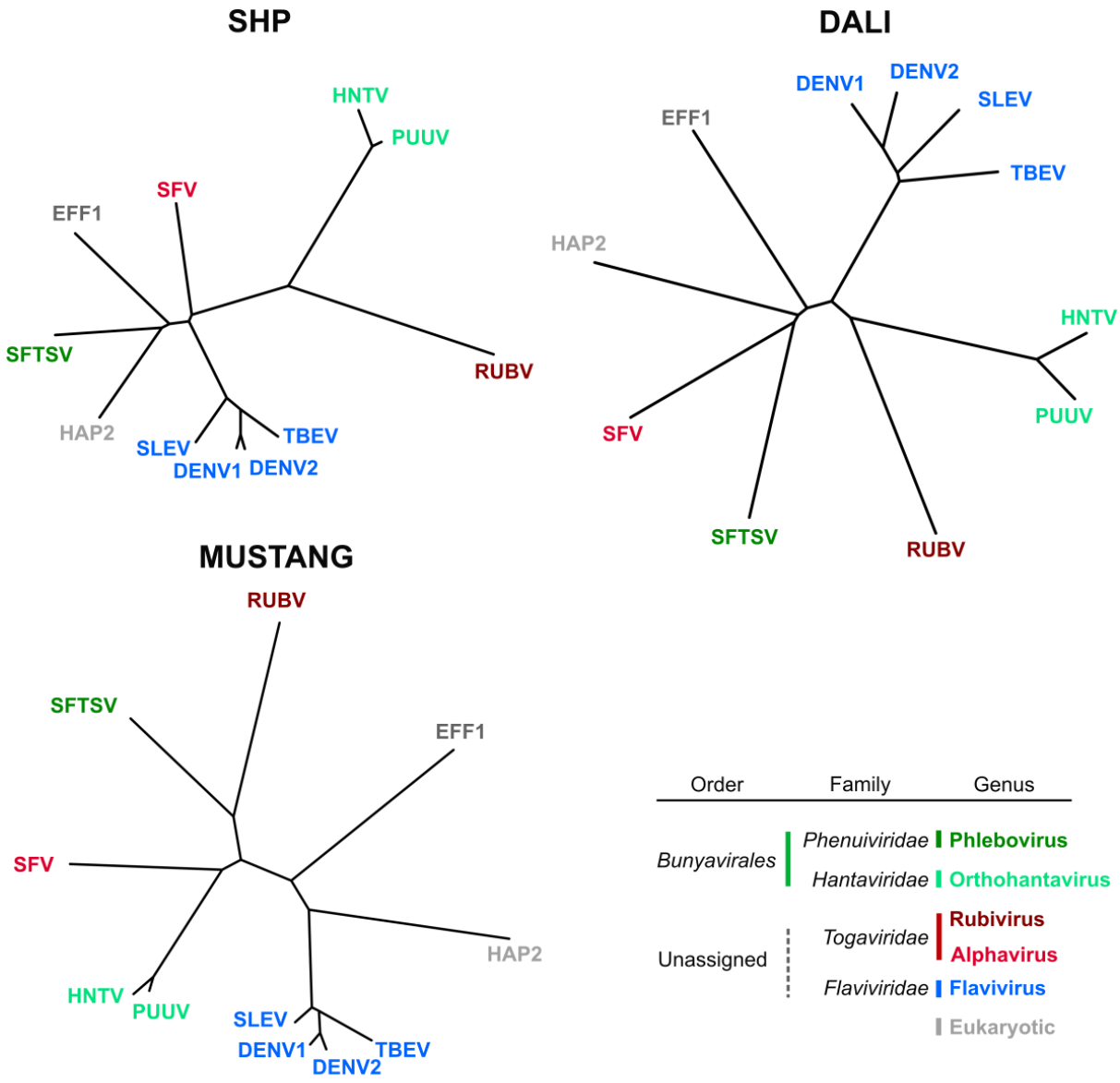


Figure 6.16 Structural phylogeny of class II fusion proteins.
Eleven post-fusion structures of class II fusion proteins were used to generate structure-based cladograms (see Methods **2.3.5**). Three different approaches were compared which gave varied results. SHP [116] is based on cross-correlation of atomic coordinates and optimises structural overlap to infer relatedness. The DALI server [120] searches for secondary structure similarities and uses atomic coordinates to generate a Z-score which was used to infer relatedness. MUSTANG [118] uses chemical properties of amino acids and atomic coordinates to generate a structure-based sequence alignment which is then used to infer relatedness. Virus names are coloured according to ICTV taxonomy [29]. Only post-fusion structures were included to minimise artefacts introduced by differences in functional states.

The class II fusion protein fold is widely spread and we may yet discover many more instances of this fold in unexpected places. Unless we figure out how to turn back the

clock and observe the evolutionary processes in action, or find another way to measure virus relatedness, we may never know the true origin of this curious protein fold.

7. Discussions and future directions

7.1 Foreword

This final chapter brings together the work presented in previous chapters. As previous chapters had each their own concluding section, I will focus on the impact of the findings of the thesis in a broader context. Some preliminary results of receptor binding and host immune response will be presented as a prelude to what future research in the field of phleboviruses may bring.

7.2 Entry and assembly

The work presented in this thesis has broadened our understanding of the entry mechanism and assembly of phleboviruses. The localised reconstructions of the RVFV capsomers constitute the first sub-nm cryo-EM structures of any virus from the *Bunyavirales* order. Furthermore, the fitting of the Gn and Gc crystal structures into those densities and the resulting model is the first pseudo-atomic account of any bunyavirus glycoprotein assembly. The pseudo-atomic model is a treasure trove of information and, as any new structure, it establishes a new platform for future research. And specifically for structural studies, avenues such as fusion intermediates, receptor-virus complexes, humoral antibody response and mechanism of neutralisation can be explored in the broader context of the virion assembly.

The arrangements of the glycoproteins on the RVFV surface has many implications for the viral entry mechanism. Firstly, the arrangement of N-linked glycosylation sites may give us clues about how DC-SIGN mediates entry by binding to the glycans displayed on the virion surface. The binding has not been characterised in any structural sense and as previously mentioned, there is no real consensus on which glycoproteins are involved in

the DC-SIGN binding. Secondly, the mechanism of fusion can now be rationalised in the context of the assembly. Previous attempts at placing glycoprotein structures into EM densities of the glycoprotein layer have been limited by low resolution [47, 101] but the model presented in this thesis is based on a much more detailed cryo-EM map. The shielding of the Gc fusion loop by the Gn and the placement of the fusion loop near the top part of the capsomers is a snapshot of the 'predator-like' readiness of the Gc to lunge forwards and attack the nearest membrane. Undoubtedly, future research will explore the conformational transitions the glycoproteins undergo to achieve membrane merger and undoubtedly a model of the glycoprotein layer in the shielded native state will serve as a fundamental starting point. Transient intermediate states are best captured by cryo-ET and sub-tomogram averaging, but these techniques are currently limited by relatively low resolution compared to single-particle cryo-EM. In these instances, a good starting model can be very beneficial to interpret any low resolution intermediate structures.

The post-fusion structure of the Gc from SFTSV has firmly tied together the fusion mechanism of seemingly unrelated viruses by revealing common conformational changes required for transitioning from pre- to post-fusion. In a way, the structure tells us that there is a lot to be learned about the phleboviral fusion mechanism from the wealth of information which is already available out there for flavi- and alphaviruses. Equally, what is discovered about phleboviral fusion may be applicable to flavi- and alphaviral fusion.

A model of the phleboviral glycoprotein layer and the Gn-Gc heterodimer is an important step towards understanding the assembly of other bunyaviruses. As described in Introduction **1.4.2,** bunyaviruses are largely expected to contain a class II fusion protein [47, 73, 94, 95], and additionally they also contain a Gn protein. The nature of the other bunyaviral Gn proteins is not clear but the mechanism of Gc-mediated fusion is anticipated to be highly similar [82, 95, 171]. If the mechanism of fusion is similar among

bunyaviruses, we can also assume that the function of the bunyaviral Gn is similar. The crystal structure of the RVFV Gn and the link to PUUV Gn and the CHIV E2 fusion chaperone suggests that the Gn proteins of the bunyaviruses have a common function as fusion chaperones. A common function, however, does not offer an explanation as to why bunyaviruses assemble in such diverse ways (see Introduction **1.4.2**).

The bunyaviral Gn proteins may have a common function but the sequence similarity of the Gn proteins, even just between phleboviruses, is significantly lower than for the Gc proteins (see Chapter **3.5**). As the pseudo-atomic model revealed, the Gn sits on top of the Gc and protects it from solvent *in vitro,* and *in vivo* the Gn would be shielding the Gc from the immune system. As discussed in Chapters **3.5** and **3.6**, domain A is the most diverse part of the Gn protein and based on the pseudo-atomic model, it is the part most exposed to the immune system. And as predicted based on N-linked glycosylation sites, it only serves a minor role in protein-protein contacts on the virion surface. If we assume that similar pressures are present for other bunyaviral Gn proteins, then perhaps due to the pressure of immune evasion, the constant evolution and diversification of the bunyaviral Gn as led to mutations which ultimately affected the glycoprotein organisation and have given rise to various assemblies. Since the Gc fusion fold is highly conserved, it is perhaps the Gn which drives of these diverse assemblies, as predicted in a very recent paper reporting on a hantaviral Gn structure [154]. Currently, we lack a more complete structure of any bunyaviral Gn. Both the phleboviral and the hantaviral Gn structures are truncated ectodomains and a structure of a putative domain C is still missing. Obtaining a structure of a full phleboviral Gn ectodomain, or a domain C fragment, would greatly improve our understanding of the viral assembly. As more Gn crystal structures and cryo-EM densities of various bunyaviruses become available, hopefully the question of the diversity of bunyaviral assembly can be addressed.

7.3 Pushing the resolution limits of flexible structures

The RVFV is an interesting example of an unusual quasi-equivalent icosahedron. As discussed in Chapter **5.7.2** the T=12 icosahedral symmetry has a high hexamer complexity number, which is predicted to be a particularly unstable assembly [153]. Forcing the Gn-Gc heterodimer to adopt twelve quasi-equivalent permutations is asking for a lot of assembly flexibility, which is most likely the source of the sample heterogeneity observed Chapter **4.3** in the raw micrographs. Perhaps this sort of flexibility really means that very few perfectly icosahedral particles are ever assembled *in vivo*. And if they are, maybe they are so unstable that they 'crumble' quickly into heterogeneous virions. Obtaining an atomic-level structure of such a sample using traditional single-particle methods is doomed to fail but methods such as localized reconstruction are great tools which can account for some of the inherent flexibility of the sample. In the case of the RVFV, the jump from 13.3 to sub-nm resolutions is a testament of the power of sub-particle or focused image processing which the method allows.

Localized reconstruction has been used successfully for other biological assemblies, such as receptor bound foot and mouth disease virus and COP II vesicles [127, 172]. Few other similar methods have been developed for specific samples, which has boosted resolution beyond expectations [128, 173, 174]. Without cryo-EM and the computational tools developed to deal with conformational and compositional heterogeneity, we would still be stuck with the same problem as X-ray crystallography: structural homogeneity is a limiting factor. The excitement of cryo-EM as a high-resolution structural tool cannot be overstated. As more and more researcher goes into cryo-EM, better tools and better techniques to deal with flexibility will hopefully be developed and the current 'status quo' of flexible structures is likely to change before we know it.

7.4 Glycan mediated receptor binding

Now that the phleboviral glycoprotein layer has been structurally characterised, one of the most obvious avenues of research is to look at virus-receptor complexes. DC-SIGN is the receptor for phleboviruses [37, 100] and a high affinity oligo-mannose binder [167]. Since the binding is glycan dependent, it is largely expected to be as heterogeneous as the glycosylation. This rules out X-ray crystallography as a structural approach, but cryo-EM has been used before on DENV in complex with DC-SIGN [175]. The 25 Å resolution DENV map was very revealing and showed unambiguously which glycosylation sites were bound by comparison with the unbound native virus structure. Such structural clarity of DC-SIGN binding was sought after and here a preliminary account for the phleboviral receptor binding is presented.

A construct of DC-SIGN with 4.5 neck repeats (Gly158-Ala404, see Appendix 10.8) was expressed in a mammalian expression system and purified to homogeneity. DC-SIGN is a type II single-span transmembrane protein which tetramerises via its seven neck-repeat domains and the N-terminal carbohydrate recognition domain (CRD) binds glycans [167, 176]. The expressed construct was expected to form oligomers but no assays had been carried out to determine the oligomeric state. The purified DC-SIGN was incubated with the fixed RVFV for 5 mins prior to plunge-freezing and the virus-receptor complexes were imaged on the 'Polara' cryo-EM microscope. A single-particle data set was processed and localized reconstruction was used to reconstruct the pentamer to 11.9 Å resolution (see Methods **2.6**). See Table **10.2** for data collection and refinement details.

The reconstructed density of the pentamer shows unambiguously the main binding site of the DC-SIGN (Figure **7.1A**). A rigid body fit of the pseudo-atomic model reveals tentacle-like densities near the N438 glycosylation site emanating from a starfish-like

density at the centre of the pentamer (Figure **7.1B**). The density for the DC-SIGN is most likely averaged into such odd densities due to the symmetry miss-match between the five-fold symmetrical pentamer and the putative four-fold symmetrical DC-SIGN construct. A five-fold symmetry applied during reconstruction and hence the density for the DC-SIGN should not be considered for any structural interpretations. The miss-match problem had not been addressed prior to submission of this thesis and the data is thus presented only as preliminary. Interestingly, the DC-SIGN 'tentacles' are close to the N438 site, but not quite at the site as one might expect (Figure **7.1B**). Of course, the glycan composition at this site is not known and it may be that the glycan has a preferred structure and orientation which may account for the unexpected location of the 'tentacle' density

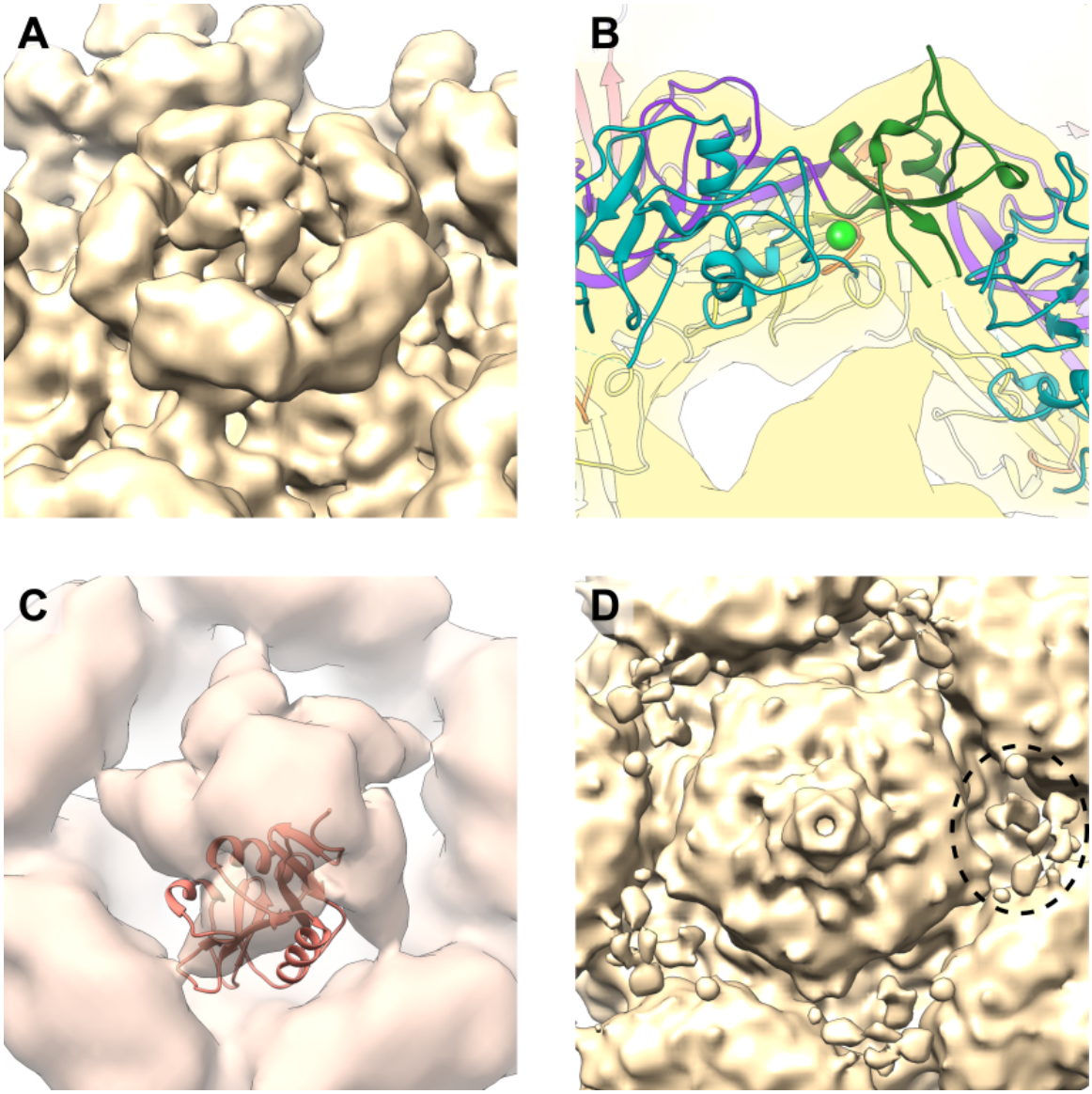


Figure 7.1 DC-SIGN binding to RVFV.
Single-particle cryo-EM and localized reconstruction of RVFV in complex with 4.5 neck repeat DC-SIGN construct reveals receptor binding sites. A) Localized reconstruction of the pentamer to 11.9 Å resolution shows a starfish-like density in the middle of the pentamer. B) The pseudo-atomic model of the pentamer fit into the RVFV-DC-SIGN reconstruction. The reconstruction map has been flattened. A green sphere indicates the N438 glycosylation site on the Gn. The 'tentacle' of the DC-SIGN 'starfish' protrudes close to but not quite at the N438 site. C) The crystal structure of the DC-SIGN CRD (PDB code 1SL5) placed manually near the binding site. This was done only to assess how many copies of the CRD could theoretically fit into the pentamer lumen. The small size of the CRD means that a tetramer of the CRD may fit inside the pentamer lumen. D) The pentamer map shown at a low threshold reveals additional floating densities, shown with a dotted circle, at the quasi-three-fold axes between capsomers, which may be DC-SIGN binding to glycans of domain III of Gc.

Despite the shortcomings of the data presented, there is still something to be learned. The single glycan of the Gn appears to be the main focal point for DC-SIGN binding, which may be rationalised by an avidity effect. The arrangements of the five N438 glycans in a ring like fashion might be an ideal ligand for DC-SIGN as the affinity for an object with two oligo-mannose type glycan is ~10 fold higher than a ligand with only a single oligo-mannose glycan [167]. Additionally, the inner lumen of the pentamer appears to be able to accommodate several CRDs (Figure **7.1C**) so perhaps all four CRDs of a DC-SIGN tetramer could be engaged in glycan binding at a single pentamer. However, this is not the only binding site observed. When the map contour level is lowered, densities start to appear near the quasi-three-fold axis, where the glycans of domain III are located (Figure **7.1D**). It may be that phleboviruses have evolved a focal point for DC-SIGN binding at the centre of each capsomer, but it may also be that the presence of oligomannose at any accessible glycan site across the phleboviral surface is enough for effective host cell entry.

Phlebovirus receptor binding is a highly important and interesting avenue of research which now has been shown to be feasible to study by structural approaches. This work would benefit from comprehensive *in vitro* binding assays to determine affinity and binding kinetics. Such studies could help to optimise the sample preparation and improve future cryo-EM data. Additionally, the DC-SIGN construct design could be optimised. The number of neck-repeats is most likely a determinant of affinity and stability, and the CRD alone may prove to be the best tool to get around the symmetry miss-match. The obvious draw back to these explorations is the containment issue of the RVFV which makes such studies difficult. However, one could use the non-pathogenic UUKV as a model system but the glycan content and spatial organisation may be different and would need to be established. If an oligomeric form of DC-SIGN proves to be the best binding partner, computational methods to deal with the symmetry miss-match need to be explored in much

detail. Methods within the Huiskonen group are being developed to deal with 'relaxed symmetry', i.e. when objects are related by imperfect symmetry such as the heterodimers within the type 2 hexamer of the RVFV. Ultimately, a way to structurally characterise the receptor binding of RVFV has now been established and this promising approach in combination with function studies may give rise to important findings.

7.5 Mechanism of neutralisation

Structural biology has proved to be an important tool in developing therapeutics and rationalising mechanism of neutralisation by small molecules, antibodies or other neutralising agents [177]. As mentioned in the introduction to this thesis, vaccines which are currently in development have shown extremely promising results by activating the B-cell response and inducing production of neutralising antibodies [14, 65]. Furthermore, neutralising antibodies have become a therapeutic tool against viruses in cases where non-immunised individuals get infected and need an effective targeted treatment [178, 179]. A few studies have reported on neutralising antibodies against phleboviruses [180–182] but only one study has gone as far as reporting on a neutralising epitope [101]. The Bowden group is interested in understanding the mechanism of antibody mediated neutralisation and has obtained a large library of neutralising and non-neutralising antibodies directed against the Gn of RVFV. This work has been done with the help of several collaborators including Prof. Dennis Burton (The Scripps Research Institute, California), the late Prof. Richard Elliott (Glasgow University) and Dr. Katie Doores (King's College London). A short account towards understanding such a mechanism comes from the study of a single neutralising antibody.

Two antibodies selected for neutralising potential, termed 49-6 and 93-4, were rescued from hybridomas and cloned into the RabFab vector, which was developed as a

variant of the pHLSec vector to allow for easier cloning of variable regions from rabbit antibodies (see Methods **2.1.12**). The 49-6 and 93-4 antibodies have almost identical complementarity-determining region (CRD), the loops responsible for antigen binding (Figure **7.2**). In fact, the fab-fragment parts of the antibodies have a sequence identity of 96% and as such the antibodies were assumed to have largely the same binding mode. The antibody Fab-fragments were either expressed on their own or co-transfected with the RVFV Gn in a transient mammalian expression system. The fab-fragments or the fab-Gn complexes were purified to homogeneity. The antibody-Gn complexes were used for high-throughput crystallisation, which resulted in a single 49-6-Gn complex structure. The data was only collected a few weeks prior to submission and as a result the structure has not been refined to a publication standard but it will be included in here to complement cryo-EM data. The apo 93-4 antibody Fab-fragment was incubated with the fixed RVFV for 5 mins and the Fab-fragment bound virus was used to collect a single particle data and localized reconstruction was used to reconstruct the pentamer. See Table **10.1** and Table **10.2** and for data collection and refinement parameters.

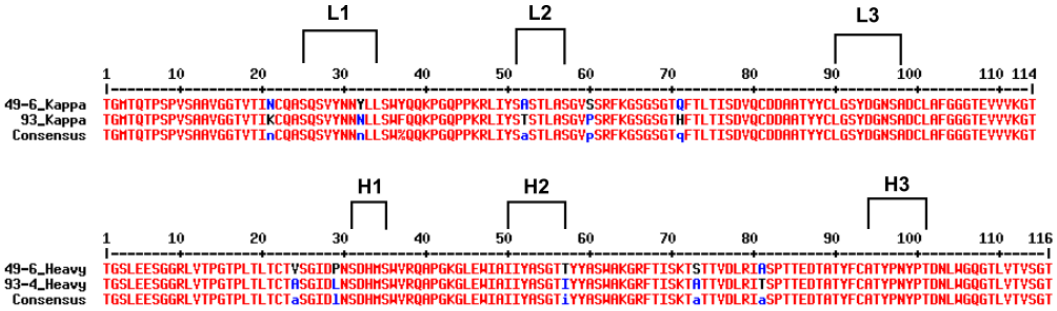


Figure 7.2 Sequence alignment between 49-6 and 93-4 fab-fragment variable region. The variable region of the 49-6 and 93-4 fab-fragments was aligned using MultAlin [183]. Amino acids in red are fully conserved while blue amino acids are not conserved. The canonical CRD loops of each chain have been marked and labelled according to the Kabat scheme, e.g. L1 is the first loop of the light (kappa) chain.

The preliminary crystal structure was roughly refined at the epitope binding interface and the current structure shows good agreement with the electron density map at this site (Figure **7.3A**). Interestingly, as with the previous Gn crystal structure, proteolysis was an important factor in obtaining crystals. However, the only part of Gn which is present in the crystal is a domain B fragment (Figure **7.3B**). The tight crystal packing (~ 30% solvent content) does not allow for any more of the Gn molecule. The antibody has a binding interface of 825 $Å^2$ which involves all six CRD loops of the antibody which are typically involved in epitope binding. The few residues which differ between the 49-6 and 93-4 fab-fragments were not responsible for epitope binding according to the crystal structure and hence the Fab-fragments are assumed to have an identical binding interface. The epitope on the Gn is focused on the part which is not involved in Gn-Gn contacts according to the pseudo-atomic model. As this data was very new at the time of writing this thesis, the details of the binding had not yet been explored.

The localised reconstruction of the Fab-fragment decorated virus shows strong densities for the Fab-fragments, which decorates the pentamers near the Gn-Gn contact sites (Figure **7.3C**). When the pseudo-atomic model of the pentamer is fitted into the density and the crystal structure of the Fab-Gn complex is aligned to the pseudo-atomic model, the constant region of the Fab-fragment slots into the extra density (Figure **7.3D**). The neighbouring Gn appears not to clash with the Fab-fragment and the binding appears not to cause any dramatic conformational changes to the capsomer assembly. From these structures alone it is not immediately clear why the antibody is neutralising. An antibody which would force conformational changes might not have bound to the virus since it has been chemically fixed, so perhaps capturing this particular structure by cryo-EM was a stroke of luck. Undoubtedly, the Gn must change conformation to allow for the Gc to interact with a target membrane and maybe the antibody somehow inhibits this kind of

confomational change. Perhaps the bridging of two binding sites within a single antibody locks two opposing Gn proteins. Or maybe it is something completely different.

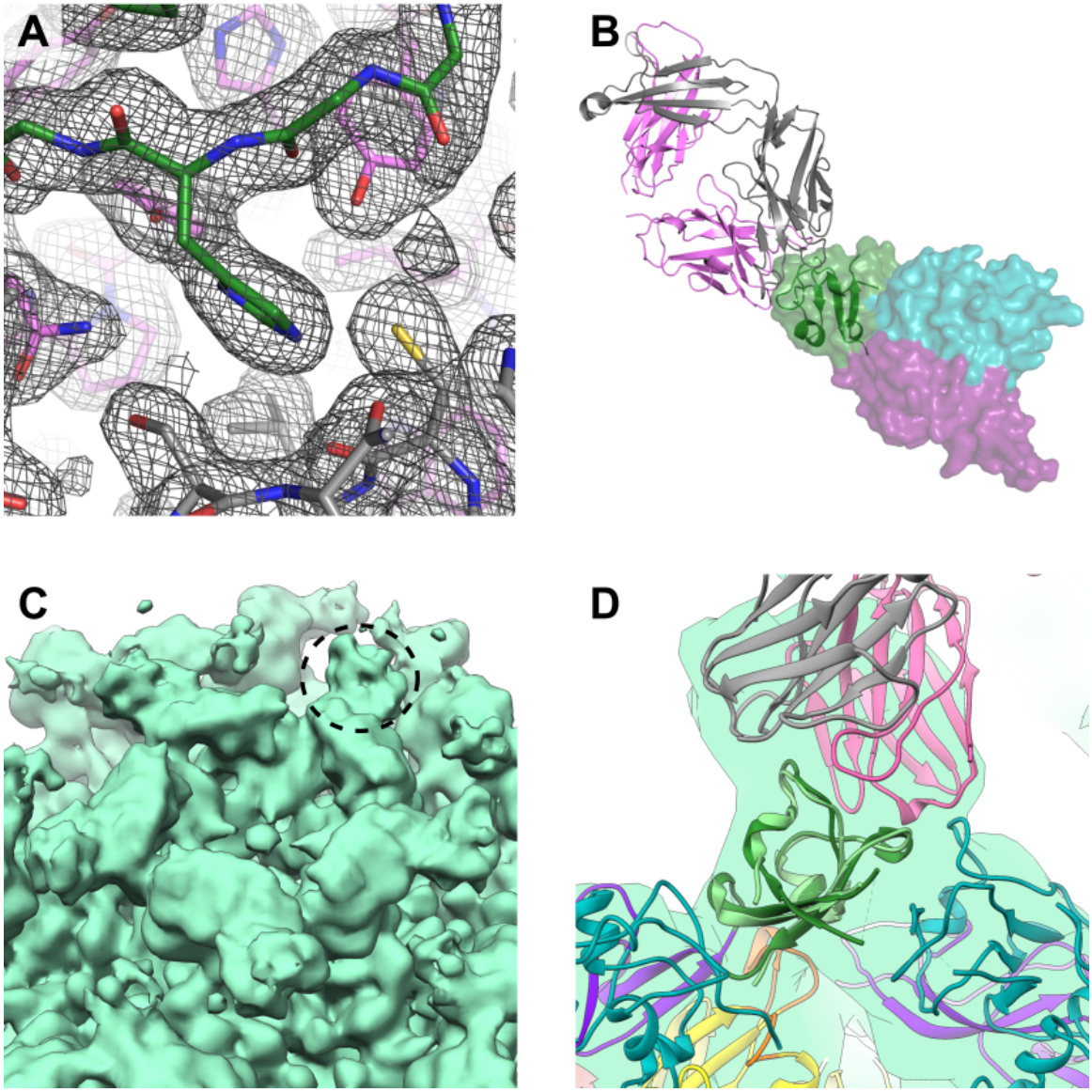


Figure 7.3 Antibody neutralisation of RVFV.
Crystal structure of RVFV-Gn-RabFab49-6 complex and single-particle cryo-EM and localized reconstruction of RVFV-RabFab93-4 complex reveal a neutralising epitope. Fab-fragment heavy chain is coloured in violet and the light (kappa) chain is coloured in grey. A) A zoom-in of the crystal structure of RVFV-Gn-RabFab49-6 at the interface between the Gn domain B (in forest green) and the Fab-fragment. Structure is shown as sticks and an *2Fo-Fc* map is shown as a grey mesh. B) An overview of the crystal structure shown in ribbon. The crystal structure of the RVFV Gn is overlayed on domain B and shown as a transparent surface. C) Localized reconstruction of the pentamer shown in cream-green at 8.7 Å resolution. The extra density for RabFab93-4 is highlighted with a dotted circle. D) A zoom-in of the Fab-fragment density, shown here as a flattened surface. The pseudo-atomic model of the pentamer was fitted into the density and the crystal structure was aligned to the pseudo-atomic model by aligning domain B from the crystal structure to the

pseudo-atomic model. The extra density accounts for a part of the variable region of the Fab-fragment.

On another note, these findings are an independent source of validation for the pseudo-atomic model. Or at least they confirm the location and the orientation of the Gn.

It is evident that the current data needs much more work before the full picture of this mode of neutralisation can be described with confidence. A combination of further neutralisation assays *in vivo* and *in vitro*, structural refinement and binding affinity measurements of both the Fab-fragment and the full-length antibody could ultimately lead to a solid understanding of the neutralisation mechanism. The exciting aspect of this project is the combination of cryo-EM and X-ray crystallography which has the potential to describe epitope binding at the larger whole virus scale and at the detailed atomic-level. This in turn could help rationalise a more general mechanism of neutralisation by studying the various neutralising and non-neutralising antibodies available to the Bowden group. Ultimately, this could lead to the development of therapeutic agents specific to the RVFV and perhaps even cross-neutralising antibodies which might constitute a phlebovirus-wide therapeutic agent.

7.6 Final concluding remarks

This thesis has delivered novel structural data on the phleboviruses. The pseudo-atomic account of the phleboviral glycoprotein layer and the post-fusion structure of the Gc are important milestones in understanding the shielding and priming of the Gc for entry and the final low energy state, a universal theme among class II fusion proteins. The preliminary work on receptor binding and the mechanism of neutralisation gives a flavour of the future of RVFV research and demonstrates how the pseudo-atomic model is a fundamental cornerstone in interpreting such data.

One thing to keep in mind is that phleboviruses are only a part of the larger order of *Bunyavirales*. and what we learn about that single genus does not necessarily apply to the entire order. There is yet much to be discovered about other bunyaviruses. So far there are crystal structures of glycoproteins and cryo-EM maps available for phleboviruses, hantaviruses and orthobunyaviruses, which are only a fraction of the entire order. If we are to better understand these pathobiologically important viruses we must invest more time and effort into studying them. Until more is known about other bunyaviruses, we are forced to make general assumptions based on limited data. The pseudo-atomic model of the RVFV will hopefully serve as a reference point and spark more research into these remarkable viruses.

Finally, the author of this thesis hopes that the work presented herein will be valuable to the field of phleboviruses, bunyaviruses and class II viral fusion research.

8. List of Abbreviations

BMP	Bis(monoacylglycero)phosphate
BUNV	Bunyamwera virus
BVDV	Bovine viral diarrhoea virus
CDR	Complementarity-determining region
CHIV	Chikunguya virus
CMC	Carboxymethylcellulose
Cryo-EM	Cryo electron microscopy
Cryo-ET	Cryo electron tomography
CRD	Carbohydrate recognition domain
DC-SIGN	Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin
DENV	Dengue Virus
DMEM	Dulbecco's modified Eagle media
EM	Electron Microscopy
FCS	Fetal calf serum
IMAC	Immobilised metal affinity chromatography
kDa	Kilodalton
LB	Luria broth
m.o.i.	Multiplicity of infection
mAb	monoclonal antibody
PCR	Polymerase chain reaction
PUUV	Puumala virus
PEI	Polyethylenimine

RT	Room temperature
RVFV	Rift Valley fever virus
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate poly-acryl amide gel electrophoresis
SEC	Size exclusion chromatography
SFV	Semliki forest virus
SINV	Sindbis virus
TBEV	Tick-borne encephalitis virus
TULV	Tula virus
UUKV	Uukuniemi virus

9. References

1. Elliott RM, Brennan B (2014) Emerging phleboviruses. *Curr Opin Virol* 5:50–7.
2. Ikegami T, Makino S (2011) The Pathogenesis of Rift Valley Fever. *Viruses* 3(12):493–519.
3. SCOTT GR, COACKLEY W, ROACH RW, COWDY NR (1963) Rift Valley fever in camels. *J Pathol Bacteriol* 86:229–31.
4. Daubney R, Hudson JR, Garnham PC (1931) Enzootic hepatitis or rift valley fever. An undescribed virus disease of sheep cattle and man from east africa. *J Pathol Bacteriol* 34(4):545–579.
5. Easterday B, Murphy L, Bennett D (1962) Experimental rift valley fever in lambs and sheeps. *Am J Vet Res* 23:1231–1240.
6. Swanepoel R, Manning B, Watt JA (1979) Fatal Rift Valley fever of man in Rhodesia. *Cent Afr J Med* 25(1):1–8.
7. Dodd KA, McElroy AK, Jones MEB, Nichol ST, Spiropoulou CF (2013) Rift Valley fever virus clearance and protection from neurologic disease are dependent on CD4+ T cell and virus-specific antibody responses. *J Virol* 87(11):6161–71.
8. Meegan JM (1979) The Rift Valley fever epizootic in Egypt 1977–1978 1. Description of the epizootic and virological studies. *Trans R Soc Trop Med Hyg* 73(6):618–623.
9. Laughlin LW, Meegan JM, Strausbaugh LJ, Morens DM, Watten RH (1979) Epidemic Rift Valley fever in Egypt: observations of the spectrum of human illness. *Trans R Soc Trop Med Hyg* 73(6):630–3.
10. Seufi AM, Galal FH (2010) Role of Culex and Anopheles mosquito species as potential vectors of rift valley fever virus in Sudan outbreak, 2007. *BMC Infect Dis*

10:65.

11. Outbreak of Rift Valley fever--Yemen, August-October 2000. (2000) *MMWR Morb Mortal Wkly Rep* 49(47):1065–6.
12. Balkhy HH, Memish ZA (2003) Rift Valley fever: an uninvited zoonosis in the Arabian peninsula. *Int J Antimicrob Agents* 21(2):153–7.
13. Chevalier V, Pépin M, Plée L, Lancelot R (2010) Rift Valley fever--a threat for Europe? *Euro Surveill* 15(10):19506.
14. Warimwe GM, et al. (2013) Immunogenicity and efficacy of a chimpanzee adenovirus-vectored Rift Valley fever vaccine in mice. *Virol J* 10:349.
15. Dar O, Hogarth S, McIntyre S (2013) Tempering the risk: Rift Valley fever and bioterrorism. *Trop Med Int Heal* 18(8):1036–1041.
16. Balenghien T (2016) *VMERGE (Emerging viral vector borne diseases) Report Summary* (France) Available at: http://cordis.europa.eu/result/rcn/199643_en.html.
17. Yu X-J, et al. (2011) Fever with thrombocytopenia associated with a novel bunyavirus in China. *N Engl J Med* 364(16):1523–32.
18. Zhang Y-Z, et al. (2012) Hemorrhagic fever caused by a novel Bunyavirus in China: pathogenesis and correlates of fatal outcome. *Clin Infect Dis* 54(4):527–33.
19. Cui N, et al. (2014) Severe fever with thrombocytopenia syndrome bunyavirus-related human encephalitis. *J Infect*. doi:10.1016/j.jinf.2014.08.001.
20. Liu S, et al. (2014) Systematic review of severe fever with thrombocytopenia syndrome: virology, epidemiology, and clinical characteristics. *Rev Med Virol* 24(2):90–102.
21. Yoshikawa T, et al. (2015) Phylogenetic and Geographic Relationships of Severe Fever With Thrombocytopenia Syndrome Virus in China, South Korea, and Japan. *J Infect Dis* 212(6):889–98.

22. Yun S-M, et al. (2014) Severe fever with thrombocytopenia syndrome virus in ticks collected from humans, South Korea, 2013. *Emerg Infect Dis* 20(8):1358–61.

23. Li D, et al. (2015) SFTS Virus in Ticks in an Endemic Area of China. *Am J Trop Med Hyg* 92(4):684–689.

24. Hayasaka D, et al. (2016) Seroepidemiological evidence of severe fever with thrombocytopenia syndrome virus infections in wild boars in Nagasaki, Japan. *Trop Med Health* 44(1):6.

25. Oh S-S, et al. (2016) Detection of Severe Fever with Thrombocytopenia Syndrome Virus from Wild Animals and Ixodidae Ticks in the Republic of Korea. *Vector-Borne Zoonotic Dis* 16(6):408–414.

26. Savage HM, et al. (2013) First detection of heartland virus (Bunyaviridae: Phlebovirus) from field collected arthropods. *Am J Trop Med Hyg* 89(3):445–52.

27. McMullan LK, et al. (2012) A new phlebovirus associated with severe febrile illness in Missouri. *N Engl J Med* 367(9):834–41.

28. Zhang L, et al. (2014) Antibodies against Severe Fever with Thrombocytopenia Syndrome Virus in Healthy Persons, China, 2013. *Emerg Infect Dis* 20(8):1355–7.

29. Adams MJ, et al. (2017) Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2017). *Arch Virol* 162(8):2505–2538.

30. Plyusnin A, Elliott RM (2011) *Bunyaviridae : molecular and cellular biology* (Caister Academic Press) Available at: http://www.caister.com/bunyaviridae [Accessed August 3, 2017].

31. Bouloy M, et al. (2001) Genetic Evidence for an Interferon-Antagonistic Function of Rift Valley Fever Virus Nonstructural Protein NSs. *J Virol* 75(3):1371–1377.

32. Raymond DD, Piper ME, Gerrard SR, Smith JL (2010) Structure of the Rift Valley

fever virus nucleocapsid protein reveals another architecture for RNA encapsidation. *Proc Natl Acad Sci* 107(26):11769–11774.

33. Collett MS, et al. (1985) Complete nucleotide sequence of the M RNA segment of Rift Valley fever virus. *Virology* 144(1):228–45.

34. Gerrard SR, Nichol ST (2007) Synthesis, proteolytic processing and complex formation of N-terminally nested precursor proteins of the Rift Valley fever virus glycoproteins. *Virology* 357(2):124–133.

35. Terasaki K, Won S, Makino S (2013) The C-terminal region of Rift Valley fever virus NSm protein targets the protein to the mitochondrial outer membrane and exerts antiapoptotic function. *J Virol* 87(1):676–82.

36. Hofmann H, et al. (2013) Severe fever with thrombocytopenia virus glycoproteins are targeted by neutralizing antibodies and can use DC-SIGN as a receptor for pH-dependent entry into human and animal cell lines. *J Virol* 87(8):4384–94.

37. Lozach P-Y, et al. (2011) DC-SIGN as a receptor for phleboviruses. *Cell Host Microbe* 10(1):75–88.

38. Svajger U, Anderluh M, Jeras M, Obermajer N (2010) C-type lectin DC-SIGN: an adhesion, signalling and antigen-uptake molecule that guides dendritic cells in immunity. *Cell Signal* 22(10):1397–405.

39. Navarro-Sanchez E, et al. (2003) Dendritic-cell-specific ICAM3-grabbing non-integrin is essential for the productive infection of human dendritic cells by mosquito-cell-derived dengue viruses. *EMBO Rep* 4(7):723–728.

40. Klimstra WB, Nangle EM, Smith MS, Yurochko AD, Ryman KD (2003) DC-SIGN and L-SIGN can act as attachment receptors for alphaviruses and distinguish between mosquito cell- and mammalian cell-derived viruses. *J Virol* 77(22):12022–32.

41. Crispin M, et al. (2014) Uukuniemi Phlebovirus assembly and secretion leave a functional imprint on the virion glycome. *J Virol* 88(17):10244–51.

42. Phoenix I, et al. (2016) N-Glycans on the Rift Valley Fever Virus Envelope Glycoproteins Gn and Gc Redundantly Support Viral Infection via DC-SIGN. *Viruses* 8(5). doi:10.3390/v8050149.

43. de Boer SM, et al. (2012) Heparan sulfate facilitates Rift Valley fever virus entry into the cell. *J Virol* 86(24):13767–71.

44. Harmon B, et al. (2012) Rift Valley fever virus strain MP-12 enters mammalian host cells via caveola-mediated endocytosis. *J Virol* 86(23):12954–70.

45. Lozach P-Y, et al. (2010) Entry of bunyaviruses into mammalian cells. *Cell Host Microbe* 7(6):488–99.

46. Wichgers Schreur PJ, Kortekaas J (2016) Single-Molecule FISH Reveals Non-selective Packaging of Rift Valley Fever Virus Genome Segments. *PLOS Pathog* 12(8):e1005800.

47. Dessau M, Modis Y (2013) Crystal structure of glycoprotein C from Rift Valley fever virus. *Proc Natl Acad Sci U S A* 110(5):1696–701.

48. Bitto D, Halldorsson S, Caputo A, Huiskonen JT (2016) Low pH and anionic lipid-dependent fusion of uukuniemi phlebovirus to liposomes. *J Biol Chem* 291(12). doi:10.1074/jbc.M115.691113.

49. Gallala HD, Sandhoff K (2011) Biological Function of the Cellular Lipid BMP—BMP as a Key Activator for Cholesterol Sorting and Membrane Digestion. *Neurochem Res* 36(9):1594–1600.

50. Drake MJ, et al. (2017) A role for glycolipid biosynthesis in severe fever with thrombocytopenia syndrome virus entry. *PLOS Pathog* 13(4):e1006316.

51. Silvas JA, Popov VL, Paulucci-Holthauzen A, Aguilar P V. (2016) Extracellular

Vesicles Mediate Receptor-Independent Transmission of Novel Tick-Borne Bunyavirus. *J Virol* 90(2):873–886.

52. Elliott RM, Schmaljohn C. (2013) Bunyaviridae. *Fileds Virology*, eds Knipe DM, Howley PM (Wolters Kluwer Business), pp 1244–1282.
53. Gerrard SR, Nichol ST (2002) Characterization of the Golgi retention motif of Rift Valley fever virus G(N) glycoprotein. *J Virol* 76(23):12200–10.
54. Melin L, et al. (1995) The membrane glycoprotein G1 of Uukuniemi virus contains a signal for localization to the Golgi complex. *Virus Res* 36(1):49–66.
55. Overby AK, Popov VL, Pettersson RF, Neve EPA (2007) The cytoplasmic tails of Uukuniemi Virus (Bunyaviridae) G(N) and G(C) glycoproteins are important for intracellular targeting and the budding of virus-like particles. *J Virol* 81(20):11381–91.
56. Rönnholm R (1992) Localization to the Golgi complex of Uukuniemi virus glycoproteins G1 and G2 expressed from cloned cDNAs. *J Virol* 66(7):4525–31.
57. Kuismanen E, Hedman K, Saraste J, Pettersson RF (1982) Uukuniemi virus maturation: accumulation of virus particles and viral antigens in the Golgi complex. *Mol Cell Biol* 2(11):1444–58.
58. Mir MA, Panganiban AT (2006) The bunyavirus nucleocapsid protein is an RNA chaperone: possible roles in viral RNA panhandle formation and genome replication. *RNA* 12(2):272–82.
59. Piper ME, Sorenson DR, Gerrard SR (2011) Efficient Cellular Release of Rift Valley Fever Virus Requires Genomic RNA. *PLoS One* 6(3):e18070.
60. McElroy AK, Nichol ST (2012) Rift Valley fever virus inhibits a pro-inflammatory response in experimentally infected human monocyte derived macrophages and a pro-inflammatory cytokine response may be associated with patient survival during

natural infection. *Virology* 422(1):6–12.

61. Le May N, et al. (2004) TFIIH transcription factor, a target for the Rift Valley hemorrhagic fever virus. *Cell* 116(4):541–50.
62. Billecocq A, et al. (2004) NSs protein of Rift Valley fever virus blocks interferon production by inhibiting host gene transcription. *J Virol* 78(18):9798–806.
63. Habjan M, et al. (2009) NSs protein of rift valley fever virus induces the specific degradation of the double-stranded RNA-dependent protein kinase. *J Virol* 83(9):4365–75.
64. Song P, et al. (2017) Downregulation of Interferon-β and Inhibition of TLR3 Expression are associated with Fatal Outcome of Severe Fever with Thrombocytopenia Syndrome. *Sci Rep* 7(1):6532.
65. Warimwe GM, et al. (2016) Chimpanzee Adenovirus Vaccine Provides Multispecies Protection against Rift Valley Fever. *Sci Rep* 6(1):20617.
66. Rey FA, Heinz FX, Mandl C, Kunz C, Harrison SC (1995) The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution. *Nature* 375(6529):291–298.
67. Kuhn RJ, et al. (2002) Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* 108(5):717–25.
68. Modis Y, Ogata S, Clements D, Harrison SC (2004) Structure of the dengue virus envelope protein after membrane fusion. *Nature* 427(6972):313–9.
69. Allison SL, et al. (1995) Oligomeric rearrangement of tick-borne encephalitis virus envelope proteins induced by an acidic pH. *J Virol* 69(2):695–700.
70. Lescar J, et al. (2001) The Fusion glycoprotein shell of Semliki Forest virus: an icosahedral assembly primed for fusogenic activation at endosomal pH. *Cell* 105(1):137–48.

71. DuBois RM, et al. (2013) Functional and evolutionary insight from the crystal structure of rubella virus protein E1. *Nature* 493(7433):552–556.

72. El Omari K, Iourin O, Harlos K, Grimes JM, Stuart DI (2013) Structure of a pestivirus envelope glycoprotein E2 clarifies its role in cell entry. *Cell Rep* 3(1):30–5.

73. Willensky S, et al. (2016) Crystal Structure of Glycoprotein C from a Hantavirus in the Post-fusion Conformation. *PLOS Pathog* 12(10):e1005948.

74. Pérez-Vargas J, et al. (2014) Structural Basis of Eukaryotic Cell-Cell Fusion. *Cell* 157(2):407–419.

75. Fédry J, et al. (2017) The Ancient Gamete Fusogen HAP2 Is a Eukaryotic Class II Fusion Protein. *Cell* 168(5):904–915.e10.

76. Frame IG, Cutfield JF, Poulter RT (2001) New BEL-like LTR-retrotransposons in Fugu rubripes, Caenorhabditis elegans, and Drosophila melanogaster. *Gene* 263(1–2):219–30.

77. Zhang X, et al. (2013) Cryo-EM structure of the mature dengue virus at 3.5-Å resolution. *Nat Struct Mol Biol* 20(1):105–10.

78. Voss JE, et al. (2010) Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* 468(7324):709–12.

79. Sun S, et al. (2013) Structural analyses at pseudo atomic resolution of Chikungunya virus and antibodies show mechanisms of neutralization. *Elife* 2:e00435.

80. Zhang R, et al. (2011) 4.4 Å cryo-EM structure of an enveloped alphavirus Venezuelan equine encephalitis virus. *EMBO J* 30(18):3854–63.

81. Li S, et al. (2016) A Molecular-Level Account of the Antigenic Hantaviral Surface. *Cell Rep* 15(5):959–67.

82. Guardado-Calvo P, Rey FA (2017) The Envelope Proteins of the Bunyavirales.

Advances in Virus Research, pp 83–118.

83. Harrison SC (2008) Viral membrane fusion. *Nat Struct Mol Biol* 15(7):690–8.

84. Cao S, Zhang W (2013) Characterization of an early-stage fusion intermediate of Sindbis virus using cryoelectron microscopy. *Proc Natl Acad Sci U S A* 110(33):13362–7.

85. Chao LH, Klein DE, Schmidt AG, Peña JM, Harrison SC (2014) Sequential conformational rearrangements in flavivirus membrane fusion. *Elife* 3:e04389.

86. Espósito DLA, Nguyen JB, DeWitt DC, Rhoades E, Modis Y (2015) Physico-chemical requirements and kinetics of membrane fusion of flavivirus-like particles. *J Gen Virol* 96(Pt 7):1702–11.

87. Modis Y (2014) Relating structure to evolution in class II viral membrane fusion proteins. *Curr Opin Virol* 5C:34–41.

88. Fuller SD (1987) The T=4 envelope of sindbis virus is organized by interactions with a complementary T=3 capsid. *Cell* 48(6):923–934.

89. Rosenthal PB, Henderson R (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J Mol Biol* 333(4):721–45.

90. Huiskonen JT, Overby AK, Weber F, Grünewald K (2009) Electron cryo-microscopy and single-particle averaging of Rift Valley fever virus: evidence for GN-GC glycoprotein heterodimers. *J Virol* 83(8):3762–9.

91. Modis Y, Ogata S, Clements D, Harrison SC (2003) A ligand-binding pocket in the dengue virus envelope glycoprotein. *Proc Natl Acad Sci U S A* 100(12):6986–91.

92. Freiberg AN, Sherman MB, Morais MC, Holbrook MR, Watowich SJ (2008) Three-dimensional organization of Rift Valley fever virus revealed by cryoelectron tomography. *J Virol* 82(21):10341–8.

93. Overby AK, Pettersson RF, Grünewald K, Huiskonen JT (2008) Insights into bunyavirus architecture from electron cryotomography of Uukuniemi virus. *Proc Natl Acad Sci U S A* 105(7):2375–9.

94. Plassmeyer ML, et al. (2007) Mutagenesis of the La Crosse Virus glycoprotein supports a role for Gc (1066-1087) as the fusion peptide. *Virology* 358(2):273–82.

95. Guardado-Calvo P, et al. (2016) Mechanistic Insight into Bunyavirus-Induced Membrane Fusion from Structure-Function Analyses of the Hantavirus Envelope Glycoprotein Gc. *PLOS Pathog* 12(10):e1005813.

96. Bowden TA, et al. (2013) Orthobunyavirus ultrastructure and the curious tripodal glycoprotein spike. *PLoS Pathog* 9(5):e1003374.

97. Huiskonen JT, et al. (2010) Electron Cryotomography of Tula Hantavirus Suggests a Unique Assembly Paradigm for Enveloped Viruses. *J Virol* 84(10):4889–4897.

98. Mangala Prasad V, Klose T, Rossmann MG (2017) Assembly, maturation and three-dimensional helical structure of the teratogenic rubella virus. *PLOS Pathog* 13(6):e1006377.

99. Battisti AJ, et al. (2012) Cryo-electron tomography of rubella virus. *J Virol* 86(20):11078–85.

100. Tani H, et al. (2016) Characterization of Glycoprotein-Mediated Entry of Severe Fever with Thrombocytopenia Syndrome Virus. *J Virol* 90(11):5292–301.

101. Wu Y, et al. (2017) Structures of phlebovirus glycoprotein Gn and identification of a neutralizing antibody epitope. *Proc Natl Acad Sci U S A*:201705176.

102. Aricescu AR, Lu W, Jones EY (2006) A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr D Biol Crystallogr* 62(Pt 10):1243–50.

103. Heckman KL, Pease LR (2007) Gene splicing and mutagenesis by PCR-driven

overlap extension. *Nat Protoc* 2(4):924–32.

104. Pan R, et al. (2013) Rabbit anti-HIV-1 monoclonal antibodies raised by immunization can mimic the antigen-binding modes of antibodies derived from HIV-1-infected humans. *J Virol* 87(18):10221–31.

105. Chang VT, et al. (2007) Glycoprotein Structural Genomics: Solving the Glycosylation Problem. *Structure* 15(3):267–273.

106. Walter TS, et al. (2005) A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta Crystallogr Sect D Biol Crystallogr* 61(6):651–657.

107. Winter G (2009) xia2 : an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* 43(1):186–190.

108. Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* 364:215–30.

109. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62(Pt 9):1002–11.

110. McCoy AJ, et al. (2007) *Phaser* crystallographic software. *J Appl Crystallogr* 40(4):658–674.

111. Vagin AA, et al. (2004) REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2184–95.

112. Adams PD, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–21.

113. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of

Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.

114. Chen VB, et al. (2010) *MolProbity* : all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D Biol Crystallogr* 66(1):12–21.
115. Berman HM, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28(1):235–42.
116. Stuart DI, Levine M, Muirhead H, Stammers DK (1979) Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 A. *J Mol Biol* 134(1):109–42.
117. Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5. Available at: http://www.citeulike.org/group/894/article/2344765 [Accessed July 17, 2017].
118. Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM (2006) MUSTANG: A multiple structural alignment algorithm. *Proteins Struct Funct Bioinforma* 64(3):559–574.
119. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–21.
120. Holm L, Laakso LM (2016) Dali server update. *Nucleic Acids Res* 44(W1):W351-5.
121. Grassucci RA, Taylor DJ, Frank J (2007) Preparation of macromolecular complexes for cryo-electron microscopy. *Nat Protoc* 2(12):3239–46.
122. Li X, et al. (2013) Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* 10(6):584–90.
123. Mindell JA, Grigorieff N (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* 142(3):334–47.
124. Scheres S (2013) Single-particle processing in RELION. 1–21.
125. Scheres SHW (2012) RELION: implementation of a Bayesian approach to cryo-EM

structure determination. *J Struct Biol* 180(3):519–30.

126. Chen S, et al. (2013) High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* 135:24–35.

127. Ilca SL, et al. (2015) Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nat Commun* 6:8843.

128. Bai X, Rajendra E, Yang G, Shi Y, Scheres SHW (2015) Sampling the conformational space of the catalytic subunit of human γ-secretase. *Elife* 4. doi:10.7554/eLife.11182.

129. Sun Z, et al. (2017) Double-stranded RNA virus outer shell assembly by bona fide domain-swapping. *Nat Commun* 8:14814.

130. Pettersen EF, et al. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–12.

131. Trabuco LG, Villa E, Schreiner E, Harrison CB, Schulten K (2009) Molecular dynamics flexible fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* 49(2):174–180.

132. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26(16):1781–1802.

133. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–10.

134. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10(6):845–58.

135. Ge P, Zhou ZH (2014) Chaperone fusion proteins aid entropy-driven maturation of class II viral fusion proteins. *Trends Microbiol* 22(2):100–106.

136. Wei X, et al. (2003) Antibody neutralization and escape by HIV-1. *Nature*

422(6929):307–12.

137. Sommerstein R, et al. (2015) Arenavirus Glycan Shield Promotes Neutralizing Antibody Evasion and Protracted Infection. *PLoS Pathog* 11(11):e1005276.

138. Tate MD, et al. (2014) Playing hide and seek: how glycosylation of the influenza virus hemagglutinin can modulate the immune response to infection. *Viruses* 6(3):1294–316.

139. Gavrilovskaya IN, Shepley M, Shaw R, Ginsberg MH, Mackow ER (1998) beta3 Integrins mediate the cellular entry of hantaviruses that cause respiratory failure. *Proc Natl Acad Sci U S A* 95(12):7074–9.

140. Choi Y, et al. (2008) A hantavirus causing hemorrhagic fever with renal syndrome requires gC1qR/p32 for efficient cell binding and infection. *Virology* 381(2):178–183.

141. Krautkramer E, Zeier M (2008) Hantavirus Causing Hemorrhagic Fever with Renal Syndrome Enters from the Apical Surface and Requires Decay-Accelerating Factor (DAF/CD55). *J Virol* 82(9):4257–4264.

142. Muller R, et al. (1995) Characterization of clone 13, a naturally attenuated avirulent isolate of Rift Valley fever virus, which is altered in the small segment. *Am J Trop Med Hyg* 53(4):405–11.

143. Swint-Kruse L, Brown CS (2005) Resmap: automated representation of macromolecular interfaces as two-dimensional networks. *Bioinformatics* 21(15):3327–3328.

144. Caspar DL, Klug A (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* 27:1–24.

145. Johnson JE, Speir JA (1997) Quasi-equivalent viruses: a paradigm for protein assemblies 1 1Edited by T. Richmond. *J Mol Biol* 269(5):665–675.

146. Tang L, Gilcrease EB, Casjens SR, Johnson JE (2006) Highly Discriminatory Binding of Capsid-Cementing Proteins in Bacteriophage L. *Structure* 14(5):837–845.

147. Fokine A, et al. (2004) Molecular architecture of the prolate head of bacteriophage T4. *Proc Natl Acad Sci* 101(16):6003–6008.

148. Zhao G, et al. (2013) Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* 497(7451):643–6.

149. Habjan M, et al. (2009) Efficient production of Rift Valley fever virus-like particles: The antiviral protein MxA can inhibit primary transcription of bunyaviruses. *Virology* 385(2):400–8.

150. Persson R, Pettersson RF (1991) Formation and intracellular transport of a heterodimeric viral spike protein complex. *J Cell Biol* 112(2):257–66.

151. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* 16(5):673–83.

152. R Core Team (2013) R: A language and environment for statistical computing. *R Foundation for Statistical Computing* (Vienna, Austria). Available at: http://www.r-project.org/.

153. Mannige R V, Brooks CL, III (2010) Periodic table of virus capsids: implications for natural selection and design. *PLoS One* 5(3):e9423.

154. Rissanen I, et al. (2017) Structural transitions of the conserved and metastable hantaviral glycoprotein envelope. *J Virol*:JVI.00378-17.

155. Halldorsson S, et al. (2016) Structure of a phleboviral envelope glycoprotein reveals a consolidated model of membrane fusion. *Proc Natl Acad Sci U S A* 113(26):7154–9.

156. Chothia C, Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J* 5(4):823–6.

157. Gulick AM, Horswill AR, Thoden JB, Escalante-Semerena JC, Rayment I (2002) Pentaerythritol propoxylate: a new crystallization agent and cryoprotectant induces crystal growth of 2-methylcitrate dehydratase. *Acta Crystallogr Sect D Biol Crystallogr* 58(2):306–309.

158. Gibbons DL, et al. (2004) Conformational change and protein-protein interactions of the fusion protein of Semliki Forest virus. *Nature* 427(6972):320–5.

159. Roussel A, et al. (2006) Structure and interactions at the viral surface of the envelope protein E1 of Semliki Forest virus. *Structure* 14(1):75–86.

160. Brennan B, et al. (2015) Reverse Genetics System for Severe Fever with Thrombocytopenia Syndrome Virus. *J Virol* 89(6):3026–3037.

161. Mueller D, et al. (2008) Histidine protonation and the activation of viral fusion proteins. *Biochem Soc Trans* 36(1):43–45.

162. Kampmann T, Mueller DS, Mark AE, Young PR, Kobe B (2006) The Role of histidine residues in low-pH-mediated viral membrane fusion. *Structure* 14(10):1481–7.

163. de Boer SM, et al. (2012) Acid-activated structural reorganization of the Rift Valley fever virus Gc fusion protein. *J Virol* 86(24):13642–52.

164. Qin Z -l., Zheng Y, Kielian M (2009) Role of Conserved Histidine Residues in the Low-pH Dependence of the Semliki Forest Virus Fusion Protein. *J Virol* 83(9):4670–4677.

165. Nayak V, et al. (2009) Crystal structure of dengue virus type 1 envelope protein in the postfusion conformation and its implications for membrane fusion. *J Virol* 83(9):4338–44.

166. Fritz R, Stiasny K, Heinz FX (2008) Identification of specific histidines as pH sensors in flavivirus membrane fusion. *J Cell Biol* 183(2):353–361.

167. Mitchell DA, Fadden AJ, Drickamer K (2001) A Novel Mechanism of Carbohydrate Recognition by the C-type Lectins DC-SIGN and DC-SIGNR: SUBUNIT ORGANIZATION AND BINDING TO MULTIVALENT LIGANDS. *J Biol Chem* 276(31):28939–28945.

168. Behrens A-J, et al. (2016) Composition and Antigenic Effects of Individual Glycan Sites of a Trimeric HIV-1 Envelope Glycoprotein. *Cell Rep* 14(11):2695–706.

169. Tassaneetrithep B, et al. (2003) DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells. *J Exp Med* 197(7):823–9.

170. Mazelier M, et al. (2016) Uukuniemi Virus as a Tick-Borne Virus Model. *J Virol* 90(15):6784–98.

171. Simon M, Johansson C, Mirazimi A (2009) Crimean-Congo hemorrhagic fever virus entry and replication is clathrin-, pH- and cholesterol-dependent. *J Gen Virol* 90(Pt 1):210–5.

172. Kotecha A, et al. (2017) Rules of engagement between αvβ6 integrin and foot-and-mouth disease virus. *Nat Commun* 8:15408.

173. Passos DO, et al. (2017) Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science (80-)* 355(6320):89–92.

174. Bostina M, et al. (2007) Single particle cryoelectron tomography characterization of the structure and structural variability of poliovirus–receptor–membrane complex at 30Å resolution. *J Struct Biol* 160(2):200–210.

175. Pokidysheva E, et al. (2006) Cryo-EM reconstruction of dengue virus in complex with the carbohydrate recognition domain of DC-SIGN. *Cell* 124(3):485–93.

176. Feinberg H, Guo Y, Mitchell DA, Drickamer K, Weis WI (2005) Extended neck

regions stabilize tetramers of the receptors DC-SIGN and DC-SIGNR. *J Biol Chem* 280(2):1327–35.

177. Thomas SE, et al. (2017) Structural Biology and the Design of New Therapeutics: From HIV and Cancer to Mycobacterial Infections: A Paper Dedicated to John Kendrew. *J Mol Biol* 429(17):2677–2693.

178. Mendoza EJ, Racine T, Kobinger GP (2017) The ongoing evolution of antibody-based treatments for Ebola virus infection. *Immunotherapy* 9(5):435–450.

179. Nachbagauer R, Krammer F (2017) Universal influenza virus vaccines and therapeutic antibodies. *Clin Microbiol Infect* 23(4):222–228.

180. Besselaar TG, Blackburn NK (1992) The synergistic neutralization of Rift Valley fever virus by monoclonal antibodies to the envelope glycoproteins. *Arch Virol* 125(1–4):239–50.

181. Pifat DY, Osterling MC, Smith JF (1988) Antigenic analysis of Punta Toro virus and identification of protective determinants with monoclonal antibodies. *Virology* 167(2):442–50.

182. Besselaar TG, Blackburn NK (1991) Topological mapping of antigenic sites on the Rift Valley fever virus envelope glycoproteins using monoclonal antibodies. *Arch Virol* 121(1–4):111–24.

183. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16(22):10881–90.

184. Swei A, et al. (2013) The genome sequence of Lone Star virus, a highly divergent bunyavirus found in the Amblyomma americanum tick. *PLoS One* 8(4):e62083.

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

186. Robert X, Gouet P (2014) Deciphering key features in protein structures with the

new ENDscript server. *Nucleic Acids Res* 42(Web Server issue):W320-4.

10. Appendix

10.1 cDNA sequences of codon optimised genes by GenArt (5' to 3')

10.1.1 SFTSV M-segment

>SFTSV_Mseg
CACTATAGGGCGAATTGAAGGAAGGCCGTCAAGGCCGCATACCGGTATGATG
AAAGTGATCTGGTTCAGCAGCCTGATCTGCTTCGTGATCCAGTGCAGCGGCG
ACAGCGGCCCCATCATTTGTGCCGGCCCTATCCACAGCAACAAGAGCGCCGA
CATCCCCCATCTGCTGGGCTACAGCGAGAAGATCTGCCAGATCGACCGGCTG
ATCCACGTGTCCAGCTGGCTGAGAAACCACAGCCAGTTCCAGGGCTACGTGG
GACAGAGAGGCGGCAGATCCCAGGTGTCCTACTACCCCGCCGAGAACAGCTA
CTCCAGGTGGAGCGGACTGCTGAGCCCTTGCGACGCTGATTGGCTGGGCATG
CTGGTCGTGAAGAAGGCCAAGGGCAGCGACATGATCGTGCCTGGCCCTAGCT
ACAAGGGCAAGGTGTTCTTCGAGCGGCCCACCTTCGACGGCTATGTGGGCTG
GGGATGTGGCAGCGGCAAGAGCAGAACAGAGTCCGGCGAGCTGTGCAGCAG
CGATTCTGGCACAAGCAGCGGCCTGCTGCCCAGCGATAGAGTGCTGTGGATC
GGCGACGTGGCCTGCCAGCCTATGACCCCTATCCCCGAGGAAACCTTTCTGG
AACTGAAGTCCTTCAGCCAGAGCGAGTTCCCCGACATCTGCAAGATTGACGG
CATCGTGTTCAACCAGTGCGAGGGCGAGAGCCTGCCCCAGCCTTTTGATGTG
GCCTGGATGGACGTGGGCCACTCCCACAAGATCATCATGCGCGAGCACAAGA
CCAAATGGGTGCAGGAAAGCAGCAGCAAGGACTTCGTGTGCTACAAAGAGG
GCACCGGCCCCTGCAGCGAGAGCGAGGAAAAGACCTGCAAGACCAGCGGCA
GCTGCAGAGGCGACATGCAGTTCTGCAAGGTGGCCGGATGCGAGCACGGCGA
AGAGGCCAGCGAGGCCAAGTGCAGATGTAGCCTGGTGCACAAGCCCGGCGA
GGTGGTGGTGTCTTACGGCGGCATGCGGATCCGGCCTAAGTGCTACGGCTTC
AGCCGGATGATGGCCACCCTGGAAGTGAACCAGCCCGAGCAGAGAATCGGC
CAGTGCACCGGCTGCCACCTGGAATGCATCAATGGCGGCGTGCGGCTGATCA
CCCTGACCAGCGAACTGAAAAGCGCCACCGTGTGCGCCAGCCACTTCTGTAG
CAGCGCCACCTCCGGCAAGAAGTCCACCGAGATCCAGTTCCACAGCGGCTCC
CTCGTGGGCAAGACAGCCATCCATGTGAAGGGGGCCCTGGTGGACGGCACCG
AGTTTACCTTTGAGGGCTCCTGCATGTTCCCCGATGGCTGCGACGCCGTGGAC
TGTACCTTCTGCAGAGAGTTCCTGAAGAACCCCCAGTGCTACCCTGCCAAGA
AATGGCTGTTCATCATCATCGTGATCCTGCTGGGATACGCCGGCCTGATGCTG
CTGACCAACGTGCTGAAGGCCATCGGCATCTGGGGCAGCTGGGTCATCGCCC
CCGTGAAGCTGATGTTCGCCATCATCAAGAAACTGATGCGGACCGTGTCCTG
CCTGATGGGCAAGCTGATGGACCGGGGCAGACAAGTGATCCACGAGGAAAT
CGGCGAGAACAGAGAGGGCAACCAGGACGACGTGCGGATCGAGATGGCCAG
ACCTAGAAGAGTGCGGCACTGGATGTACTCCCCAGTGATCCTGACCTTCCTGG
CCATCGGACTGGCCGAGAGCTGCGACGAAATGGTGCACGCCGACAGCAAGCT
GGTGTCCTGCAGACAGGGCTCCGGCAACATGAAGGAATGCGTGACCACCGGC
AGAGCCCTGCTGCCTGCTGTGAACCCTGGACAGGAAGCCTGCCTGCACTTTAC
CGCCCCTGGCAGCCCTGACAGCAAGTGCCTGAAGATCAAAGTGAAGCGGATC
AACCTGAAGTGCAAGAAAAGCAGCTCCTACTTCGTGCCCGACGCCCGGTCCA
GATGTACCAGCGTGCGGAGATGTAGATGGGCTGGCGATTGCCAGAGCGGCTG
CCCTCCTCACTTCACCAGCAACAGCTTCAGCGACGACTGGGCCGGCAAGATG

GACAGAGCCGGACTGGGCTTTAGCGGCTGCTCTGATGGATGTGGCGGCGCTG
CCTGCGGCTGCTTTAATGCCGCCCCTAGCTGCATCTTCTGGCGGAAATGGGTG
GAAAACCCCCACGGCATCATCTGGAAGGTGTCCCCATGTGCCGCCTGGGTGC
CATCTGCCGTGATCGAGCTGACAATGCCCAGCGGCGAAGTGCGGACCTTCCA
CCCTATGAGCGGCATCCCCACCCAGGTGTTCAAGGGCGTGTCCGTGACATAC
CTGGGCTCCGACATGGAAGTGTCCGGCCTGACCGACCTGTGCGAGATTGAGG
AACTGAAATCCAAGAAGCTGGCCCTGGCCCCCTGTAACCAGGCCGGAATGGG
AGTCGTGGGAAAAGTGGGCGAGATTCAGTGCAGCTCCGAGGAAAGCGCCCG
GACCATCAAGAAGGACGGCTGCATCTGGAACGCCGACCTCGTGGGAATTGAG
CTGAGAGTGGATGACGCCGTGTGTTACAGCAAGATCACCAGCGTGGAAGCCG
TGGCCAACTACAGCGCCATCCCTACCACAATCGGCGGCCTGAGATTCGAGCG
GAGCCACGATAGCCAGGGCAAGATCAGCGGAAGCCCCCTGGACATCACCGCC
ATTAGAGGCAGCTTCTCCGTGAACTACCGGGGCCTGAGACTGAGCCTGAGCG
AGATCACAGCCACCTGTACCGGGGAAGTGACCAATGTGTCCGGCTGTTACAG
CTGCATGACCGGCGCCAAGGTGTCCATCAAGCTGCACTCCAGCAAGAACAGC
ACCGCCCACGTGCGGTGCAAGGGCGACGAGACAGCCTTCTCTGTGCTGGAAG
GCGTGCACAGCTACACCGTGTCCCTGAGCTTCGACCACGCCGTGGTGGATGA
GCAGTGCCAGCTGAATTGTGGCGGCCACGAGTCCCAAGTGACCCTGAAGGGA
AACCTGATCTTTCTGGACGTGCCCAAGTTCGTGGACGGCTCCTACATGCAGAC
CTACCACAGCACCGTGCCCACAGGCGCCAACATCCCTAGCCCTACCGACTGG
CTGAACGCCCTGTTCGGCAACGGCCTGAGCAGATGGATCCTGGGCGTGATCG
GAGTGCTGCTGGGAGGACTGGCCCTGTTCTTTATGATCATGAGCCTGTTCAAG
CTGGGCACCAAACAGGTGTTCCGGTCCCGGACAAAGCTGGCCGGTACCCTGG
GCCTCATGGGCCTTCCTTTCACTGCCCGCTTTCCAG

10.1.2 RVFV M-segment

>RVFV_Mseg
CGAATTGAAGGAAGGCCGTCAAGGCCGCATGAGCTCACCGGTATGTACGTGC
TGCTGACCATCCTGATCAGCGTGCTGGTGTGCGAGGCCGTGATCAGAGTGTCC
CTGAGCAGCACCCGGGAGGAAACCTGCTTCGGCGACAGCACCAACCCCGAGA
TGATCGAGGGCGCCTGGGACAGCCTGCGGGAGGAAGAGATGCCCGAGGAAC
TGAGCTGCAGCGTGTCCGGCATCCGGGAAGTGAAAACCAGCAGCCAGGAACT
GTACCGGGCCCTGAAGGCCATCATTGCCGCCGACGGCCTGAACAACATCACC
TGTCACGGCAAGAACCCCGAGGACAAGATCAGCCTGATCAAGGGCCCTCCCC
ACAAGAAACGCGTGGGCATCGTGCGCTGCGAGCGGAGAAGGGACGCCAAGC
AGATCGGCAGAGAAACCATGGCCGGCATTGCCATGACAGTGCTGCCCGCCCT
GGCCGTGTTTGCTCTGGCCCCCGTGGTGTTCGCCGAGGACCCCCACCTGAGAA
ACAGACCCGGCAAGGGCCACAACTACATCGACGGCATGACCCAGGAAGATG
CCACTTGCAAGCCCGTGACCTACGCTGGCGCCTGCAGCAGCTTCGATGTGCTG
CTGGAAAAGGGCAAGTTCCCCCTGTTCCAGAGCTACGCCCACCACAGAACCC
TGCTGGAAGCCGTGCACGACACCATTATCGCCAAGGCCGACCCCCCTAGCTG
TGATCTGCAGAGCGCCCACGGCAACCCCTGCATGAAAGAAAAGCTGGTGATG
AAGACCCACTGCCCCAACGACTACCAGTCCGCCCACTACCTGAACAACGACG
GCAAGATGGCCAGCGTGAAGTGCCCCCCCAAGTACGAGCTGACCGAGGACTG
CAACTTCTGTCGGCAGATGACCGGCGCCAGCCTGAAGAAGGGCAGCTACCCA
CTGCAGGACCTGTTCTGCCAGAGCAGCGAGGACGACGGCAGCAAGCTGAAA

ACAAAGATGAAGGGCGTGTGCGAAGTGGGCGTGCAGGCCCTGAAGAAGTGC
GACGGCCAGCTGAGCACAGCCCACGAGGTGGTGCCCTTCGCCGTGTTCAAGA
ACAGCAAGAAGGTGTACCTGGACAAGCTGGACCTGAAAACCGAGGAAAACC
TGCTGCCCGACTCCTTCGTGTGCTTCGAGCACAAGGGCCAGTACAAGGGCAC
CATGGACAGCGGCCAGACCAAGCGGGAGCTGAAGTCCTTCGACATCAGCCAG
TGCCCCAAGATCGGCGGCCACGGCAGCAAAAAGTGTACCGGCGACGCCGCCT
TCTGTAGCGCCTACGAGTGCACCGCCCAGTACGCCAACGCCTACTGCAGCCA
CGCCAACGGCAGCGGCATCGTGCAGATCCAGGTGTCCGGCGTGTGGAAGAAA
CCCCTGTGCGTGGGCTACGAGCGGGTGGTGGTGAAGAGAGAGCTGAGCGCCA
AGCCCATCCAGAGAGTGGAGCCCTGCACCACCTGTATCACCAAGTGCGAGCC
CCATGGCCTGGTGGTGCGGAGCACCGGCTTCAAGATCAGCAGCGCCGTGGCC
TGTGCTAGCGGCGTGTGTGTGACCGGCAGCCAGAGCCCTAGCACCGAGATCA
CCCTGAAGTACCCCGGCATCTCTCAGAGCAGCGGCGGAGACATCGGAGTGCA
CATGGCCCACGACGACCAGAGCGTGTCCAGCAAGATCGTGGCCCACTGCCCA
CCCCAGGATCCTTGCCTGGTGCACGGCTGCATCGTGTGTGCCCACGGCCTGAT
CAACTACCAGTGCCACACCGCCCTGAGCGCCTTCGTGGTGATCTTCGTGTTCA
GCAGCACCGCCATCATCTGCCTGGCCGTGCTGTATCGGGTGCTGAAGTGCCTG
AAGATCGCCCCCAGAAAGGTGCTGAACCCCCTGATGTGGATCACAGCCTTCA
TCCGCTGGGTGTACAAGAAAATGGTGGCCCGGGTGGCCGACAACATCAACCA
GGTGAACCGGGAGATCGGCTGGATGGAAGGCGGACAGCTGGCCCTGGGCAA
CCCTGCCCCTATCCCCAGACACGCCCCCATCCCCCGGTACAGCACCTACCTGA
TGCTGCTGCTGATCGTGTCCTACGCCAGCGCTTGCTCTGAGCTGATCCAGGCC
AGCAGCCGGATCACCACATGCAGCACCGAGGGCGTGAACACCAAGTGCAGA
CTGAGCGGCACTGCCCTGATCAGAGCTGGCTCTGTGGGAGCCGAGGCCTGCC
TGATGCTGAAAGGCGTGAAAGAGGACCAGACCAAGTTTCTGAAGATCAAGAC
CGTGTCCAGCGAGCTGTCCTGCAGAGAGGGCCAGAGCTACTGGACCGGCAGC
TTCAGCCCCAAGTGCCTGAGCAGCCGCAGATGTCACCTGGTGGGCGAGTGCC
ACGTGAACAGATGCCTGAGCTGGCGGGACAACGAGACAAGCGCCGAGTTCA
GCTTCGTGGGCGAAAGCACCACCATGCGGGAGAACAAGTGCTTTGAGCAGTG
CGGCGGCTGGGGCTGCGGCTGCTTCAACGTGAACCCCAGCTGCCTGTTCGTGC
ACACCTACCTGCAGAGCGTGCGGAAAGAGGCCCTGCGGGTGTTCAACTGCAT
CGACTGGGTGCACAAGCTGACCCTGGAAATCACCGACTTCGACGGCTCCGTG
AGCACAATCGACCTGGGCGCCAGCAGCTCCAGATTCACCAACTGGGGCTCCG
TGTCCCTGTCTCTGGACGCCGAGGGCATCAGCGGCAGCAACAGCTTCAGCTTC
ATCGAGAGCCCTGGCAAGGGCTACGCCATCGTGGACGAGCCCTTCAGCGAGA
TCCCCAGGCAGGGCTTTCTGGGCGAGATCCGGTGCAACAGCGAGAGCAGCGT
GCTGTCCGCCCACGAGAGCTGCCTGAGAGCCCCCAACCTGGTGTCCTACAAG
CCCATGATCGACCAGCTGGAATGCACCACCAACCTGATCGACCCCTTTGTGGT
GTTCGAGAGAGGCAGCCTGCCCCAGACCCGGAACGACAAGACCTTCGCCGCC
AGCAAGGGCAATCGGGGGGTGCAGGCCTTTAGCAAGGGCTCCGTGCAGGCCG
ACCTGACCCTGATGTTCGACAACTTCGAGGTGGACTTCGTGGGAGCCGCCGT
GAGCTGTGATGCCGCCTTCCTGAACCTGACCGGCTGCTACAGCTGCAATGCCG
GCGCTAGAGTGTGTCTGTCCATCACCAGCACCGGCACAGGCACACTGAGCGC
CCACAACAAGGACGGCAGCCTGCACATCGTGCTGCCCAGCGAGAACGGCACC
AAGGACCAGTGCCAGATCCTGCACTTCACCGTGCCCGAGGTGGAAGAGGAGT
TCATGTACAGCTGCGACGGCGACGAAAGACCCCTGCTGGTGAAGGGCACCCT
GATCGCCATCGACCCTTTCGACGACAGAAGAGAGGCTGGCGGCGAGTCCACC
GTGGTGAACCCCAAGAGCGGCAGCTGGAACTTCTTCGACTGGTTCAGCGGCC
TGATGTCTTGGTTTGGCGGCCCTCTGAAAACCATCCTGCTGATCTGCCTGTAC
GTGGCCCTGAGCATCGGCCTGTTCTTTCTGCTGATCTATCTGGGACGGACCGG

CCTGTCAAAAATGTGGCTGGCTGCCACCAAGAAGGCTAGCGGTACCCTGGGC
CTCATGGGCCTTCCTTTCACTGCC

10.1.3 Human DC-SIGN

>DC-SIGN
CTCACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCTAGGCGCGCCAATG
AGCGACAGCAAAGAGCCCCGGCTGCAGCAGCTGGGCCTGCTGGAAGAGGAA
CAGCTGCGGGGCCTGGGCTTCCGGCAGACCAGAGGCTACAAGAGCCTGGCCG
GCTGTCTGGGCCACGGCCCTCTGGTGCTGCAGCTGCTGAGCTTTACCCTGCTG
GCCGGACTGCTGGTCCAGGTGTCCAAGGTGCCCAGCAGCATCAGCCAGGAAC
AGAGCCGGCAGGACGCCATCTACCAGAACCTGACCCAGCTGAAGGCCGCCGT
GGGCGAGCTGAGCGAGAAGTCCAAGCTGCAGGAAATCTATCAGGAACTCACC
CAGCTCAAAGCAGCTGTGGGAGAACTGCCCGAGAAAAGCAAACTCCAGGAA
ATCTACCAGGAACTGACAAGACTGAAAGCTGCCGTCGGAGAGCTGCCTGAGA
AATCTAAGCTGCAGGAAATCTACCAGGAACTGACCTGGCTGAAGGCTGCAGT
GGGGGAACTGCCTGAAAAGTCAAAAATGCAGGAAATCTACCAGGAACTCAC
AAGGCTCAAGGCCGCTGTCGGGGAGCTGCCAGAAAAGAGTAAACAGCAGGA
AATCTATCAGGAACTGACTCGCCTGAAGGCCGCTGTCGGCGAACTCCCTGAG
AAAAGCAAGCAGCAGGAAATCTACCAGGAACTCACTCGCCTCAAAGCTGCAG
TCGGAGAACTCCCCGAAAAATCCAAACAGCAGGAAATCTACCAGGAACTGAC
CCAGCTCAAGGCTGCCGTCGAGAGACTGTGCCACCCCTGCCCCTGGGAGTGG
ACCTTCTTCCAGGGCAACTGCTACTTCATGAGCAACAGCCAGCGGAACTGGC
ACGACAGCATCACCGCCTGCAAAGAAGTGGGCGCCCAGCTGGTGGTCATCAA
GAGCGCCGAGGAACAGAACTTCCTGCAGCTCCAGAGCAGCAGAAGCAACAG
ATTCACCTGGATGGGCCTGAGCGACCTGAACCAGGAAGGCACCTGGCAGTGG
GTGGACGGCAGCCCCCTGCTGCCCAGCTTCAAGCAGTATTGGAACCGGGGCG
AGCCCAACAACGTGGGCGAAGAGGATTGCGCCGAGTTCAGCGGCAACGGCT
GGAACGACGACAAGTGCAACCTGGCCAAGTTCTGGATCTGCAAGAAGTCCGC
CGCCAGCTGCTCCCGGGATGAGGAACAGTTTCTGAGCCCTGCCCCAGCCACC
CCCAATCCTCCACCAGCTATTAATTAACTGGCCTCATGGGCCTTCCGCTCACT
GCCCGCTTTCCAGT

10.2 List of constructs and primers for SFTSV Gc and RVFV Gn

Construct	Start site – end site (length)[a]	Forward primer sequence (5' to 3')[b,c]	Reverse primer sequence (5' to 3')
SFTSV Gc			
D995	C563 – D995 (433)	cgcaccggtTGCGACGAAATGGTGCACGCC	gcgggtaccGTCCAGAAAGATCAGGTTTCC
V996	C563 – V996 (434)	cgcaccggtTGCGACGAAATGGTGCACGCC	cgcggtaccCACGTCCAGAAAGATCAGGTTTCCC
P997	C563 – P997 (435)	cgcaccggtTGCGACGAAATGGTGCACGCC	gcgggtaccGGGCACGTCCAGAAAG
G1002	C563 – G1002 (440)	cgcaccggtTGCGACGAAATGGTGCACGCC	gcgggtaccGCCGTCCACGAACTTG
Q1006	C563 – Q1006 (444)	cgcaccggtTGCGACGAAATGGTGCACGCC	gcgggtaccCTGCATGTAGGAGCCG
S1010	C563 – S1010 (448)	cgcaccggtTGCGACGAAATGGTGCACGCC	gcgggtaccGCTGTGGTAGGTCTGC
T1014	C563 – T1014 (452)	cgcaccggtTGCGACGAAATGGTGCACGCC	gcgggtaccTGTGGGCACGGTG
N1017	C563 – N1017 (455)	cgcaccggtTGCGACGAAATGGTGCACGCC	gcgggtaccGTTGGCGCCTGTG
S1020	C563 – S1020 (458)	cgcaccggtTGCGACGAAATGGTGCACGCC	gcgggtaccGCTAGGGATGTTGGCG
D1023	C563 – D1023 (461)	cgcaccggtTGCGACGAAATGGTGCACGCC	gcgggtaccGTCGGTAGGGCTAGG
A1027	C563 – A1027 (465)	cgcaccggtTGCGACGAAATGGTGCACGCC	gcgggtaccGGCGTTCAGCCAGT
R10353	C563 – R1037 (475)	cgcaccggtTGCGACGAAATGGTGCACGCC	gcgggtaccTCTGCTCAGGCCGT

RVFV Gn			
FL1	E154 – D560 (407)	cgcaccggtGAGGACCCCCACCTGAGAAACAG	cgcggtaccATCCTGGGGTGGGCAGTGGG
FL2	E154 – I568 (415)	cgcaccggtGAGGACCCCCACCTGAGAAACAG	cgcggtaccGATGCAGCCGTGCACCAGGC
FL3	E154 – T581 (428)	cgcaccggtGAGGACCCCCACCTGAGAAACAG	cgcggtaccGGTGTGGCACTGGTAGTTGATCAGG
FL4	Q174 – D560 (387)	cgcaccggtCAGGAAGATGCCACTTGCAAGCCC	cgcggtaccATCCTGGGGTGGGCAGTGGG
FL5	Q174 – I568 (395)	cgcaccggtCAGGAAGATGCCACTTGCAAGCCC	cgcggtaccGATGCAGCCGTGCACCAGGC
FL6	Q174 – T581 (408)	cgcaccggtCAGGAAGATGCCACTTGCAAGCCC	cgcggtaccGGTGTGGCACTGGTAGTTGATCAGG
FL7	D192 – D560 (369)	cgcaccggtGATGTGCTGCTGGAAAAGGGCAAG	cgcggtaccATCCTGGGGTGGGCAGTGGG
FL8	D192 – I568 (377)	cgcaccggtGATGTGCTGCTGGAAAAGGGCAAG	cgcggtaccGATGCAGCCGTGCACCAGGC
FL9	D192 – T581 (390)	cgcaccggtGATGTGCTGCTGGAAAAGGGCAAG	cgcggtaccGGTGTGGCACTGGTAGTTGATCAGG
NT1	E154 – Q380 (227)	cgcaccggtGAGGACCCCCACCTGAGAAACAG	cgcggtaccCTGGCCCTTGTGCTCGAAGCAC
NT2	E154 – D398 (245)	cgcaccggtGAGGACCCCCACCTGAGAAACAG	cgcggtaccGTCGAAGGACTTCAGCTCCCGC
NT3	Q174 – Q380 (207)	cgcaccggtCAGGAAGATGCCACTTGCAAGCCC	cgcggtaccCTGGCCCTTGTGCTCGAAGCAC
NT4	Q174 – D398 (225)	cgcaccggtCAGGAAGATGCCACTTGCAAGCCC	cgcggtaccGTCGAAGGACTTCAGCTCCCGC
NT5	D192 – Q380 (189)	cgcaccggtGATGTGCTGCTGGAAAAGGGCAAG	cgcggtaccCTGGCCCTTGTGCTCGAAGCAC
NT6	D192 – D398 (207)	cgcaccggtGATGTGCTGCTGGAAAAGGGCAAG	cgcggtaccGTCGAAGGACTTCAGCTCCCGC
CT1	Q380 – D560 (181)	cgcaccggtCAGTACAAGGGCACCATGGACAGC	cgcggtaccATCCTGGGGTGGGCAGTGGG
CT2	Q380 – I568 (189)	cgcaccggtCAGTACAAGGGCACCATGGACAGC	cgcggtaccGATGCAGCCGTGCACCAGGC
CT3	Q380 – T581 (202)	cgcaccggtCAGTACAAGGGCACCATGGACAGC	cgcggtaccGGTGTGGCACTGGTAGTTGATCAGG
CT4	D398 – D560 (163)	cgcaccggtGACATCAGCCAGTGCCCCAAGATC	cgcggtaccATCCTGGGGTGGGCAGTGGG
CT5	D398 – I568 (171)	cgcaccggtGACATCAGCCAGTGCCCCAAGATC	cgcggtaccGATGCAGCCGTGCACCAGGC
CT6	D398 – T581 (184)	cgcaccggtGACATCAGCCAGTGCCCCAAGATC	cgcggtaccGGTGTGGCACTGGTAGTTGATCAGG

[a]Start and end site numbers are based on amino acid sequences M-segments of SFTSV and RVFV.
[b]Underlined nucleotide sequence corresponds to restriction sites: Forward primer contains *Age* I restriction site and reverse primer contains *Kpn*I restriction site.
[c]Lowercase nucleotide sequences contains the stabilising flanking region (cgc) and restriction site, while the uppercase sequence is complimentary to region on cDNA.

10.3 List of primers for site-directed mutagenesis of SFTSV Gc V996

	Primer A (5' to 3')		**Primer D (5' to 3')**
	cgcaccggtTGCGACGAAATGGTGCACGCC		cgcggtaccCACGTCCAGAAAGATCAGGTTTCCC
Mutant[a]	**Primer B (5' to 3')**	**Primer C (5' to 3')**	**Primer O (5' to 3')**
H663M	AGGAGGGCAGCCGCTCTGG	TTCACCAGCAACAGCTTCAGCGAC	CCAGAGCGGCTGCCCTCCTATGTTCAC CAGCAACAGCTTCAGCGAC
H747M	CCGCTCATAGGCATGAAGGTCC	GGACCTTCATGCCTATGAGCGG	
H940M	GGCGGTGCTGTTCTTGCTGGAG	GTGCGGTGCAAGGGCGACG	CTCCAGCAAGAACAGCACCGCCATGG TGCGGTGCAAGGGCGACG
W652S	CAATCGCCAGCCGATCTACATCTCC	GGAGATGTAGATCGGCTGGCGATTG	
A694S	CAGGCAGAGCCGCCACATC	GATGTGGCGGCTCTGCCTG	
F699S	GGCGGCATTAGAGCAGCCGCAG	CTGCGGCTGCTCTAATGCCGCC	
A694F and F699A	CATTAGCGCAGCCGCAGGCAAAGCCG	CGGCTTTGCCTGCGGCTGCGCTAATG	
C563M	cgcaccggtatgGACGAAATGGTGCACGCCGAC		
C604M	GGCTTCCTGTCCAGGGTTCACAG	CTGCACTTTACCGCCCCTGGC	CTGTGAACCCTGGACAGGAAGCCATG CTGCACTTTACCGCCCCTGGC
C617M	CTTGCTGTCAGGGCTGCCAGG	CTGAAGATCAAAGTGAAGCGGATCAACC	CCTGGCAGCCCTGACAGCAAGATGCT GAAGATCAAAGTGAAGCGGATCAACC

[a]Numbers denote the site of mutation within the M-segment of SFTSV. Letters stand for amino acid changes, e.g. H663M is a histidine to methionine mutation. Mutated codons are underscored

10.4 List of rabbit IgG specific primers

Kappa chain forward primers (5' to 3')		Kappa chain reverse primers (5' to 3')	
RVK1	GCGCCGGAGCTCGTGATGACCCAGACTCCA	RCK1	GCGCCGTCTAGACTAACAGTCACCCCTATTGAAGC
RVK2	GCGCCGGAGCTCGATATGACCCAGACTCCA	RCK2	GCGCCGTCTAGACTAACAGTTCTTCCTACTGAAGC
		IGκ	GATGCCAGTTGTTTGGGTGGT
Heavy chain forward primers (5' to 3')		**Heavy chain reverse primers (5' to 3')**	
RHFout	ATGGAGACTGGGCTGCGCTGGCTTC	RHRout1	GTCCTTGGGTTTTGGGGGAAAGATGAA
		RHRout2	GTCCCCGCAGCAGGGGGCCAGTGGGAA
		RHRout3	CTCCTCCCGGGGAGGGCCCATGGTGTA

10.5 List of primers for RabFab cloning with InFusion

Construct name	Forward primer (5' to 3')	Reverse primer (5' to 3')
49-6 Kappa	gtagctgaaaccggtATGACCCAGACTCCATCCCCCG	ggccacaggggtaccTTTGACCACCACCTCGGTCCC
49-6 Heavy	gtagctgaaaccggtTCGCTGGAGGAGTCCGGGG	cttgggctgggtaccGGAGACGGTGACCAGGGTGC
93-4 Kappa	gtagctgaaaccggtATGACCCAGACTCCATCGCCCG	ggccacaggggtaccTTTGACCACCACCTCGGTCCC
93-4 Heavy	gtagctgaaaccggtTCGCTGGAGGAGTCCGGGG	cttgggctgggtaccGGAGACGGTGACCAGGGTGC

Uppercase letters in the primer are regions complimentary to the gene to be amplified, lowercase letters are regions complimentary to the plasmid used for InFusion cloning with a recombinase strategy.

10.6 List of amplified rabbit IgG sequences:

> 49-6_Kappa_DNA
ATGACCCAGACTCCATCCCCCGTGTCTGCAGCTGTGGGAGGCACAGTCACCA
TCAACTGCCAGGCCAGTCAGAGTGTTTATAATAACTACCTCTTATCCTGGTAT
CAGCAAAAACCAGGGCAGCCTCCCAAGCGCCTGATCTATTCTGCATCCACTCT
GGCATCTGGGGTCTCATCGCGGTTCAAAGGCAGTGGATCTGGGACACAGTTC
ACTCTCACCATCAGCGACGTGCAGTGTGACGATGCTGCCACTTATTATTGTCT
AGGCAGTTATGACGGTAATAGTGCTGATTGCCTTGCTTTCGGCGGAGGGACC
GAGGTGGTGGTCAAAGGTGATCCAGTTGCACCTACTGTCCTCATCTTCCCACC
AGCTGCTGATCAGGTGGCAACTGGAACAGTCACCATCGTGTGTGTGGCGAAT
AAATACTTTCCCGATGTCACCGTCACCTGGGAGGTGGATGGCACCACCCAAA
CAA

>49-6_Kappa_translated
MTQTPSPVSAAVGGTVTINCQASQSVYNNYLLSWYQQKPGQPPKRLIYSASTLAS
GVSSRFKGSGSGTQFTLTISDVQCDDAATYYCLGSYDGNSADCLAFGGGTEVVV
KGDPVAPTVLIFPPAADQVATGTVTIVCVANKYFPDVTVTWEVDGTTQT

>49-6_Heavy_DNA
GGAGACTGGGCTGCGCTGGCTTCTCCTGGTCGCTGTGCTCAAAGGTGTCCAGT
GTCAGTCGCTGGAGGAGTCCGGGGGTCGCCTGGTCACGCCTGGGACACCCCT
GACACTCACCTGCACAGTCTCTGGAATCGACCCCAATAGCGACCACATGAGC
TGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAATGGATCGCAATTATTTATG
CTAGTGGTACCACTTACTACGCGAGCTGGGCAAAAGGCCGATTCACCATCTC
CAAAACCTCGACCACGGTGGATCTGAGGATCGCCAGTCCGACAACCGAGGAC
ACGGCCACCTATTTCTGTGCCACTTATCCTAATTATCCTACTGATAACTTGTGG
GGCCAAGGCACCCTGGTCACCGTCTCCTCAGGGCAACCTAAGGCTCCATCAG
TCTTCCCACTGGCCCCCTGCTGCGGGGACA

>49-6_Heavy_translated
ETGLRWLLLVAVLKGVQCQSLEESGGRLVTPGTPLTLTCTVSGIDPNSDHMSWV
RQAPGKGLEWIAIIYASGTTYYASWAKGRFTISKTSTTVDLRIASPTTEDTATYFC
ATYPNYPTDNLWGQGTLVTVSSGQPKAPSVFPLAPCCGD

>55-3_Kappa_DNA
TGCGCCGGAGCTCGATATGACCCAGACTCCACCCTCGGTGTCTGCAGCTGTGG
GAGACACAGTCACCATCAAGTGCCAGTCCAGTCAGAGTGTTGGTGATGTCAA
CGAATTATCCTGGTATCAGCAGAAACCAGGGCAGCCTCCCAAGCTCCTGGTC
TACAGGGCATCCAATCTGGCATCTGGTATCCCATCGCGGTTCAAGGGCAGTG
GATCTGGGACCCAGTTCACTCTCGCCATTAACGGCGTGGACTGTGACGATGCT
GCCACTTACTACTGTCAAGGCACTTATTATACTAGTGATTGGTACTTGGCTTT
CGGCGGAGGGACCGAGGTGGTGGTCAAAGGTGATCCAGTTGCACCTACTGTC
CTCATCTTCCCACCAGCTGCTGATCAGGTGGCAACTGGAACAGTCACCATCGT
GTGTGTGGCGAATAAATACTTTCCCGATGTCACCGTCACCTGGGAGGTGGAT
GGCACCACCCAAACAACTGGCATCA

>55-3_Kappa_translated
APELDMTQTPPSVSAAVGDTVTIKCQSSQSVGDVNELSWYQQKPGQPPKLLVYR
ASNLASGIPSRFKGSGSGTQFTLAINGVDCDDAATYYCQGTYYTSDWYLAFGGG

TEVVVKGDPVAPTVLIFPPAADQVATGTVTIVCVANKYFPDVTVTWEVDGTTQT
TGI

>55-3_Heavy_DNA
TGGTCGCTGTGCTCAAAGGTGTCCaGTGTCAGCAGCTGGAGCAGTCCGGAGG
AGGAGCCGAAGGAGGCCTGGTCAAGCCTGGGGGATCCCTGGAACTCTGCTGC
AAAGCCTCTGGATTCATCATGAGTAGTACATACTGGATGTGTTGGGTCCGCCA
GGCTCCAGGGAAGGGGCTGGAGTGGATCGGATGTACTGCTGTTCGTAGTGGT
GGTAGGACCACCTACGCGAGCTGGGTGAATGGCCGGTTCACTCTCTCCAGAG
ACGTCGACCAGAGCACAGGCTGCCTACAACTGAACAGTCTGACAGTCGCGGA
CACGGCCATATATTTTTGTGGGAGAGATCGATCTACTGGTACCATGGATATCC
TGGACTTGTGGGGCCAAGGCACCCTGGTTACCGTCTCCTCAGGGCAACCTAA
GGCTCCATCAGTCTTCCCACTGGCCCCCTGCT

> 55-3_Heavy_translated
VAVLKGVQCQQLEQSGGGAEGGLVKPGGSLELCCKASGFIMSSTYWMCWVRQ
APGKGLEWIGCTAVRSGGRTTYASWVNGRFTLSRDVDQSTGCLQLNSLTVADT
AIYFCGRDRSTGTMDILDLWGQGTLVTVSSGQPKAPSVFPLAPC

>93-4_Kappa_DNA
ATGACCCAGACTCCATCGCCCGTGTCTGCAGCTGTGGGAGGCACAGTCACCA
TCAAGTGCCAGGCCAGTCAGAGTGTTTATAATAACAACCTCTTATCCTGGTTT
CAGCAGAAACCAGGGCAGCCTCCCAAGCGCCTGATCTATTCTACATCCACTCT
GGCATCTGGGGTCCCATCGCGGTTCAAAGGCAGTGGATCTGGGACACACTTC
ACTCTCACCATCAGCGACGTGCAGTGTGACGATGCTGCCACTTATTACTGTCT
AGGCAGTTATGATGGGAATAGCGCTGATTGCCTTGCTTTCGGCGGAGGGACC
GAGGTGGTGGTCAAAGGTGATCCAGTTGCACCTACTGTCCTCATCTTCCCACC
AGCTGCTGATCAGGTGGCAACTGGAACAGTCACCATCGTGTGTGTGGCGAAT
AAATACTTTCCCGATGTCACCGTCACCTGGGAGGTGGATGGCACCACCC

>93-4_Kappa_translated
MTQTPSPVSAAVGGTVTIKCQASQSVYNNNLLSWFQQKPGQPPKRLIYSTSTLAS
GVPSRFKGSGSGTHFTLTISDVQCDDAATYYCLGSYDGNSADCLAFGGGTEVVV
KGDPVAPTVLIFPPAADQVATGTVTIVCVANKYFPDVTVTWEVDGTT

>93-4_Heavy_DNA
GAGACTGGGCTGCGCTGGCTTCTCCTGGTCGCTGTGCTCAAAGGTGTCCAGTG
TCAGTCGCTGGAGGAGTCCGGGGGTCGCCTGGTCACGCCTGGGACACCCCTG
ACACTCACCTGCACAGCCTCTGGAATCGACCTCAATAGCGACCACATGAGCT
GGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAATGGATCGCCATCATTTATGC
TAGTGGTACCATATACTACGCGAGCTGGGCGAAAGGCCGCTTCACCATCTCC
AAAACCGCGACCACGGTGGATCTGAGGATCACCAGTCCGACGACCGAGGAC
ACGGCCACCTATTTCTGTGCCACTTACCCTAATTATCCTACTGATAACTTGTG
GGGCCAGGGCACCCTGGTCACCGTCTCCTCAGGGCAACCTAAGGCTCCATCA
GTCTTCCCACTGGCCCCCTGCTGCGGGGACAA

>93-4_Heavy_translated
ETGLRWLLLVAVLKGVQCQSLEESGGRLVTPGTPLTLTCTASGIDLNSDHMSWV
RQAPGKGLEWIAIIYASGTIYYASWAKGRFTISKTATTVDLRITSPTTEDTATYFC
ATYPNYPTDNLWGQGTLVTVSSGQPKAPSVFPLAPCCGD

10.7 List of RabFab constructs

Format of sequences:

*Age*I site – Variable region – *Kpn*I site – Constant region – (His-tag) – Termination codon – *Xho*I site

>55-3_Kappa_RabFab
ACCGGTCTGGACATGACCCAGACCCCCCCTTCTGTGTCTGCCGCCGTGGGCGA
TACCGTGACCATCAAGTGTCAGAGCAGCCAGAGCGTGGGCGACGTGAACGAG
CTGAGCTGGTATCAGCAGAAGCCCGGCCAGCCCCCTAAGCTGCTGGTGTACA
GAGCCAGCAATCTGGCCAGCGGCATCCCCAGCAGATTCAAGGGCTCTGGCTC
CGGCACCCAGTTCACCCTGGCTATCAATGGCGTGGACTGCGACGACGCCGCC
ACCTACTATTGTCAGGGCACCTACTACACCAGCGACTGGTATCTGGCCTTCGG
CGGAGGCACAGAGGTGGTCGTGAAGGGTACCCCTGTGGCCCCCACCGTGCTG
ATCTTTCCACCAGCCGCTGACCAGGTGGCCACAGGCACAGTGACCATCGTGT
GCGTGGCCAACAAGTACTTCCCCGACGTGACCGTGACCTGGGAGGTGGACGG
CACCACACAGACCACCGGCATCGAGAACAGCAAGACCCCCCAGAACAGCGC
CGACTGCACCTACAACCTGAGCAGCACCCTGACCCTGACCTCCACCCAGTAC
AACAGCCACAAAGAGTACACCTGTAAAGTGACCCAGGGCACCACCAGCGTG
GTGCAGAGCTTCAACAGAGGCGACTGCAGCTAATGATCACTCGAG

>55-3_Heavy_RabFab
ACCGGTTCTGGCGGCGGAGCTGAAGGCGGACTCGTGAAACCTGGCGGCAGCC
TGGAACTGTGCTGCAAGGCCAGCGGCTTCATCATGAGCAGCACCTACTGGAT
GTGCTGGGTGCGCCAGGCCCCTGGCAAAGGCCTGGAATGGATCGGATGTACC
GCCGTGCGGAGCGGCGGCAGAACAACATATGCCAGCTGGGTCAACGGCCGGT
TCACCCTGAGCAGAGATGTGGACCAGAGCACCGGCTGCCTGCAGCTGAATAG
CCTGACCGTGGCCGACACCGCCATCTACTTTTGCGGCAGAGACAGATCCACC
GGCACCATGGACATCCTGGACCTGTGGGGCCAGGGCACACTCGTGACAGTGT
CTGGTACCCAGCCCAAGGCCCCCAGCGTGTTCCCTCTGGCTCCTTGCTGTGGC
GACACCCCTAGCAGCACCGTGACACTGGGCTGCCTCGTGAAGGGCTACCTGC
CTGAGCCTGTGACCGTGACCTGGAACAGCGGCACCCTGACAAATGGCGTGCG
GACCTTTCCTAGCGTGCGGCAGTCTAGCGGCCTGTACAGCCTGAGCAGCGTG
GTGTCTGTGACCAGCAGCAGCCAGCCCGTGACATGCAATGTGGCCCACCCCG
CCACCAACACCAAGGTGGACAAAACCGTGGCCCCCTCCACCTGTAGCAAGCA
CCACCATCACCATCACTAATGATCACTCGAG

>49-6_Kappa_RabFab
ACCGGTATGACCCAGACTCCATCCCCCGTGTCTGCAGCTGTGGGAGGCACAG
TCACCATCAACTGCCAGGCCAGTCAGAGTGTTTATAATAACTACCTCTTATCC
TGGTATCAGCAAAAACCAGGGCAGCCTCCCAAGCGCCTGATCTATTCTGCAT
CCACTCTGGCATCTGGGGTCTCATCGCGGTTCAAAGGCAGTGGATCTGGGAC
ACAGTTCACTCTCACCATCAGCGACGTGCAGTGTGACGATGCTGCCACTTATT
ATTGTCTAGGCAGTTATGACGGTAATAGTGCTGATTGCCTTGCTTTCGGCGGA
GGGACCGAGGTGGTGGTCAAAGGTACCCCTGTGGCCCCCACCGTGCTGATCT
TTCCACCAGCCGCTGACCAGGTGGCCACAGGCACAGTGACCATCGTGTGCGT
GGCCAACAAGTACTTCCCCGACGTGACCGTGACCTGGGAGGTGGACGGCACC
ACACAGACCACCGGCATCGAGAACAGCAAGACCCCCCAGAACAGCGCCGAC
TGCACCTACAACCTGAGCAGCACCCTGACCCTGACCTCCACCCAGTACAACA

GCCACAAAGAGTACACCTGTAAAGTGACCCAGGGCACCACCAGCGTGGTGCA
GAGCTTCAACAGAGGCGACTGCAGCTAATGATCACTCGAG

>49-6_Heavy_RabFab
ACCGGTTCGCTGGAGGAGTCCGGGGGTCGCCTGGTCACGCCTGGGACACCCC
TGACACTCACCTGCACAGTCTCTGGAATCGACCCCAATAGCGACCACATGAG
CTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAATGGATCGCAATTATTTAT
GCTAGTGGTACCACTTACTACGCGAGCTGGGCAAAAGGCCGATTCACCATCT
CCAAAACCTCGACCACGGTGGATCTGAGGATCGCCAGTCCGACAACCGAGGA
CACGGCCACCTATTTCTGTGCCACTTATCCTAATTATCCTACTGATAACTTGTG
GGGCCAAGGCACCCTGGTCACCGTCTCCGGTACCCAGCCCAAGGCCCCCAGC
GTGTTCCCTCTGGCTCCTTGCTGTGGCGACACCCCTAGCAGCACCGTGACACT
GGGCTGCCTCGTGAAGGGCTACCTGCCTGAGCCTGTGACCGTGACCTGGAAC
AGCGGCACCCTGACAAATGGCGTGCGGACCTTTCCTAGCGTGCGGCAGTCTA
GCGGCCTGTACAGCCTGAGCAGCGTGGTGTCTGTGACCAGCAGCAGCCAGCC
CGTGACATGCAATGTGGCCCACCCCGCCACCAACACCAAGGTGGACAAAACC
GTGGCCCCCTCCACCTGTAGCAAGCACCACCATCACCATCACTAATGATCACT
CGAG

>93_Kappa_RabFab
ACCGGTATGACCCAGACTCCATCGCCCGTGTCTGCAGCTGTGGGAGGCACAG
TCACCATCAAGTGCCAGGCCAGTCAGAGTGTTTATAATAACAACCTCTTATCC
TGGTTTCAGCAGAAACCAGGGCAGCCTCCCAAGCGCCTGATCTATTCTACATC
CACTCTGGCATCTGGGGTCCCATCGCGGTTCAAAGGCAGTGGATCTGGGACA
CACTTCACTCTCACCATCAGCGACGTGCAGTGTGACGATGCTGCCACTTATTA
CTGTCTAGGCAGTTATGATGGGAATAGCGCTGATTGCCTTGCTTTCGGCGGAG
GGACCGAGGTGGTGGTCAAAGGTACCCCTGTGGCCCCCACCGTGCTGATCTTT
CCACCAGCCGCTGACCAGGTGGCCACAGGCACAGTGACCATCGTGTGCGTGG
CCAACAAGTACTTCCCCGACGTGACCGTGACCTGGGAGGTGGACGGCACCAC
ACAGACCACCGGCATCGAGAACAGCAAGACCCCCCAGAACAGCGCCGACTG
CACCTACAACCTGAGCAGCACCCTGACCCTGACCTCCACCCAGTACAACAGC
CACAAAGAGTACACCTGTAAAGTGACCCAGGGCACCACCAGCGTGGTGCAGA
GCTTCAACAGAGGCGACTGCAGCTAATGATCACTCGAG

>93-4_Heavy_RabFab
ACCGGTTCGCTGGAGGAGTCCGGGGGTCGCCTGGTCACGCCTGGGACACCCC
TGACACTCACCTGCACAGCCTCTGGAATCGACCTCAATAGCGACCACATGAG
CTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAATGGATCGCCATCATTTAT
GCTAGTGGTACCATATACTACGCGAGCTGGGCGAAAGGCCGCTTCACCATCT
CCAAAACCGCGACCACGGTGGATCTGAGGATCACCAGTCCGACGACCGAGGA
CACGGCCACCTATTTCTGTGCCACTTACCCTAATTATCCTACTGATAACTTGTG
GGGCCAGGGCACCCTGGTCACCGTCTCCGGTACCCAGCCCAAGGCCCCCAGC
GTGTTCCCTCTGGCTCCTTGCTGTGGCGACACCCCTAGCAGCACCGTGACACT
GGGCTGCCTCGTGAAGGGCTACCTGCCTGAGCCTGTGACCGTGACCTGGAAC
AGCGGCACCCTGACAAATGGCGTGCGGACCTTTCCTAGCGTGCGGCAGTCTA
GCGGCCTGTACAGCCTGAGCAGCGTGGTGTCTGTGACCAGCAGCAGCCAGCC
CGTGACATGCAATGTGGCCCACCCCGCCACCAACACCAAGGTGGACAAAACC
GTGGCCCCCTCCACCTGTAGCAAGCACCACCATCACCATCACTAATGATCACT
CGAG

10.8 DC-SIGN construct and primers

Human DC-SIGN with 4.5 neck repeats. Length: G158-A404

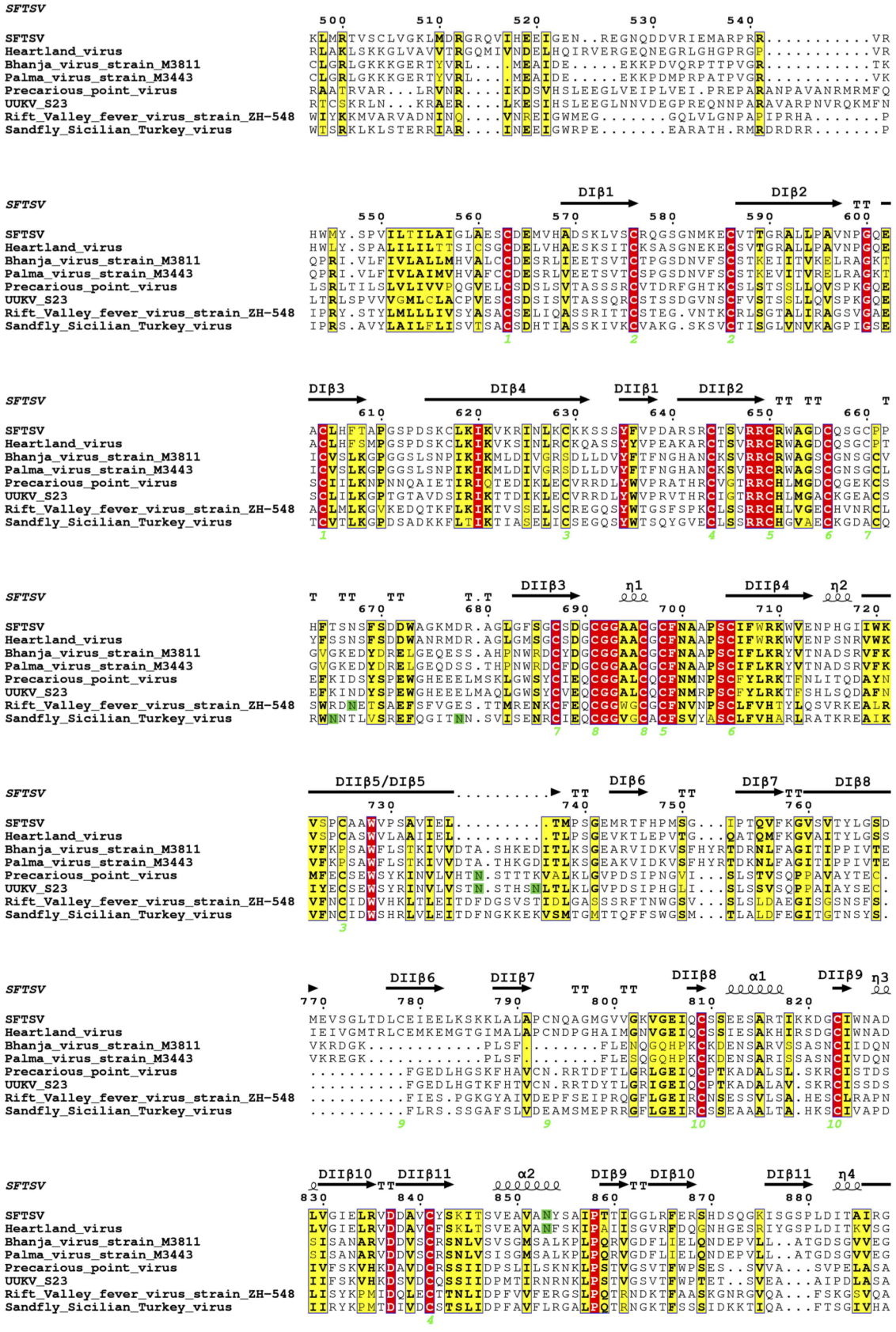
>DC_SIGN_4-5rep
GELPEKSKMQEIYQELTRLKAAVGELPEKSKQQEIYQELTRLKAAVGELPEKSKQ
QEIYQELTRLKAAVGELPEKSKQQEIYQELTQLKAAVERLCHPCPWEWTFFQGN
CYFMSNSQRNWHDSITACKEVGAQLVVIKSAEEQNFLQLQSSRSNRFTWMGLSD
LNQEGTWQWVDGSPLLPSFKQYWNRGEPNNVGEEDCAEFSGNGWNDDKCNLA
KFWICKKSAASCSRDEEQFLSPAPATPNPPPA

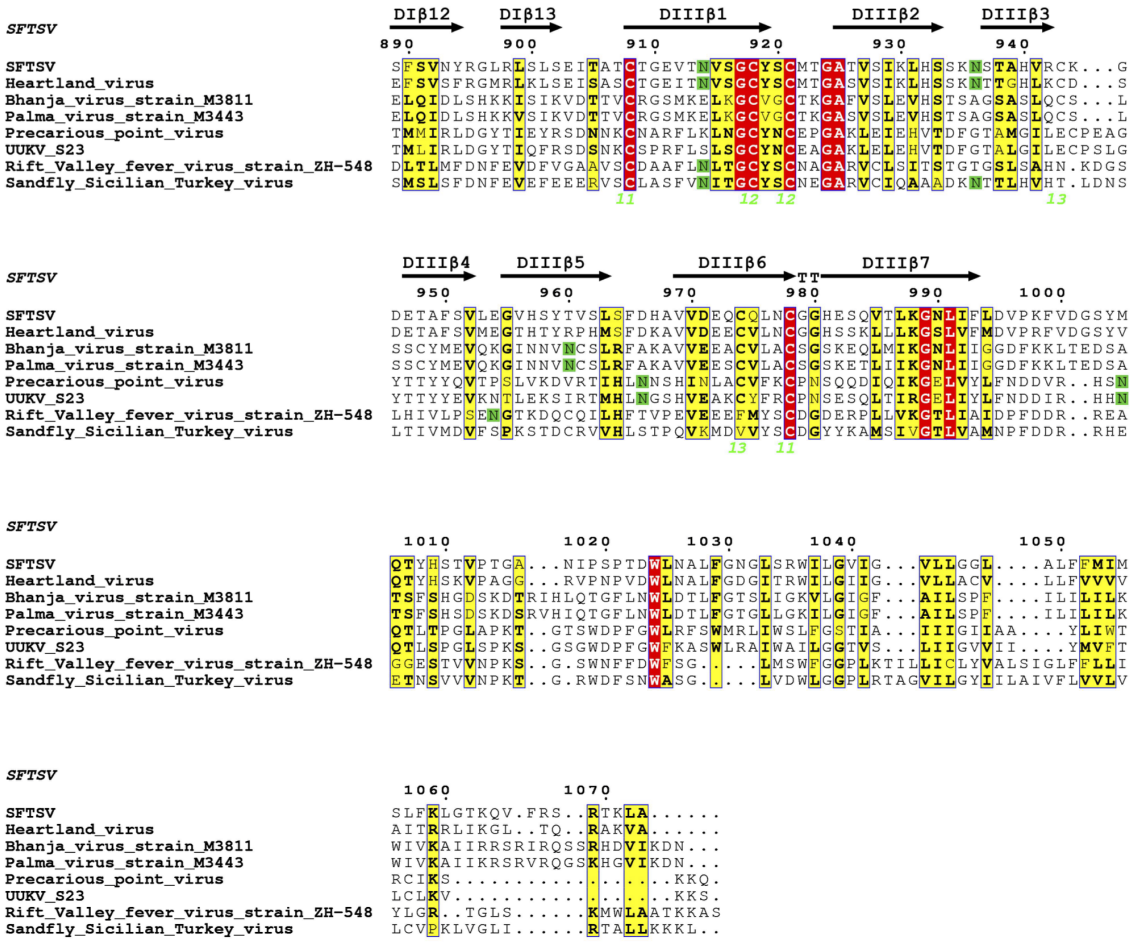
Primers for pHLSec cloning:

Forward: 5'-cgcaccggtGGGGAACTGCCTGAAAAGTCAAAAATGC-3'

Reverse: 5'-gcggtaccAGCTGGTGGAGGATTGGGGGTG-3'

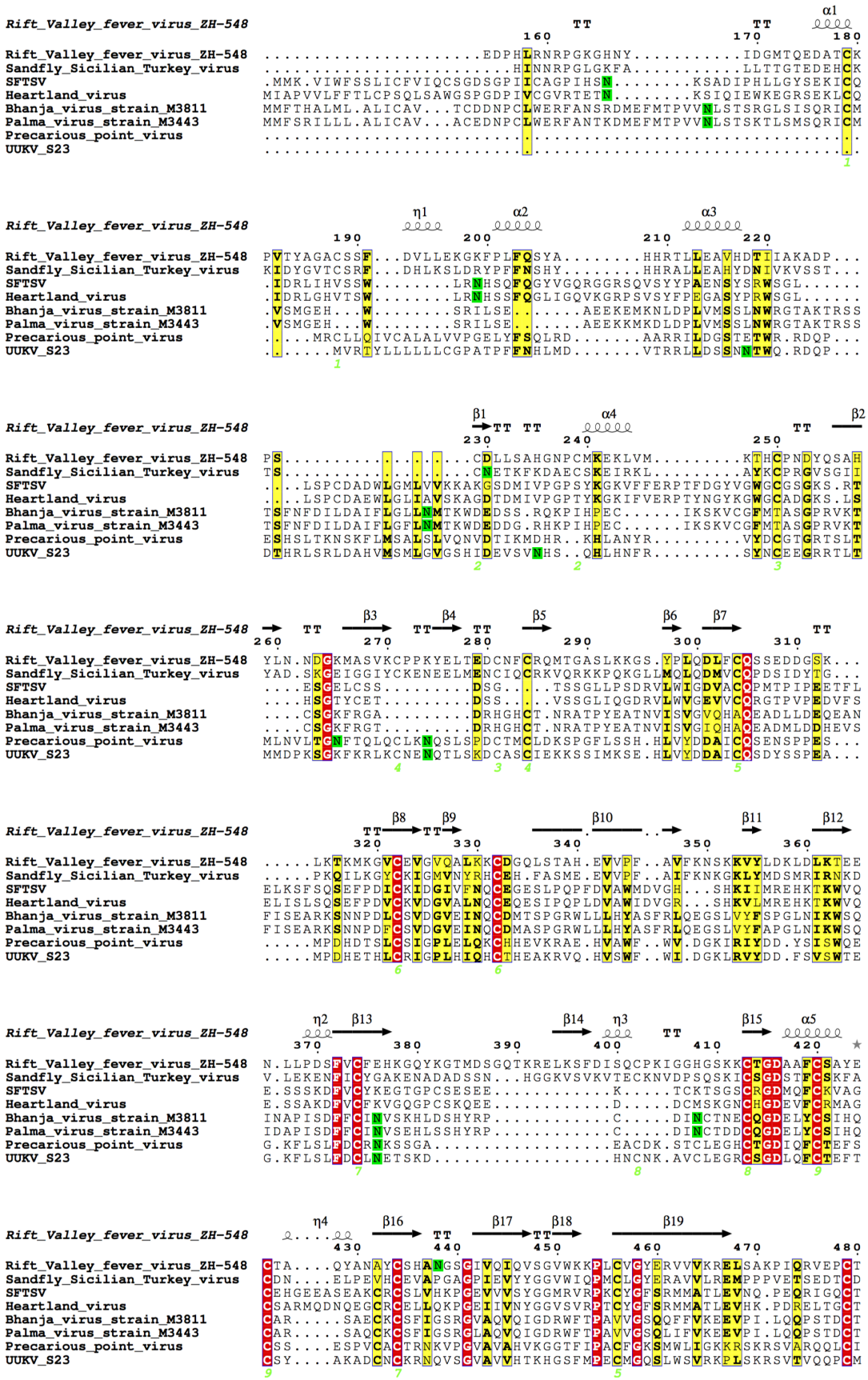
10.9 Sequence alignment of representative Gc phleboviral sequences

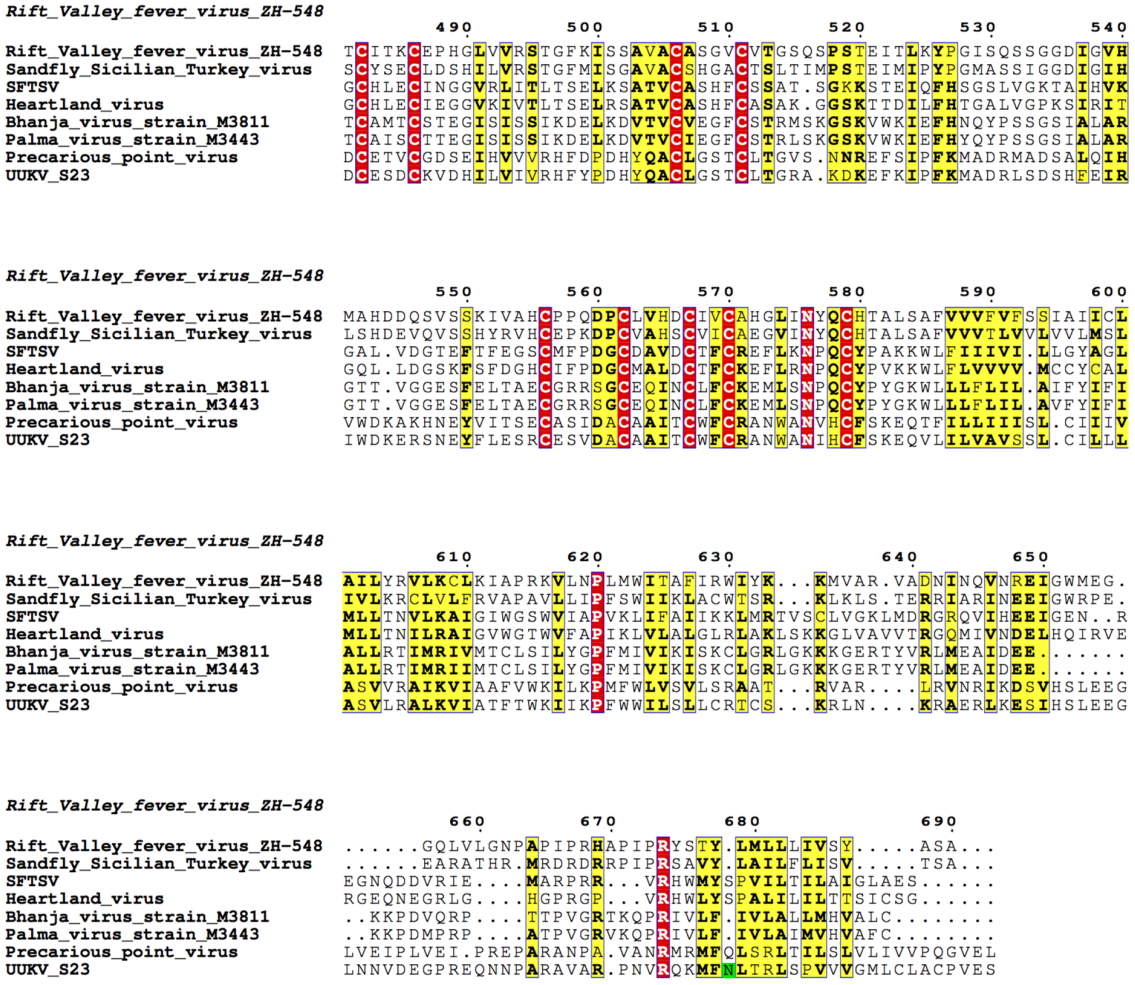




A list of representative phlebovirus sequences was based on previously published work [184]. The sequences were obtained from Genbank (accession numbers are given in parenthesis): Bhanja virus strain M3811 (JQ956377), Heartland virus (JX005845), Palma virus strain M3443 (JQ956380), Precarious Point virus (HM566179), Rift Valley fever virus strain Smithburn (DQ380193.1), Sandfly fever Turkey virus (NC_015411), SFTSV strain HN6 (HQ141596) and Uukuniemi virus (UUKGPM). The sequences were aligned using the online Clustal Omega [185] server available at the EMBL-EBI (http://www.ebi.ac.uk/Tools/msa/clustalo/). The sequence alignment was rendered using the online server ESPript 3.0 [186] (http://espript.ibcp.fr).

10.10 Sequence alignment of representative Gn phleboviral sequences





Alignment was made as described in Appendix **10.9**

10.11 Fourier Shell Correlations for cryo-EM maps

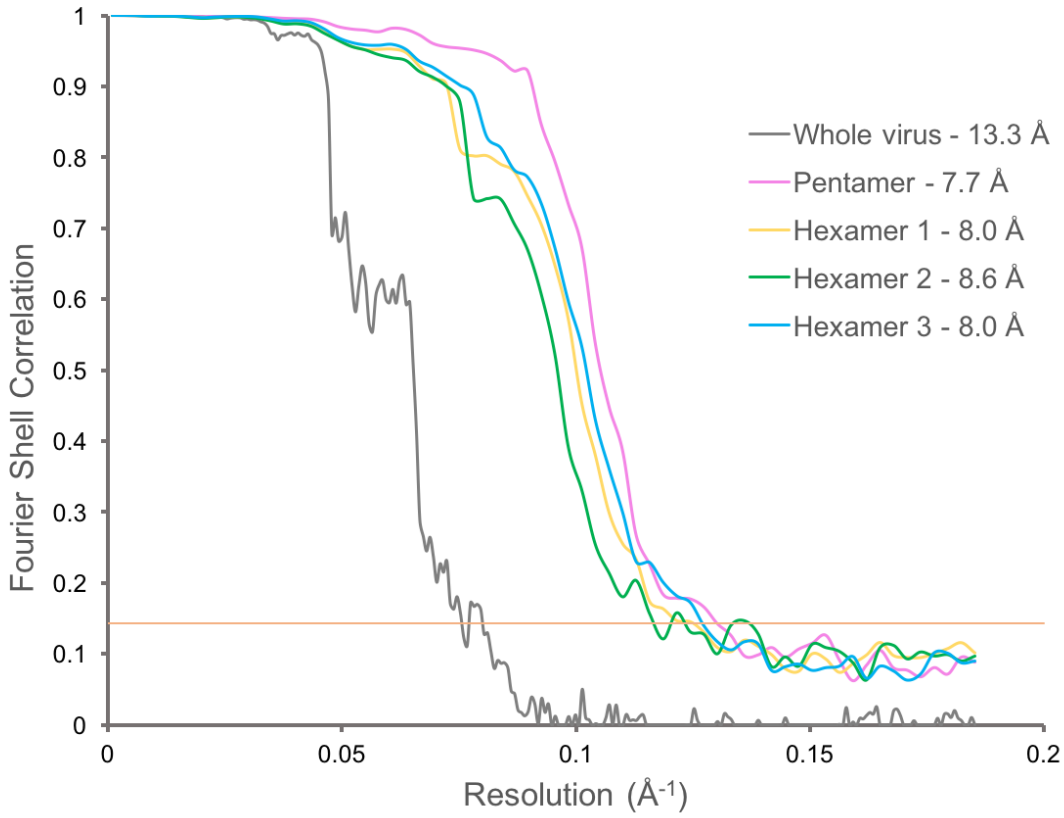
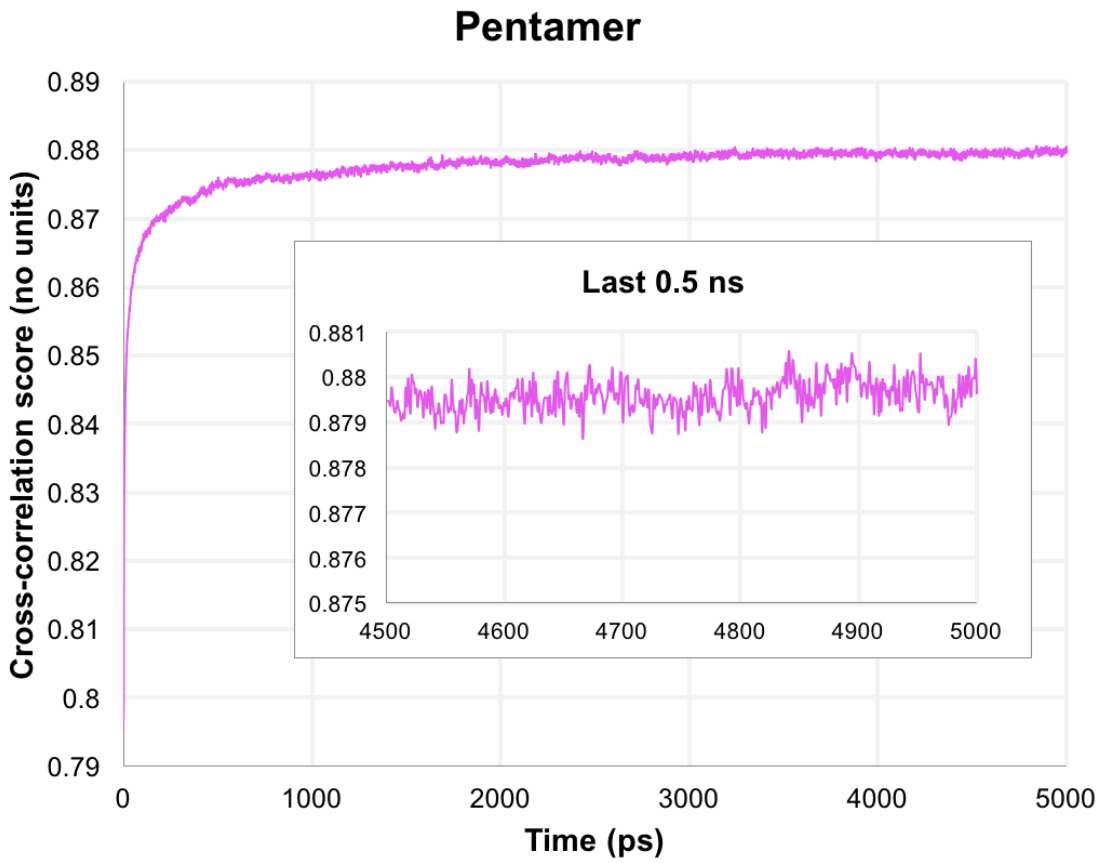
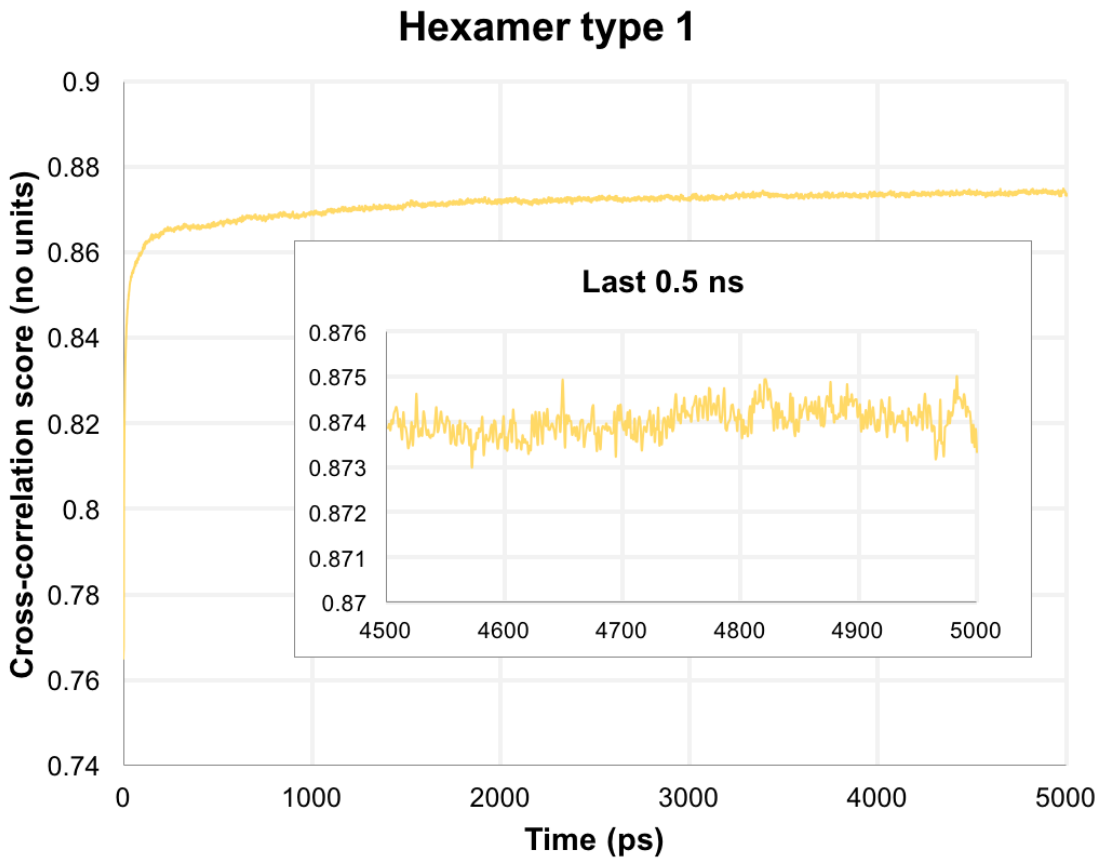
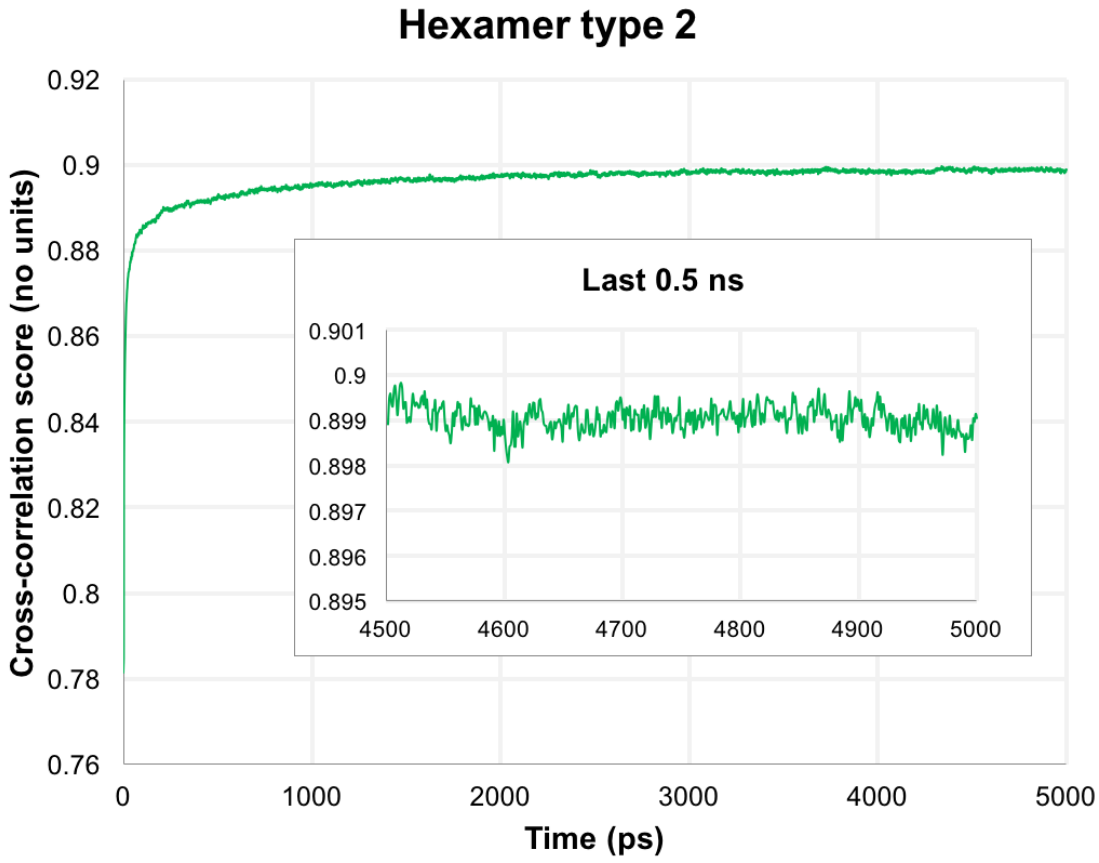


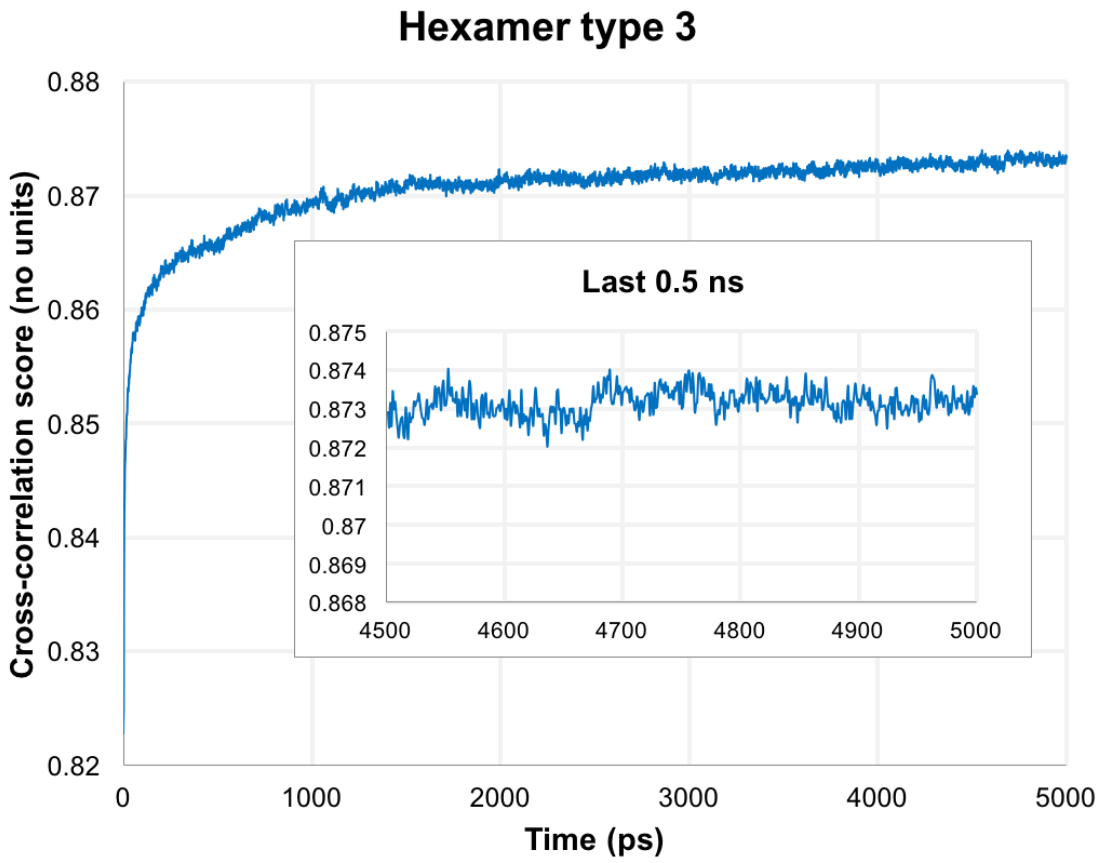
Figure 10.1 Resolution estimation of whole RVFV and sub-particles.

10.12 Cross-correlation scores for MDFF simulations




Hexamer type 1
Cross-correlation score (no units)
0.9
0.88
0.86
0.84
0.82
0.8
0.78
0.76
0.74
0
1000
2000
3000
4000
5000
Time (ps)
Last 0.5 ns
0.876
0.875
0.874
0.873
0.872
0.871
0.87
4500
4600
4700
4800
4900
5000


Hexamer type 2
Cross-correlation score (no units)
0.92
0.9
0.88
0.86
0.84
0.82
0.8
0.78
0.76
0
1000
2000
3000
4000
5000
Time (ps)
Last 0.5 ns
0.901
0.9
0.899
0.898
0.897
0.896
0.895
4500
4600
4700
4800
4900
5000


Hexamer type 3
Cross-correlation score (no units)
0.88
0.87
0.86
0.85
0.84
0.83
0.82
0
1000
2000
3000
4000
5000
Time (ps)
Last 0.5 ns
0.875
0.874
0.873
0.872
0.871
0.87
0.869
0.868
4500
4600
4700
4800
4900
5000

10.13 RMSD matrix from all-against-all alignment of heterodimers

	Hex1A	Hex1B	Hex1C	Hex1D	Hex1E	Hex1F	Hex2A	Hex2B	Hex2C	Hex2D	Hex2E	Hex2F	Hex3A	Hex3B	Hex3C	Hex3D	Hex3E	Hex3F	PenA	PenB	PenC	PenD	PenE
Hex1A	0	1.742	2.116	1.112	1.226	1.454	1.195	2.165	2.046	1.182	2.349	2.077	1.571	1.34	1.402	1.427	1.352	1.373	2.04	2.01	1.947	1.991	1.948
Hex1B	1.742	0	1.496	1.545	1.409	1.7	1.647	1.497	1.767	1.433	1.585	1.921	1.562	1.341	1.653	1.432	1.543	1.285	1.753	1.738	1.603	1.674	1.847
Hex1C	2.116	1.496	0	1.373	1.653	1.7	2.228	1.663	1.728	1.3	1.74	1.746	1.558	1.778	1.719	1.691	1.657	1.566	1.89	1.843	1.785	1.844	1.922
Hex1D	1.112	1.545	1.373	0	1.14	1.226	1.01	0.858	1.261	1.071	0.979	1.257	1.388	1.434	1.402	1.447	1.489	1.481	1.024	1.006	1.02	1.025	1.042
Hex1E	1.226	1.409	1.653	1.14	0	1.33	1.278	2.026	1.837	1.292	2.164	1.811	1.166	1.243	1.228	1.348	1.123	1.273	2.153	2.07	2.06	2.034	2.109
Hex1F	1.454	1.7	1.7	1.226	1.33	0	1.294	1.786	2.142	1.191	1.876	1.963	1.457	1.236	1.522	1.259	1.337	1.141	1.893	1.752	1.764	1.738	1.837
Hex2A	1.195	1.647	2.228	1.01	1.278	1.294	0	2.368	2.321	0.845	2.374	2.329	1.645	1.216	1.62	1.274	1.447	1.229	2.278	2.258	2.281	2.204	2.242
Hex2B	2.165	1.497	1.663	0.858	2.026	1.786	2.368	0	1.727	0.994	0.845	1.926	1.604	1.941	1.586	1.976	1.611	1.919	2.043	1.948	2.01	1.99	2.1
Hex2C	2.046	1.767	1.728	1.261	1.837	2.142	2.321	1.727	0	1.321	1.738	0.75	1.688	1.932	1.703	2.007	1.844	1.861	2.209	2.005	2.034	2.084	2.145
Hex2D	1.182	1.433	1.3	1.071	1.292	1.191	0.845	0.994	1.321	0	1.216	1.289	1.166	1.23	1.214	1.215	1.153	1.084	0.889	0.901	0.901	0.893	0.969
Hex2E	2.349	1.585	1.74	0.979	2.164	1.876	2.374	0.845	1.738	1.216	0	1.902	1.649	1.987	1.633	1.958	1.711	1.928	1.932	1.874	1.888	1.862	2.062
Hex2F	2.077	1.921	1.746	1.257	1.811	1.963	2.329	1.926	0.75	1.289	1.902	0	1.643	1.962	1.667	2.031	1.797	1.964	2.2	2.025	2.024	2.055	2.102
Hex3A	1.571	1.562	1.558	1.388	1.166	1.457	1.645	1.604	1.688	1.166	1.649	1.643	0	1.404	0.799	1.673	0.771	1.499	1.881	1.859	1.786	1.77	1.886
Hex3B	1.34	1.341	1.778	1.434	1.243	1.236	1.216	1.941	1.932	1.23	1.987	1.962	1.404	0	1.424	0.735	1.283	0.716	2.017	1.933	1.801	1.804	1.909
Hex3C	1.402	1.653	1.719	1.402	1.228	1.522	1.62	1.586	1.703	1.214	1.633	1.667	0.799	1.424	0	1.601	0.76	1.44	1.915	1.905	1.843	1.838	1.944
Hex3D	1.427	1.432	1.691	1.447	1.348	1.259	1.274	1.976	2.007	1.215	1.958	2.031	1.673	0.735	1.601	0	1.549	0.823	1.863	1.773	1.602	1.652	1.738
Hex3E	1.352	1.543	1.657	1.489	1.123	1.337	1.447	1.611	1.844	1.153	1.711	1.797	0.771	1.283	0.76	1.549	0	1.321	1.938	1.885	1.849	1.816	1.901
Hex3F	1.373	1.285	1.566	1.481	1.273	1.141	1.229	1.919	1.861	1.084	1.928	1.964	1.499	0.716	1.44	0.823	1.321	0	1.92	1.864	1.728	1.758	1.844
PenA	2.04	1.753	1.89	1.024	2.153	1.893	2.278	2.043	2.209	0.889	1.932	2.2	1.881	2.017	1.915	1.863	1.938	1.92	0	0.609	0.601	0.576	0.637
PenB	2.01	1.738	1.843	1.006	2.07	1.752	2.258	1.948	2.005	0.901	1.874	2.025	1.859	1.933	1.905	1.773	1.885	1.864	0.609	0	0.588	0.578	0.627
PenC	1.947	1.603	1.785	1.02	2.06	1.764	2.281	2.01	2.034	0.901	1.888	2.024	1.786	1.801	1.843	1.602	1.849	1.728	0.601	0.588	0	0.546	0.604
PenD	1.991	1.674	1.844	1.025	2.034	1.738	2.204	1.99	2.084	0.893	1.862	2.055	1.77	1.804	1.838	1.652	1.816	1.758	0.576	0.578	0.546	0	0.596
PenE	1.948	1.847	1.922	1.042	2.109	1.837	2.242	2.1	2.145	0.969	2.062	2.102	1.886	1.909	1.944	1.738	1.901	1.844	0.637	0.627	0.604	0.596	0

10.14 Additional crystallographic and cryo-EM tables

Table 10.1 Crystallographic data and refinement statistics for the RVFV-Gn-RabFab49-6 complex.

Parameters	Data
Data collection	
Beamline	I03, DLS
Resolution range (Å)	60.78–2.05 (2.09–2.05)
Space group	*P*1
Cell dimensions	
a, b, c (Å)	51.46, 61.64, 78.64
α, β, γ (°)	81.36, 78.09, 84.11
Wavelength (Å)	0.97622
Unique reflections	56,934 (2,660)
Completeness (%)	97.38 (92.72)
R_{merge} (%)[a]	6.96 (84.59)
I/σI	9.64 (1.14)
CC-half (%)	99.78 (48.65)
Avg. redundancy	3.41 (2.94)
Refinement	
Resolution range	60.77–2.05 (2.10–2.05)
Number of reflections	54,160 (2,050)
R_{work} (%)[b]	24.92
R_{free} (%)[c]	30.84
RMSD	
Bonds (Å)	0.015
Angles (°)	1.78
Molecules per a.s.u.	6
Atoms per a.s.u (protein)	6,589
Average *B*-factors (Å^2)	47.26
Ramachandran plot (%)	
Most favoured region	92.34
Allowed region	5.60
Outliers	2.06

Numbers in parentheses refer to the relevant outer resolution shell.
RMSD: root mean square deviation from ideal geometry.
[a]$R_{merge} = \Sigma_{hkl} \Sigma_i |I(hkl;i) - \langle I(hkl) \rangle| / \Sigma_{hkl} \Sigma_i I(hkl;i)$, where $I(hkl;i)$ is the intensity of an individual measurement and $\langle I(hkl) \rangle$ is the average intensity from multiple observations.
[b]$R_{factor} = \Sigma_{hkl} ||F_{obs}| - k|F_{calc}|| / \Sigma_{hkl} |F_{obs}|$
[c]R_{free} equals the *R*–factor as calculated above but using against 5% of the data removed prior to refinement.

Table 10.2 Single-particle cryo-EM and localized reconstruction parameters for RVFV-DC-SIGN and RVFV-RabFab93-4 complexes

Parameters	RVFV DC-SIGN	Localized DC-SIGN	RVFV Fab93-4	Localized Fab93-4
Data acquisition				
Frames per movie	88	N/A	88	N/A
Exposure per frame (s)	0.2	N/A	0.2	N/A
Dose rate (e^-/ pix / s)	2.5	N/A	2.5	N/A
Total dose (e^- / $Å^2$)	22	N/A	22	N/A
Defocus* (μm)	1.0–3.0	N/A	1.0–3.0	N/A
Data processing				
Micrographs	291	N/A	963	N/A
Particles**	343 (685)	8,220 (8,220)	2,770 (3,065)	33,142 (33,240)
Box size (pixels)	512	128	512	128
Symmetry	I1	C5	I1	C5
Pixel size (Å)	2.7	2.7	2.7	2.7
Resolution (Å)***	28.2	11.9	20.3	8.7
B-factor applied	-800	–300	–800	–300

* Positive value denotes underfocus.
** Number of particles used for the reconstruction (numbers of all extracted particles are parenthesis).
*** Resolution (Fourier shell correlation = 0.143).

10.15 List of publications and manuscripts in preparation

Halldorsson S, Li S, Li M, Harlos K, Bowden TA, Huiskonen JT. (2018) Shielding and activation of a membrane fusion protein. *Accepted in Nature Communications*

Halldorsson S, Behrens AJ, Harlos K, Huiskonen JT, Elliott RM, Crispin M, Brennan B, Bowden TA (2016) Structure of a phleboviral envelope glycoprotein reveals a consolidated model of membrane fusion. *PNAS.* 113(8):7154-9.

Bitto D, **Halldorsson S**, Caputo A, Huiskonen JT. (2016) Low pH and Anionic Lipid-dependent Fusion of Uukuniemi Phlebovirus to Liposomes. *J. Biol. Chem.* 291(12):6412-22.

Bitto D, Harvey DJ, **Halldorsson S**, Doores KJ, Pritchard LK, Huiskonen JT, Bowden TA, Crispin M. (2015) Determination of N-linked Glycosylation in Viral Glycoproteins by Negative Ion Mass Spectrometry and Ion Mobility. *Methods Mol. Biol.* 1331:93-121.

Crispin M, Harvey DJ, Bitto D, **Halldorsson S**, Bonomelli C, Edgeworth M, Scrivens JH, Huiskonen JT, Bowden TA (2014) Uukuniemi Phlebovirus assembly and secretion leave a functional imprint on the virion glycome. *J. Virol*. 88(17):10244-51.