STRUCTURE AND ENTRY OF HAEMORRHAGIC FEVER-CAUSING ARENAVIRUSES

Marie-Laure Parsy Szikola
Wellcome Trust programme in Structural Biology

UNIVERSITY OF OXFORD

Supervised by: **Dr. Juha Huiskonen**
Oxford Particle Imaging Centre
Division of Structural Biology
Wellcome Trust Centre for Human Genetics
University of Oxford
Roosevelt Drive, Oxford OX3 7BN

Introduction

1. *Viral haemorrhagic fevers: emerging threats and global challenge*

1.1. General Aspects of VHFs

Viral haemorrhagic fevers (VHFs) are defined as a group of systemic emerging viral diseases often associated with high-mortality rates. They occur predominantly in underdeveloped parts of the world where they appear as sporadic outbreaks of potentially fatal illnesses characterized by fever, bleeding diathesis (an abnormal propensity towards bleeding), organ failure and circulatory shock (Geisbert & Jahrling 2004).

VHFs agents belong to four different families of single-stranded enveloped RNA viruses whose distribution is defined by the ecology of their vectors or reservoirs:

- the *Filoviridae*, which contain the well-studied Ebola virus, are endemic in Africa.

- the *Bunyaviridae*, which can be transmitted either by mosquito (Rift Valley fever virus: RVFV) or by ticks (Crimean-Congo haemorrhagic fever virus: CCHFV), are respectively located in Africa, Yemen and Saudi Arabia; and in Africa, Central Asia, Europe and throughout the Middle East.

- the *Flaviviridae* (Yellow fever virus), also transmitted by mosquitoes, are found in Africa and in the tropical Americas

- the family of *Arenaviridae*, whose natural reservoirs are rodents, has a global distribution with a presence in North and South America, Africa, Asia and Europe.

VHFs present an incubation period that ranges from 2 to 21 days and a subsequent onset of varied, unspecific clinical symptoms that renders early diagnosis difficult. A flu-like

illness with fever and myalgia is typically observed during the early stages of the disease. This flu-like illness may develop into a life-threatening shock-like syndrome depending on agent virulence, mode of transmission, initial dose and host factors (Table 01) (Geisbert & Jahrling 2004).

	Clinical description of acquired disease	Incidence of disease
Filoviridae		
Ebola virus* (Central and West Africa)	Sudden fever, intense weakness, muscle pain, headache, sore throat, vomiting, diarrhoea, rash, impaired kidney and liver functions, in some cases internal and external bleeding.	Fatality rate up to 90%.
Marburg virus* (Central and South Africa)	Abrupt high fever, severe headache and malaise, muscle aches, severe watery diarrhoea, abdominal pain and cramping, nausea and vomiting, mucosal bleeding, confusion, irritability, aggression.	Fatality rates in outbreaks range from 24% to 88%.
Bunyaviridae		
Rift Valley fever virus* (Sub-Saharan and North Africa, Saudi Arabia, Yemen)	Flu-like fever, muscle pain, joint pain, headache, sometimes neck stiffness, sensitivity to light, loss of appetite and vomiting. In severe RVF: Ocular form: retinal lesions; Meningoencephalitis form: loss of memory, hallucinations, confusion, disorientation, vertigo, convulsions, lethargy, coma; Haemorrhagic form: liver impairment, haemorrhage with purpuric rash, ecchymoses, menorrhagia.	**Ocular form:** Uncommon. **Meningoencephalitis form:** Low with common residual neurological deficit. **Haemorrhagic form:** Approx. 50%.
Crimean-Congo haemorrhagic fever virus* (Africa, Balkans, Middle East, Asia: south of the 50th parallel north)	Fever, myalgia, dizziness, neck pain and stiffness, backache, sore eyes, photophobia, nausea, vomiting, diarrhoea, abdominal pain, mood swings, confusion, depression, liver enlargement, tachycardia, lymphadenopathy, petechial rash on internal mucosal surfaces and skin, liver and pulmonary failure.	From 10 to 40%.
Flaviviridae		
Yellow fever virus* (tropical Africa and Latin America)	Fever, muscle pain and prominent backache, headache, shivers, loss of appetite, jaundice, abdominal pain, vomiting, bleeding from the mouth, nose, eyes, stomach.	Up to 50%.
Arenaviridae	***(Old-World)***	
LCMV ** (Worldwide)	Fever, cough, malaise, myalgia, headache, photophobia, nausea, vomiting, adenopathy, sore throat, thrombocytopenia, leukopenia, meningitis, meningoencephalitis.	Less than 1%.
Lassa virus**	Fever, weakness, malaise, cough, severe headache,	1-5% of cases

(West Africa)	sore throat, nausea, vomiting, diarrhoea, sensoneural deafness, facial oedema, pleural effusion, thrombocytopenia, leukopenia, pulmonary oedema, respiratory distress, shock, encephalopathy, seizures, coma, mucosal bleeding.	develop into severe haemorrhagic fever.
Lujo virus** (South Africa)	Diarrhoea, vomiting, fever, chest pain, sore throat, rash, myalgia, facial swelling, respiratory distress, cerebral oedema, thrombocytopenia, elevated liver transaminases, fever, mild bleeding, leukopenia.	
Arenaviridae	***(New-World)***	
Junín virus** (Argentina)	Fever, myalgia, mild hypotension, conjunctivitis, soft palate or gingival margin, irritability, lethargy, lyporeflexia, haemorrhage, leukopenia, thrombocytopenia, shock, seizures.	15-30% mortality
Machupo virus** (Bolivia)	Fever, gingival haemorrhage, petechiae, nausea, gastrointestinal haemorrhage, thrombocytopenia, leukopenia, hematuria, tremor, asthenia, anorexia, respiratory distress.	Approx. 20% fatality rate.
Sabiá virus** (Brazil)	Fever headache, myalgia, nausea, vomiting, weakness, leukopenia, elevated liver transaminases, hematemesis, mucosal bleeding, conjunctival petechiae, tremors, coma, shock, pulmonary edema, hepatic haemorrhage and necrosis, gastrointestinal haemorrhage.	1 case, fatal
Guanarito virus** (Venezuela)	Fever, malaise, headache, sore throat, vomiting, abdominal pain, diarrhoea, convulsions, leukopenia, thrombocytopenia, haemorrhage.	23% fatality rate.
Chapare virus** (Bolivia)	Fever, arthralgia, myalgia, vomiting, haemorrhage.	1 case, fatal.

***Table 01:* Human viral haemorrhagic fevers**

Clinical description and fatality rate of human viral haemorrhagic fevers caused by members of the Filoviridae, Bunyaviridae, Flavividae and Arenaviridae families.

***:* http://www.who.int/**
****:* (Mclay et al. 2013)

1.2. Pathogenesis of fatal VHFs

Although the mechanisms of cell entry by VHF viruses have yet to be fully understood, a wide tropism based on virus affinity with widespread cellular receptors ensures the infection of a variety of cell types, contributing to the dissemination of the pathogen in the body (Takada et al. 2004; Gallaher et al. 2001).

A correlation has been shown to exist between receptor affinity and severity of disease in the case of acute infections (Li et al. 2003; Li et al. 2006). In the case of arenaviruses, this was exemplified by investigating the worldwide-distributed lymphocytic choriomeningitis virus (LCMV) (Radoshitzky et al. 2007). LCMV Clone 13, which is associated with chronic infection, was shown to display a higher affinity for its cellular receptor α-dystroglycan than LCMV Amstrong strain, responsible for a self-limiting infection in immunocompetent mice. This high affinity was attributed a single point mutation at residue 260 (GP1: F260L) on the receptor-binding sub-unit of the viral surface glycoprotein complex. In this particular case, an increased affinity with α-dystroglycan, a cell receptor preferentially expressed on dendritic cells, altered the functions of the host cells in the immune system, which led to viral persistence (Sullivan et al. 2011).

However, despite the rapid dissemination of the infectious agent through the body thanks to wide viral tropism, it was noted that cell and organ damage alone were insufficient to explain the terminal circulatory shock and subsequent death of patients suffering from acute VHFs. Fatal cases of VHFs are characterized by a combination of infectious and physiological conditions that result in immunological and haematological imbalances. This suggests that biochemical dysfunctions might be the prime factors leading to lethal shock.

1.2.1. High viremia in fatal VHFs

Cases of fatal VHFs generally show high viremia, a condition in which viruses enter the bloodstream, hence gaining access to the entire body. This high virus burden might be the result of a common preferential tropism for cells and tissues involved in the antiviral immune response. Several studies indicated that, among the wide variety of cell types permeable to VHF agents, monocytes, macrophages and dendritic cells were likely to act as the first ports of viral replication after the initial infection (Geisbert, Hensley, et al. 2003; Geisbert, Young, et al. 2003). An impairment of pro-inflammatory chemokines in these cell lines was therefore proposed as a contributing factor to the establishment of infection in the body.

When investigating the influence of arenavirus infection on cell functions, it was shown that the infection by Lassa virus of dendritic cells and macrophages did not affect cell viability (Baize et al. 2004). One structural feature of Lassa virions is the presence of a matrix protein Z that acts as a driving force for the budding and release of viral particles (Strecker et al. 2003). This matrix protein Z contains a RING zinc-binding domain and was shown to interact with another RING cellular protein called promyelocytic leukemia protein (PML). PML is induced by the production of interferons, a family of secretory proteins known for their antiviral and antiproliferative functions. This protein is expressed in the nucleus where it localizes within nuclear bodies (Lavau et al. 1995; Daniel et al. 1993; Dyck et al. 1994; Koken et al. 1994; Weis et al. 1994). Mutational analysis and transfection experiments unravelled the involvement of PML proteins in cell survival (Borden et al. 1997). Whereas transfection of cells with PML resulted in a two-fold decrease in cell survival in conditions of serum starvation, transfection of cells with PML mutants carrying a double-mutation in the RING-finger showed similar levels of cell survival as the controls. The observation of cell shrinkage and nuclear condensation during these experiments confirmed that cell death was triggered by apoptosis, and established PML as a pro-apoptotic factor whose functions are

mediated by its RING domain. By interacting with the PML RING-domain via their matrix protein Z, it is possible that arenaviruses could display the ability to disrupt the apoptosis machinery triggered by PML proteins, leading to a cell survival necessary to the establishment of an infection and the production of a high virus burden (Baize et al. 2004).

1.2.2. VHF-induced Immunosuppression

Common aspects among fatal VHFs are the presence of lymphopenia, a condition of having abnormal low levels of lymphocytes in the blood, and a pattern of immune responses where early infection shows an uncontrolled secretion of inflammatory mediators and late infection is characterized by suppression of the adaptive immunity (Geisbert & Jahrling 2004; González et al. 1980; Terrell et al. 1973; Yun & Walker 2012; Mahanty et al. 2003; McCormick & Fisher-Hoch 2002; Mahanty et al. 2001).

Investigation of the role of inflammatory mediators in fatal and non-fatal cases of acute Lassa fever uncovered an immunologic dysfunction correlating with fatality. Patients with a non-fatal Lassa fever presented an elevated serum level of interleukin-8 (IL)-8, a chemoattractant of neutrophiles and T-cells, that progressively declined as the illness resolved (Mahanty et al. 2001). Similarly, production of interferon-inducible protein 10 (IP)-10, a chemokine secreted in response to IFN-α, IFN-β, IFN-γ, TNF-α and lipopolysaccharide was significantly high in acute non-fatal Lassa fever cases. However, fatal cases of Lassa fever displayed low or undetectable serum levels of (IL)-8 and (IP)-10. This impairment of pro-inflammatory chemokines production was proposed to be the result of high viremia, or the result of depletion of chemokine-producing cells.

Interestingly, a similar hyporesponse was observed in fatal cases of Ebola haemorrhagic fever (Baize et al. 1999). When investigating the consequences of VHFs on the innate immune system, (Mahanty et al. 2003) demonstrated that immunosuppression could result from the infection of

dendritic cells (DC) by either Lassa or Ebola virus. Dendritic cells are antigen-presenting cells that play a pivotal role in the initiation of the T-cell response, thus acting as a link between innate and adaptive immunity. Despite being taxonomically different, both viruses were shown to readily infect DC in early stages of the infection, and most importantly to impair the DC-mediated proliferation of allogenic T-cells. Analysis of CD83 expression, a marker of DC maturation, of the ability of the cells to take-up antigens, showed a lack of maturation of DC upon infection with Lassa virus (Baize et al. 2004). Further chemotaxis assays revealed impairment in the ability of Lassa virus-infected DC to migrate towards MIP-3β, a chemokine involved in the migration of mature DC in secondary lymphoid organs. Cell entry of Lassa virus in monocyte-derived immature DCs was recently shown to be facilitated by the C-type lectin DC-SIGN at the virus-cell attachment level. DC-SIGN being present on the surface of macrophages and dendrictic cells, this facilitation mechanism clearly posed itself as a potential early determinant for efficient infection (Goncalves et al. 2013).

Taken together, these data show that immunosuppression by impairment of the mechanisms involved in the activation of adaptive immunity processes occurs in VHFs and is crucial to the pathogenicity of the disease.

1.2.3. Coagulation abnormalities

Another characteristic of VHFs is a clinical presentation that firstly suggests an impairment of the body coagulation system; and secondly suggests abnormalities in the mechanisms of fibrinolysis, a process that prevents the growth of blood clots (Geisbert & Jahrling 2004; Fisher-Hoch et al. 1987; Fisher-Hoch et al. 1988). While VHFs display haemorrhagic symptoms, mainly in the form of purpuric skin eruptions, mucosal bleeding and uncontrolled bleeding at venepuncture sites, these symptoms are generally insufficient to cause death.

In the case of Lassa virus, a study conducted in infected rhesus monkeys uncovered a link between the development of circulatory shock and the presence of biochemical dysfunctions of platelets and endothelial cells (Fisher-Hoch et al. 1987). In the experiment, three adult rhesus monkeys were inoculated with the Josiah strain of Lassa virus and blood samples were collected by venipuncture on day 0, and then daily from day 4 until death. At the time of death, which happened at day 10, 13 and 19, none of the animals presented a rash or any apparent signs of haemorrhage. Extremely low titres of fibrin degradation products, a series of blood components resulting from blood clot degeneration; the absence of significant disturbances in coagulation and the absence of platelet consumption showed that disseminated intravascular coagulation, a pathological activation of clotting mechanisms, was not an important feature of Lassa fever. However this is the case for Junin and Machupo virus infections. Although platelets counts stayed within normal haematologic ranges over the course of the disease in the three animal subjects, *in vitro* platelet aggregation assays showed a progressive depression of response to aggregation-inducing factors such as ADP and collagen. This pattern of platelet abnormalities could be found among other VHFs, with similarities observed in Ebola haemorrhagic fever (Fisher-Hoch et al. 1985), Dengue haemorrhagic fever (Mitrakul et al. 1977), and in the epizootic haemorrhagic disease of the deer (Debbie & Abelseth 1971).

In platelets and in edothelial cells, homeostasis is self-regulated by a balanced production of prostacyclin and thromboxane A_2. Prostacyclin is an independent mediator produced by endothelial cells that inhibits platelet activation and prevents the formation of the platelet plug involved in primary hemostasis. On the other hand, thromboxane A_2 is produced by activated platelets. It stimulates the activation of new platelets and increases platelet aggregation by mediating expression of the glycoprotein complex GP IIb/IIIa on their cell membrane. *Post-mortem* sampling of aortic tissues in the three rhesus monkeys infected with Lassa virus showed a decrease of prostacyclin production that hinted at a possible impairment of the endothelial functions. The animal model-derived observations were later confirmed by a prospective study of haematologic dysfunction in Sierra Leone in patients infected by Lassa virus (Fisher-Hoch et al. 1988). Patients presenting severe cases of Lassa fever were reported to suffer from mild thrombocytopenia (decrease of platelets in the blood) and depressed platelet function that occurred even when the platelet count stayed within normal ranges.

The nature of VHFs as growing threats to global health therefore lies in their high pathogenicity which seems based on efficient immunosuppressive strategies. To date, the lack of vaccine and effective treatment to treat or prevent VHF infections pose them as major medical challenges and potential bio-terrorism agents.

2. Arenaviruses

2.1. General description

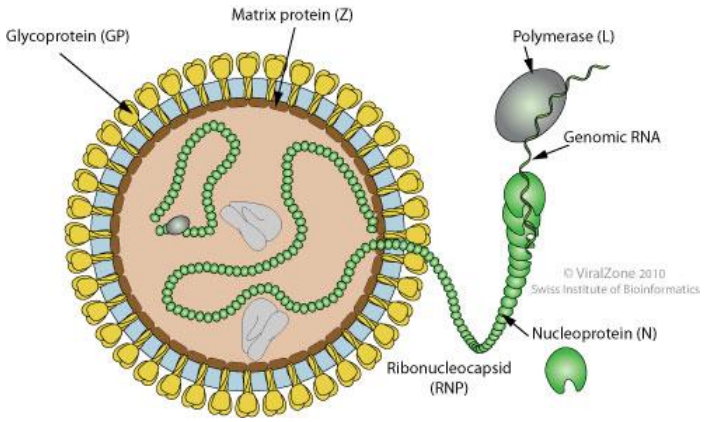
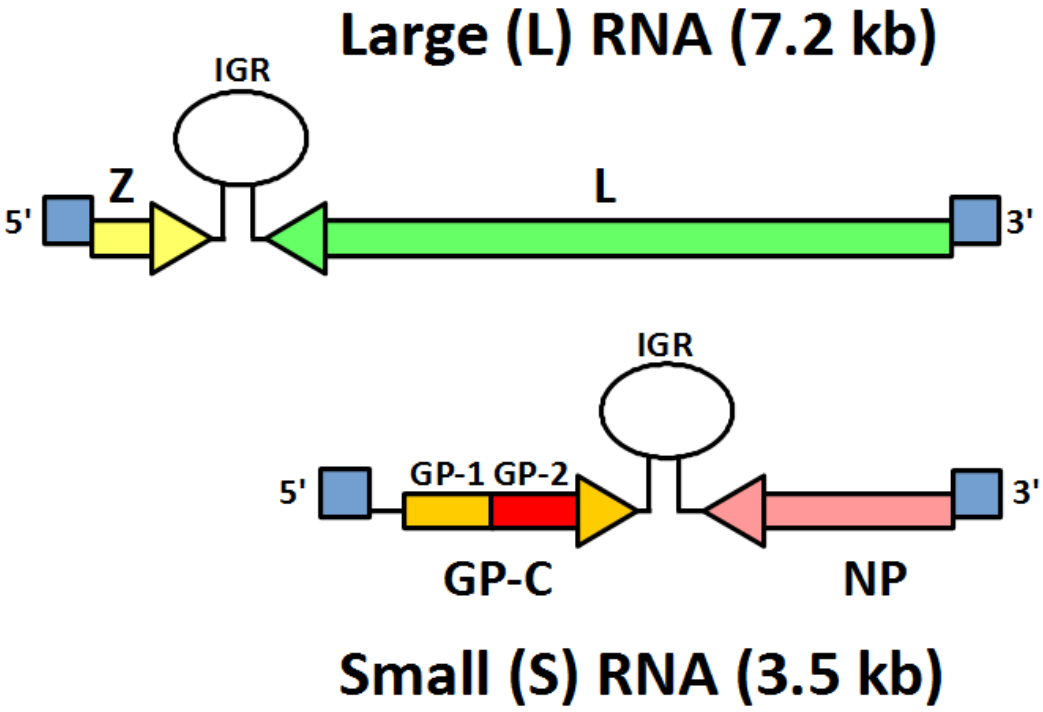


Figure 01: Schematic representation of Arenaviridae

Schematic detailing of Arenaviruses. The spherical enveloped virion displays a diameter that ranges from 60 to 300 nm. It is composed of the surface glycoprotein complex GP, the underlying matrix protein Z, the viral polymerase L and the nucleoprotein N that associates with the segmented linear RNA genome. *: http://viralzone.expasy.org*

Among VHF agents, members of the Arenaviridae merit significant attention as highly virulent emerging human pathogens (Sevilla et al. 2002; FIELDS et al. 1993; Zinkernagel 2002). The present taxonomic structure of the *Arenaviridae* comprises one unique genus *Arenavirus* (Buchmeier, M. J. 1995) that encompasses 23 recognised viruses. Arenaviruses are composed of a nucleocapsid surrounded by a lipid envelop acquired when budding out of the host cell membrane bi-layer. They contain a single-stranded bi-segmented RNA genome that encodes four structural proteins in an ambisense manner (Auperin et al. 1984). Each segment carries two open-reading-frames, separate by a non-coding intergenic region (de la Torre 2009). The viral RNA-dependent RNA polymerase and the zinc-binding matrix protein are encoded by the large genomic segment L (approx. 7.2 kb); while the surface glycoprotein precursor GP-C, subsequently cleaved into the envelope GP-1 and GP-2, is encoded alongside the nucleocapsid protein NP by two non-overlapping open reading frames located on the small genomic segment S (approx. 3.4 kb) (Figures 01 and 02).



<u>Figure 02:</u> Schematic view of the arenavirus genomic organisation

The arenavirus genome is composed of bi-segmented ambisense RNA that encodes five structural proteins. The matrix protein Z and the RNA-dependent RNA-polymerase L are encoded by the 7.2 kb-long Large segment. The glycoproteins GP-1 and GP-2 and the nucleoprotein NP are carried by a Small RNA segment of 3.5 kb. The Z and L open-reading frames on the L segment and the GP-C and NP open-reading frames located on the S segment are each separated by a non-coding inter-genomic region (IGR).

2.2. Old-World and New-World Arenaviruses

The *Arenavirus* genus can be further divided into two complexes based on serological, phylogenetical and geographical criteria: the Old-World complex and the New-World complex (Buchmeier, M. J. 1995; Charrel et al. 2008). To date, the definition of species within the arenavirus family is based on the sharing of at least two of the following criteria (i) the nature of the reservoir species, (ii) the geographic distribution of the virus, (iii) the existence of human pathogenicity, (iv) the antigenic cross-reactivity with other members of the genus and (v) the amino-acid sequence divergence from other species in the genus. As an example, a 12% criterion in the partial NP has been used to define a species within the Old-World clade (Bowen et al. 2000; Wulff et al. 1975; Bowen et al. 1997).

The Old-World arenaviruses, which include the prototype lymphocytic choriomeningitis virus (LCMV) the VHF-causing Lassa virus, the Mopeia, Mobala and Ippy viruses, are endemic in Europe, Asia and in Africa. The larger clade of New-World arenaviruses can be found in North and South America and is further divided into three major Clades (A, B and C), with Clade B comprising all VHF-associated agents: the Argentinian Junín virus (JUNV), the Brazilian Sabiá virus (SABV), the Bolivian Machupo (MACV) and Chapare viruses (CHAV) and the Venezuelan Guanarito (GUNV) virus (Clegg 2002; Charrel et al. 2008; Peters 2002; Delgado et al. 2008). With the exception of LCMV, whose natural reservoir is the house mouse, all arenaviruses are geographically restrained by the ecology of their host species. These host species are nearly exclusively rodents with the exception of the bat-specific New-World Tacaribe virus (Salazar-Bravo et al. 2002).

In 2009, a novel pathogen called Lujo virus (LUJV) was first-isolated in South Africa from an outbreak of high fatality rate haemorrhagic fever that occurred in September 2008 (80% fatality: 4/5) (Briese et al. 2009). A sequencing analysis revealed that LUJV was genetically distinct to

already-known arenaviruses, with sequence divergence spread over the whole genome, but particularly pronounced in the gene encoding the GP-1 surface receptor. Based on this genetic divergence and on a distant relationship to Old-World arenaviruses that could be established from the analysis of the matrix protein Z sequence, LUJV is thought to be branching off the ancestral node of the Old-World arenaviruses.

Recent studies investigated the possibility of virus-host co-evolution in the case of arenaviruses. A sequence analysis showing high degrees of nucleotide divergence between two Nigerian strains of the Old-World Lassa virus (GA391, LP) and the Sierra Leone Josiah strain was examined in conjunction with the geographic distribution of the four species of *Mastomys* rodent, Lassa virus's natural reservoir. The study provided preliminary evidence supporting the hypothesis of a co-evolution or co-speciation between the virus and its host species, and the potential role of virus transfer among hosts in virus evolution (Zapata & Salvato 2013).

Among the arenaviruses, at least 10 were shown to cause human diseases, and six were associated with severe VHFs: Lassa, Junín, Machupo and Chapare, Guanarito, and Sabiá. Asymptomatically present in their rodent hosts, these VHF agents are transmitted to humans by direct contact with rodents excreta, and have been classified as Biosafety Level 4 pathogens for their lethality and potential for misuse as biological weapons (Borio et al. 2002). As a consequence, designing effective anti-arenaviral therapeutic strategies is a priority to sustain global Health. To achieve this goal, detailed knowledge of the virus replication cycle is indispensable.

2.3. Cell Entry of VHF-associated arenaviruses

VHF-causing arenaviruses can be found in both serogroups of Old-World and New-World arenaviruses. As described previously, the bi-segmented arenavirus genome encodes the glycoprotein precursor GP-C gene at the 5' end of its short segment S. After being subjected to proteolytic processing, GP-C will yield the mature GP-1/GP-2/SSP tripartite surface glycoprotein complex, responsible for host-cell interaction and infection.

2.3.1. Proteolytic processing of GP-C by SKI-1/S1P

The first protein to be translated into the lumen of the endoplasmic reticulum is a 76-kDa pre-GPC polypeptide. Proteolytic cleavage of a long Stable Signal Peptide (SSP) yields the GP-C precursor that is further processed downstream at the recognition motif RRLA by the cellular subtilase SKI-1/S1P (subtilisin-kexin-isozyme-1/site 1 protease). This cleavage yields two mature glycoproteins: the distal N-terminal subunit GP-1 (44 kDa) and the membrane-anchored GP-2 (35 kDa) (Lenz et al. 2001; Beyer et al. 2003; Kunz et al. 2003; Rojek, Sanchez, et al. 2008). Although not required for cell surface expression, this proteolytic cleavage was suggested to be a necessary prerequisite to the effective incorporation of the surface glycoprotein into virions. Moreover, it was proposed that the proteolytic processing of GPC strongly depended on the amino-acid sequence forming the cytoplasmic tail of GP2 (Kunz et al. 2003). In LCMV, deletion of the three amino-acids WKR, a sequence located at the tail end of GP2 and conserved among Old-World arenaviruses resulted in lack of cleavage and loss of function of the GP complex without affecting of the cell-surface expression. This finding suggested that alterations in the cytoplasmic domain of GP had a direct impact on GP ectodomain, possibly through influencing the oligomerisation of GP and, as a result, the structure of the complex.

2.3.2. Role of SSP in proteolytic Maturation and Transport

One characteristic of the arenavirus surface glycoprotein complexes is the presence of a long Stable Single Peptide SSP that non-covalently associates with GP1 and GP2 to form the functional surface glycoprotein unit. SSP is a 58-amino-acid signal protein containing a hydrophobic core acting as a signal sequence (Figure 03) (Eichler, Lenz, Strecker, Eickmann, et al. 2003; Eichler, Lenz, Strecker & Garten 2003; Froeschke et al. 2003; York et al. 2004). In addition to its primary role as signal peptide involved in regulating intracellular transport, SSP has been shown to exhibit additional functions in glycoprotein structure and pH-dependant membrane fusion.

SSP is responsible for the transport of the GP-C complex from the endoplasmic reticulum (ER) into the Golgi apparatus, and is involved in subsequent transport to the cell surface. Expressions of wild-type GP-C and of a CD4sp-GPC chimera in which SSP was replaced by the signal peptide of CD4 were performed in Vero cells and analysed by confocal microscopy. GP-C was found to accumulate in the ER, localize in the Golgi apparatus and on the cell surface, while CD4sp-GPC was retained in the ER. However, expression of CD4sp-GPC in the presence of SSP resulted in a pattern of transport and localization similar to that of the wild-type GP-C, confirming SSP essential role in GP-C transport to the Golgi and cell surface. In this process, it was shown that the essential elements engaged in ER localization were localized within the cytoplasmic domain of GP-2, and that SSP-driven GP-C transport was achieved by the masking of these ER retention signals through direct interaction. Dibasic amino-acid motifs mediating ER localization, probably through retrograde transport, were identified in the cytoplasmic tail of Junin virus GP-2: the $KKPT_{479}$ and the C-terminal $RRGH_{485}$ sequences. Alanine mutations at either or both of these sites were shown to partially lift ER retention but SSP remained necessary for the effective transport of GP-C to the cell surface (Agnihothram et al. 2006). Additional investigations performed on Lassa virus showed that, in addition to an N-terminal meristoylation site on GX_3S/T, arenavirus SSPs contained two

independent hydrophobic regions with the potential of mediating membrane insertion and translocation of GP-C through the ER (Eichler et al. 2004). These two regions, named *h1* and *h2*, were predicted to extend from the amino-acid positions 18 to 32, and 43 to 52 respectively. Deletion of both domains, but not of *h1* or *h2* only, resulted in the non-glycosylation of preGP-C and thus was shown to prevent translocation into the ER lumen.

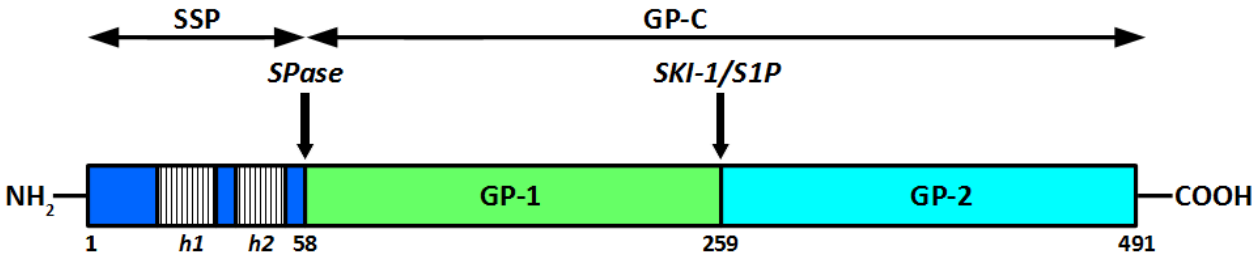
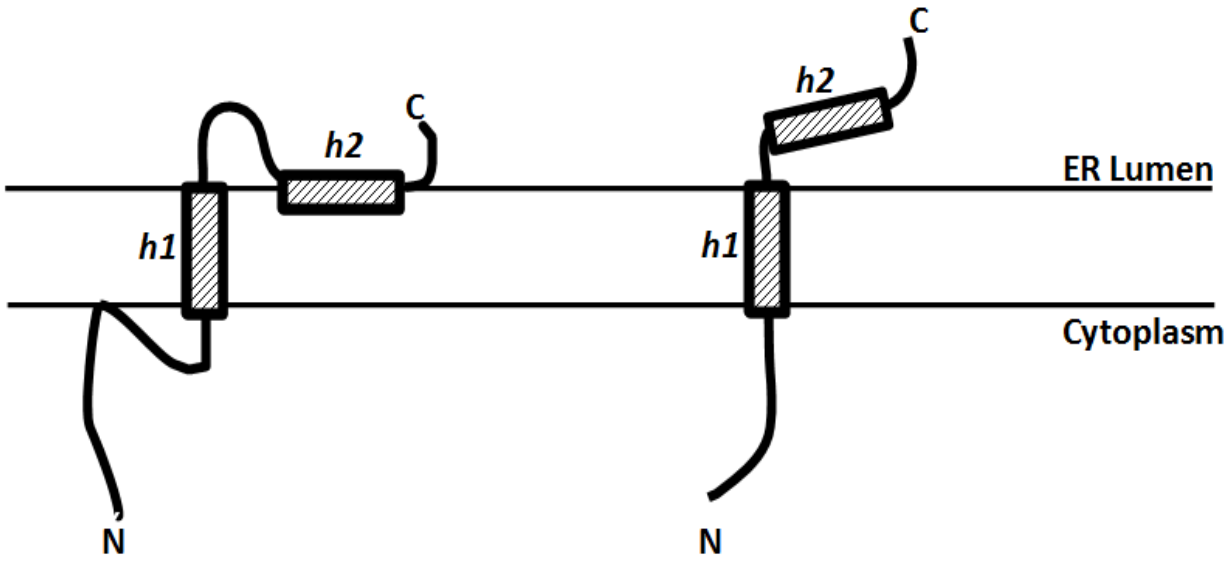


***Figure 03:* Topology of Lassa virus glycoprotein genomic segment**

Schematic representation of the Lassa virus glycoprotein genomic segment. This segment encodes the SSP (aa 1-58) and the glycoprotein precursor GPC (aa 58-491) that comprises the distal subunit GP1 (aa 58-259) and the class-I fusion transmembrane protein GP2 (259-491). SSP possesses two hydrophobic regions named h1 and h2, located respectively between aa 18-32 and aa 43-52.

Investigation of the SSP topology uncovered a novel arrangement consisting of a 17-aa N-terminal cytoplasmic peptide segment followed by the hydrophobic transmembrane domain *h1,* and a long C-terminal ER-luminal region that extends from the amino-acid 33 and comprises *h2* (Figure 04).



<u>Figure 04:</u> Putative Topology of SSP according to (Eichler et al. 2004)

Two putative alternative topology of SSP. SSP was shown to only use the N-terminal h1 hydrophobic domain for membrane insertion. Furthermore, an extended ER luminal segment is crucial for the proteolytic processing of the glycoprotein precursor GPC into GP1 and GP2.

In addition to being required for intracellular transport, SSP was found to act downstream of the ER and promote the cleavage of GP-C by SKI-1/S1P (Eichler, Lenz, Strecker, Eickmann, et al. 2003). Substitution of the wild-type signal peptide by that of the Influenza haemagglutinin (HA -SSP), or that of the cellular plasma membrane protein CD8 (CD8-SSP), in Lassa virus GP-C mutants showed that, while HA-SSP and CD8-SSP were efficiently cleaved off GP-C, the glycoprotein precursor did not undergo any further proteolytic cleavage by SKI-1/S1P to yield GP-1 and GP-2. However, when substituting Lassa-SSP by the sequence-similar LCMV-SSP, downstream cleavage of GP-C into GP-1 and GP-2 was retained, suggesting similar functions across the Arenavirus family. Further co-immunoprecipation assays and long-term pulse-chase experiments confirmed a stable direct interaction between GP-C and SSP upon cleavage by SKI-1/S1P.
By taking into account the topology of arenavirus SSP, it can be hypothesized that interaction with GP-C occurs through the SSP C-terminal luminal portion, whereas the N-terminal cytoplasm-oriented segment might be dispensable for the cleavage (Eichler et al. 2004).

Together, these studies show that the interaction of GP-C with SSP is a crucial parameter for the transport of GP-C through the secretory pathway to the cell surface; and for its proteolytic maturation leading to dissociation into GP-1 and GP-2 (Burri et al. 2013).

2.3.3. Reception recognition and Cell entry – Old-World Arenaviruses

At the viral surface, arenaviruses expose tripartite mature GP complexes formed by GP-1, GP-2 and the Stable Single Peptide. There, SSP is thought to anchor into the membrane by its N-terminal hydrophobic regions, with the C-terminal tail being part of the GP complex ectodomain (Eichler et al. 2004; Borrow & Oldstone 1992; Candurra & Damonte 1997). GP spikes are functional units responsible for interacting with host cell receptors, an attachment that leads to the virus cell entry by triggering membrane fusion (Borrow & Oldstone 1992; Eschli et al. 2006; York et al. 2004). For Old-World arenaviruses, the ubiquitously-expressed cell surface receptor for extracellular matrix proteins α-dystroglycan (α-DG) has been identified as the primary receptor for GP complexes.

Receptor recognition is mediated by the N-terminal GP-1 located at the top of the GP spike. Upon binding to α-DG, Old-world arenaviruses are internalized by endocytosis and traffic to acidified endosomes where pH-dependant fusion between the virus envelope and the cell membrane occurs. Immunoelectron microscopy of LCMV-infected rodent fibroblastic cells revealed that the entry mechanism involved large (150-300 nm diameter), smooth-walled vesicles which hinted at a clathrin-independent process (Borrow & Oldstone 1994). Subsequent studies unravelled that the Old-World arenaviruses entry in cells occurred independently of membrane cholesterol but that the process of internalization was sensitive to cholesterol-depletion, proving that Old-World aremaviruses were reliant on a cholesterol-dependant endocytic pathway (Rojek, Perez, et al. 2008). Based on the above knowledge, studies were directed towards caveolar pathways, a collection of cholestherol-dependant, clathrin-independent routes that had already been associated with virus-entry (FIELDS et al. 1993; Marsh & Helenius 2006). The overexpressions of Dominant-Negative mutants of caveolin-1, dynamin I, dynamin II and Eps15 were shown to block infections by New-World Junin virus, while having no impact on Old-World LCMV (Rojek, Perez, et al. 2008). Then, when looking at of the role of actin and microtubules in the early stages of LCMV

infection complemented, (Rojek & Kunz 2008) further characterised an actin-independent but intact microtubule-dependent process. Taken all together, Old-World arenaviruses were shown to enter permissible cell types via a cholesterol-dependant, clathrin, dynamin, and actin-independent, non-caveolar endocytic mechanism (Figure 05).

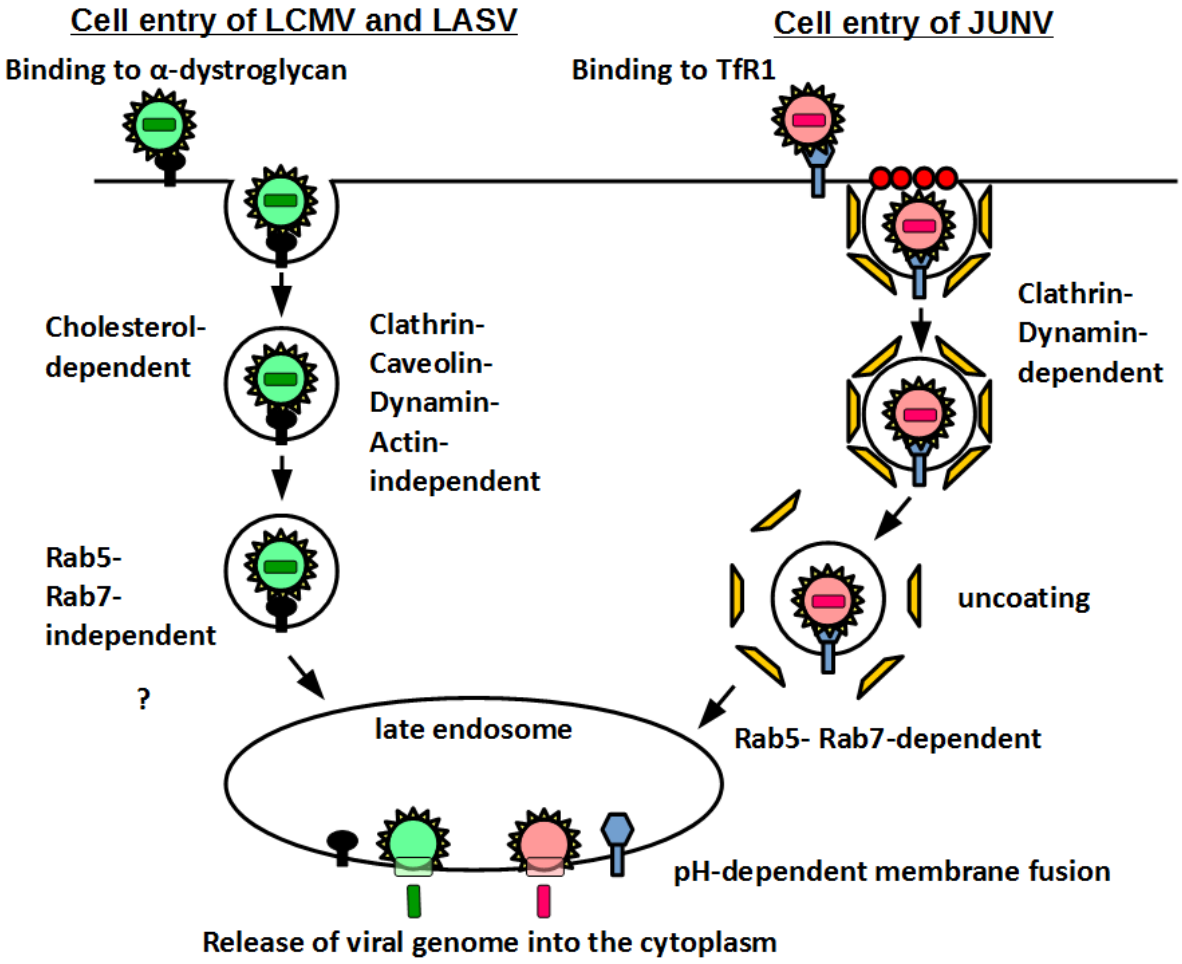


Figure 05: ***Endocytosis pathways of Old-World and New-World Arenaviruses***

LCMV and LASV enter the cell via a cholesterol-dependent, clathrin- caveolin- dynamin- actin-independent endocytosis pathway after binding to their host-cell receptor α-dystroglycan. In contrast, New-World JUNV enters the cell by clathrin-mediated endocytosis (Rojek & Kunz 2008; Kunz 2009).

2.3.4. Reception recognition and Cell entry – New-World Arenaviruses

In the case of New-World arenaviruses, a high affinity was established between the entry glycoprotein of the four VHF arenaviruses (Junín, Guanarito, Sabiá and Machupo) and Transferrin receptor 1 (TfR1), a cell surface receptor required for iron delivery into cells (Radoshitzky et al. 2007). The expression of TfR1, but not of TfR2, with which the protein shares 54% identity, was shown to increase the susceptibility of cell lines otherwise relatively non-permissive to pseudotypes of MACV, JUNV, GUNV and SABV viruses. This process was not observed in the case of Old-World LCMV and Lassa virus. Moreover, anti-TfR1 antibodies were shown to inhibit the replication of the JUNV, GUNV, SABV and MACV agents, establishing TfR1 as a cellular receptor for the South-American haemorrhagic fever arenaviruses. Interestingly, it was proposed that cell surface expression of TfR1 could become upregulated as a result of low serum iron, triggered by either iron sequestration during infection or by tissue damage (Nemeth et al. 2004; Ganz 2009). This process could be a key determinant of the viruses' high pathogenicity, and the common iron depletion that results from malnutrition in some of the VHF endemic regions of South America would become a prime factor increasing the susceptibility to VHF infections (Choe et al. 2011; Rojek & Kunz 2008).

The mechanisms of internalization of JUNV into Vero cells were investigated using compounds known to disrupt clathrin-mediated endocytosis, and compounds known to impair caveola-mediated endocytosis by altering lipid-raft microdomains (Martinez et al. 2007). Treatment of cells with chlorpromazine, a drug known to redirect the assembly of clathrin lattices to the endosome and thus causing the loss of coated-pits at the cell surface (Wang et al. 1993), was shown to inhibit virus replication in a dose-dependent manner. In the opposite way, upon exposure with Nystatin or methyl-β-cyclodextrin, two drugs promoting the sequestration of cholesterol to disrupt lipid-raft-dependant caveola-driven endocytosis (Sieczkarski & Whittaker 2002), no significant effect was

observed. Transmission electron-microscopy of resin-embedded Vero cells infected with JUNV and fixed with glutaraldehyde allowed the observation of JUNV particles in an asynchronous entry process that involved clathrin-coated pits and further internalization by clathrin-coated vesicles (Figure 05).

2.3.5. GP-mediation of membrane fusion

The first insights into the mechanisms of viral nucleocapsid delivery into the cytoplasm have been obtained by (Borrow & Oldstone 1994) who used inhibition techniques and immunoelectron microscopy to shed light on the mechanisms of LCMV entry in rodent fibroblastic cell lines. These authors utilized acidotropic weak bases like ammonium or chloroquine to raise the pH of endosomal compartments in order to assess the hypothesis that arenaviruses delivered their genomic materials in the cell via pH-triggered membrane fusion in the endosome. Acidotropic weak bases diffuse into the cell in their non-protonated form, and subsequently become protonated in acidic compartments such as the endosome, where their accumulation triggers a raise in the vesicular pH. Based on the same idea, carboxylic ionophores (nigerin and monensin) were also used to trigger a similar perturbation in the endosomal pH, this time triggered by ion-exchanges. Immunoelectron microscopy assays targeting the NP protein after treating rodent fibroblastic cells with the above-cited compounds showed that increasing the endoplasmic vesicular pH resulted in blocking LCMV infection. To confirm that acidotropic weak bases and carboxylic ionophores were disturbing the mechanism of virus entry, particular attention was drawn to linking inhibition assays and kinetics of virus entry. It was shown by electron-microscopy that the process of entry for LCMV was an asynchronous process that occurred over a period of 1 to 2 hours. In the experiment, treating the cell until the end of the virus absorption period resulted in the inhibition of LCMV infection in 75-85% of the cells. However, extending the treatment period

by 1-2h resulted this time in a complete inhibition. Together, these data show that arenavirus entry involves a pH-dependant fusion step in the acidic endosomal compartment.

Another study has utilized the R18 fluorescent dequencing assay to first establish that membrane-fusion occurring in the endosome in the context of LCMV infection was protein-mediated (Di Simone et al. 1994). In this experiment, large multilamellar liposomes were labelled using R-18 (octadecyl rhodamine B) and the occurrence of fusion with LCMV observed by dequenching. The liposomes were designed to mimic the lipid composition of an endosomal membrane with PC:SM:PE:PS:PI:Gang:Chol, 5:1:1:1:1:1:3. However, the dequenching became inhibited after pre-treatment of the LCMV virus with proteinase K, chymotrypsin or with the utilisation of a GP-1 subunit-targeting antibody. This indicated that the arenavirus fusion event was mediated by the glycoprotein complex GP. Monoclonal antibodies were subsequently used to determine whether conformational changes occurred in GP upon exposure to acidic pH. These antibodies specifically targeted epitopes located on either the GP-1 or the GP-2 subunit (4 on GP-1 including 3 conformational epitopes and 1 linear epitope. 1 single linear epitope on GP-2). Prior to the ELISA screening, LCMV was incubated in PBS in acidic (pH 5.0) or neutral (pH 7.0) conditions before being adjusted back to neutral pH. The experiment showed a sharp decrease in the binding abilities of antibodies directed against the conformational epitopes located on GP-1, while for the linear GP-1 epitope binding remained unchanged. Furthermore, antibody-binding to the GP-2 linear epitope was showed to increase by a factor of 4 after acidic treatment of the virus. Increased exposure of GP-2 during exposure of the GP-complex to acidic pH suggested a possible dissociation of GP-1 from the complex in addition to acidic-pH-induced conformational changes that resulted in the inaccessibility of GP-1 conformational epitopes.

The putative process of acidic-pH-induced dissociation of the GP-1 subunit was also assessed by gel electrophoresis. The treatment of LMCV with acidic pH resulted in the finding of GP-1 proteins in the supernatant, and in the easing of the pelleting process of the virus, suggesting a possible aggregation of the virions that could be driven by the loss of GP-1 and exposure of hydrophobic cores in GP-2. Subsequently sucrose gradients showed that GP-1 only dissociated from LCMV when the virus was pre-treated at pH 5.0 before gradient separation (Di Simone et al. 1994).

The kinetics and pH dependence of the arenavirus entry into cells have been addressed using fluorescent membrane fusion assays (Di Simone & Buchmeier 1995). Using a fluorescent pair of NBD-PE (*N*-4-nitrobenzo-2-oxa-1,3-diazole phosphatidylethanolamine) and Rd-PE (*N*-lissamine rhodamine B sulphonyl-dioleyl phosphatidylethanolamine) in target liposomes to monitor the process of membrane fusion, these authors established that LCMV fuses with target liposomes at pH values lower than 6.3, with minimal activity between pH 6.3 and 5.8; and with pronounced fusion happening when pH values fell to 5.7 and below. This suggests that membrane fusion in LCMV likely occurs in the early and late endosome. Furthermore, the authors suggested that the three distinct activity behaviours cited above (No fusion pH> 6.3 – Minimal activity 5.7<pH<6.3 – High activity pH<5.8) may be related to changes of conformation within the GP complex, where the minimal fusion activity between pH 6.3 and 5.8 would reflect a transitional state of GP where the conformational changes involved in mediating membrane fusion occurred at a slower rate (Di Simone & Buchmeier 1995).

Focusing on the kinetics of GP-1 dissociation revealed that the glycoprotein complex indeed existed in three different states as hypothesized above. The three different states were linked to the conditions of environmental pH, and irreversibly catalysed membrane exchange: (1) at neutral pH, the GP spike is inactive and cannot trigger membrane fusion (2) upon exposure to acidic pH,

conformational changes occur within the spike and catalyse the mechanisms of membrane fusion (3) exposure to low pH (5.0 and below) leads to the dissociation of GP-1 (Di Simone et al. 1994). To sum up, arenavirus envelope GPs are fusion-capable protein complexes that become activated upon exposure to low pH, and membrane fusion occurs thanks to a series of conformational changes that, in LCMV, ultimately results in the irreversible dissociation of the GP-1 subunit from the spike complex.

2.3.6. GP-2 is a putative class-I transmembrane fusion protein

When investigating the mechanisms of membrane fusion in LCMV, (Di Simone et al. 1994) drew attention on the fusion activity of viruses, which had already been shown to be dependent on highly hydrophobic regions called “fusion peptides”. These peptides are present in other viral glycoproteins, and have already been computationally predicted for the arenavirus envelope proteins GP-1 and GP-2 (Glushakova et al. 1990). Sequences of GP-C proteins from Lassa virus, LCMV, Pichinde and Tacaribe viruses were subjected to alignment and putative fusion peptides were selected based on a series of criteria. The putative signal peptide had to be (1) exposed to the extracytoplasmic environment and therefore needed to be excluded from the transmembrane segment of the complex, (2) be around 20 amino-acids in length, (3) displayed a hydrophobic index of 0.5 or higher on the scale of Kyte and Doolittle (Kyte & Doolittle 1982), (4) showed a high degree of sequence conservation, (5) be located remotely from any potential glycosylation sites, (6) be predicted to predominantly fold in α-helices, and (7) be located at the core of the glycoprotein spike structure. Analysis of the Lassa virus GP-1 and GP-2 glycoprotein sequences revealed six long hydrophobic regions, peptides I-VI, with lengths varying from 15 to 24 amino-acids. They were found to be located on GP-1 (peptides I-III), and on GP-2 (peptides IV-VI). One of them, peptide IV, drew particular attention as its location, at the N-terminus of GP-2, corresponded to that of identified fusion peptides found in different families of viruses. Peptide IV, found between the animo-acid residues 258 and 275 of LASV GPC, displayed a hydrophobic index of 0.450 and showed a high degree of remoteness to its nearest glysosylation site by being located 90 amino-acids away. It also contained a Gly-XD-Phe tripeptide, which was found to be a frequent sequence motif in enveloped virus fusion peptides (Gallaher 1987). Another potential candidate for the LASV fusion peptide was peptide V of LASV GP-2, a 25 amino-acid long sequence found to have 75% of its residues engaged in an α-helical conformation and to exhibit a high degree of sequence homology

with the other arenaviruses studied. Even if this study did not draw any final conclusion on the location of a putative fusion peptide in arenavirus GPs, it highlighted a potential role of fusion protein for the glycoprotein GP-2.

Later investigation carried out on the GP-2 protein of New-World Junín virus has revealed the existence of two hydrophobic heptad-repeat regions thought to fold into a six-helix bundle conformation, a particular structure that forms the active core of Class I viral fusion proteins (York et al. 2005). In this study, alanine-scanning mutagenesis of putative interhelical ***a*** and ***d*** position side-chains revealed that four residues in the N-terminal region of the protein (I333, L336, L347, and L350), and two residues in the C-terminal section (R392 and W395) played an important region in a JUNV envelope glycoprotein-mediated cell-cell fusion assay, supporting the idea the arenavirus GP-2 were Class I viral fusion proteins.

Additional sequence analysis undertaken in the GP-2 subunit of the LCMV GP has identified two extended α-helix regions located between aa 313 and aa 365; and between aa 402 and aa 429 (Eschli et al. 2006). These extended α-helix regions were shown to be highly conserved among Old-World and New-World arenaviruses. In the first region, a conserved heptad-repeat region was also revealed, extending between aa 327 and aa 365. Expression of a GP-2 sequence containing the heptad repeat region, a disulphide-bonded loop connected the two identified α-helices, and the helix number 2, gluteraldehyde cross-linking experiment and subsequent circular dichroism spectroscopy were used to show that the ectodomain of LCMV GP-2 spontaneously adopted an all-alpha trimeric six-helix bundle conformation.

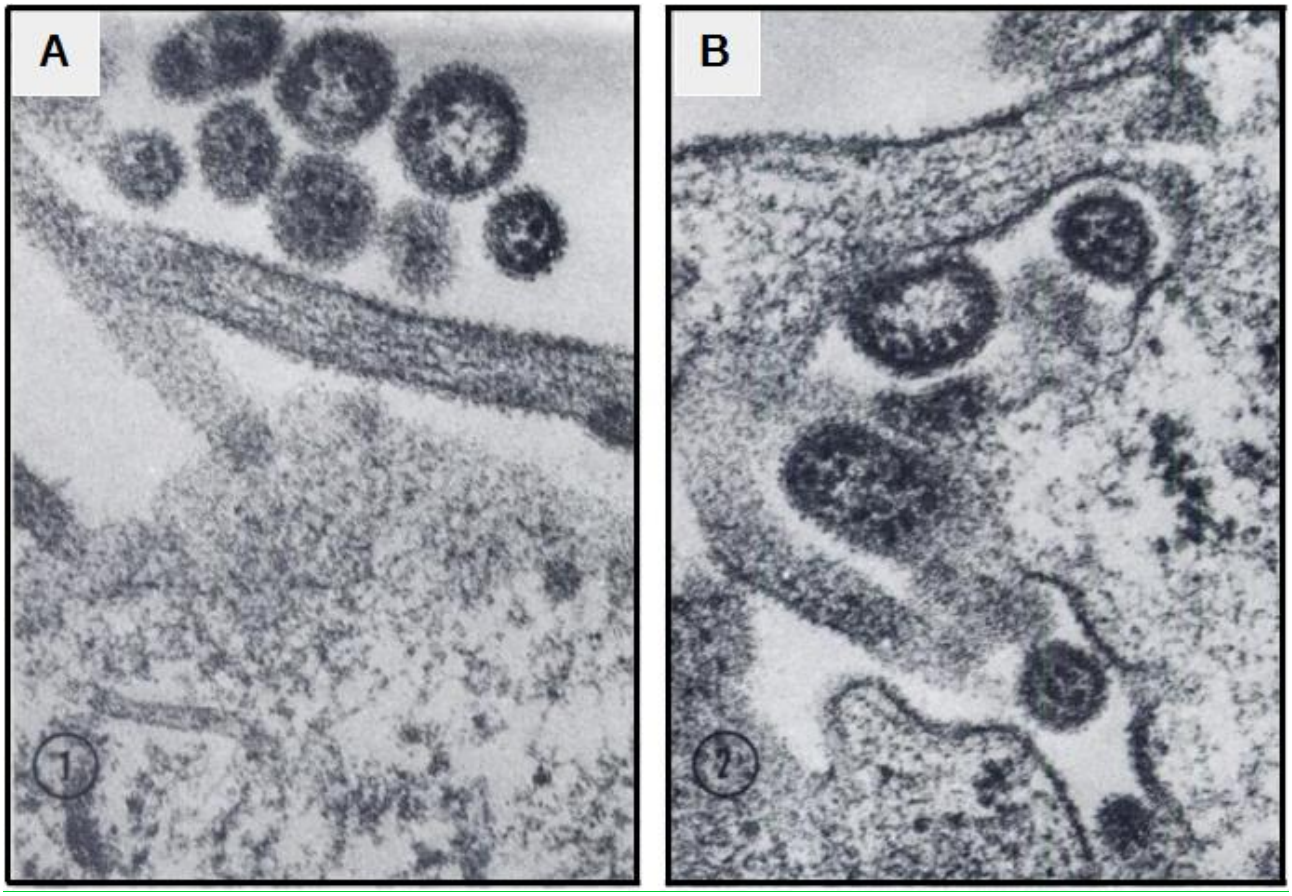
The surface glycoprotein complex is a crucial element to understand viral pathogenesis, as reflected by its structural organisation and resulting protein functions. Being involved in the earliest stages of infection, GP is a target of choice for the design of therapeutic molecules that

would prevent the virus from entering a permissible cell by either disrupting the receptor binding and/or the membrane fusion process. Among high priority targets in the search for anti-arenaviral therapeutic strategy is the Old-Wold Lassa virus, a virulent pathogen responsible for human haemorrhagic fever epidemics in West Africa.

3. Lassa virus

3.1. History and Epidemiology of Lassa Fever

Lassa virus (LASV) is an emerging zoonotic pathogen endemic in Western Africa where its natural host, the peridomestic multimammate rat, is prevalent (Maxfield 1982; Ohkuma & Poole 1978). It required its name after a small community in North Nigeria where the disease was first described in a missionary nurse in 1969 (Buckley et al. 1970; Frame et al. 1970; Leifer et al. 1970; Buckley & Casals 1970; Speir et al. 1970). The very first reports on Lassa fever were quick to establish the zoonotic character of the agent. (Frame et al. 1970) wrote: “It is likely primarily a disease of some non-human form of animal life, and its severity suggests that it is not well-adapted to the human host”. The initial clinical description of Lassa fever featured early unspecific symptoms like myalgia, weakness, lassitude, and mild pharyngitis associated with fever. This unspecific clinical presentation evolves over the course of the infection to exhibit an absolute or relative leukopenia (decrease in the number of leukocytes), an excess of fluid between the two pleural layers of the lungs, and renal failure (Frame et al. 1970). Electron microscopy of Vero cells infected with the Lassa-originating pathogen revealed viral particles that were “of variable size and shape”, “Projections or spikes [that] could sometimes be seen on the surface of particles and occasionally at a thickened area of the cell membrane that suggested a particle being formed by budding” (Figure 06) (Speir et al. 1970).



<u>Figure 06:</u> Electron microscopy of Vero cells infected by Lassa virus

Lassa virus is first observed as extracellular virions (A) and in vesicles or extracellular invagination (B) using a Philips EM-200 electron microscope at a magnification of x103,000 (Speir et al. 1970).

Since its initial discovery, Lassa virus has been responsible for series of outbreaks occurring in two geographically distinct endemic foci: the Mano River region, comprising Guinea, Sierra Leone and Liberia; and to the East the region of Nigeria (Fichet-Calvet & Rogers 2009). The distribution of Lassa virus seroprevalence in humans correlates with the ecology of its host *Mastomys natalensis*. The rodent was first described as a reservoir during an epidemic in Sierra Leone in 1972, and further isolated from rodents trapped in the Benue-Plateau and North-Eastern States of Nigeria where cases of Lassa fever had been previously reported (Monath et al. 1974; Wulff et al. 1975). More recently, *Mastomys natalensis* has been described as the only host source of Lassa virus in Guinea, ruling out an earlier assumption that two additional rodent species, *M. erythroleucus* and *M. huberti*, could be acting as co-hosts (Lecompte et al. 2006).

LASV transmission to humans occurs by direct contact with rodent excreta. Studies in Guinea established a clear connection between the incidence of Lassa fever in humans and *Mastomys natalensis* ecology, linking the rodent seasonal behavioural patterns with the prevalence of infections. Indeed, the incidence of Lassa fever in West Africa was reported to increase during dry season compared to rainy season, a specific period during which the rodent was found to be significantly aggregating inside houses, increasing the likelihood of direct contact with infected individuals (Figure 07) (Fichet-Calvet et al. 2007).

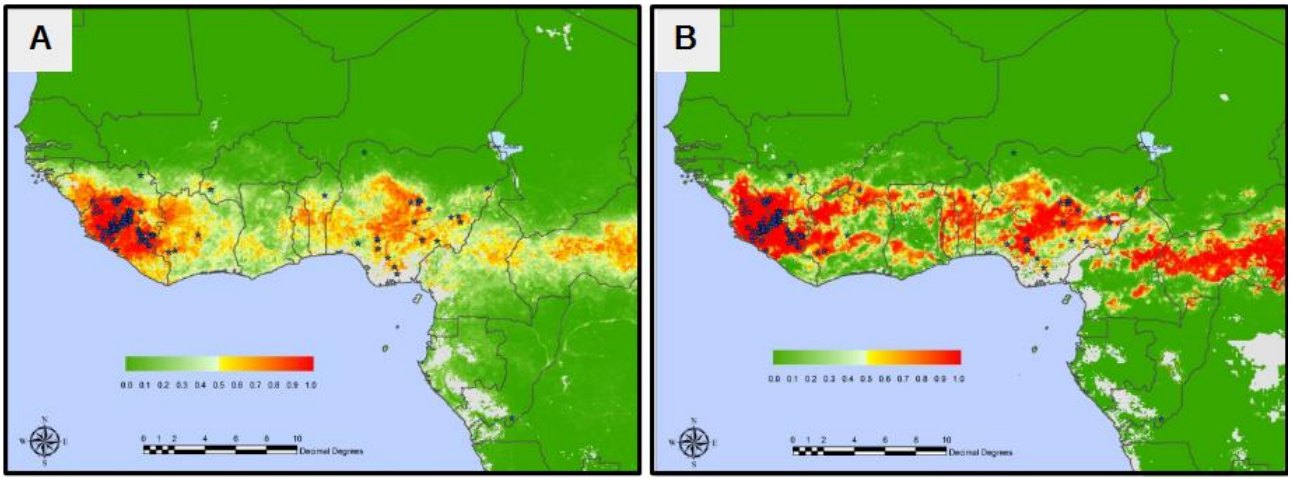


***Figure 07:* Predicted risk areas for Lassa fever in West Africa**

Two different models predict the risk areas for Lassa virus infection in West Africa according to the distribution of the Lassa virus host, and indications of temperature, rainfall and vegetation. No risk is defined by the colour green while high risk of infection is represented by a red colour. Blue stars represent epidemic outbreaks (Fichet-Calvet & Rogers 2009).

3.2. Clinical manifestation of Lassa Fever

In the reported 300,000-500,000 annual cases, 1-5% develop into a serious haemorrhagic fever marked by severe immunosuppression and multi-organ failure (Isaäcson 1975). The incubation period ranges from 1 to 24 days and from 7 to 18 days, with patients being viremic as early as day 3 of the course of illness (McCormick et al. 1987; Mertens et al. 1973; Demby et al. 1994; Trappier et al. 1993; Johnson et al. 1987).

In addition to the symptoms described earlier, Lassa fever is associated with pulmonary oedema, ascites (excessive accumulation of fluid in the abdominal cavity), and haemorrhage of the gastrointestinal mucosa; with microscopic lesions including hepatocellular necrosis, spleen necrosis, interstitial nephritis, interstitial pneumonitis and myocarditis (Günther & Lenz 2004; Edington & White 1972; Walker et al. 1982; McCormick et al. 1986; Winn et al. 1975). A peak of viremia is attained around day 4 to 9 and, in patients surviving the disease, the agent is cleared from the blood around 3 weeks after the initial onset of the symptoms (Günther & Lenz 2004; Mertens et al. 1973; Demby et al. 1994; Trappier et al. 1993; Johnson et al. 1987).

In fatal cases, death occurs after an average period of 12 days and is associated with a terminal stage of illness that clinically displays a state of circulatory shock. Several parameters were shown to increase the chance of developing fatal Lassa fever. As an example, poor prognosis was associated with LASV infection during pregnancy. A prospective study in a hospital in Sierra Leone showed that, in pregnant women, the fatality rate was increased to 21%, with significant impact on survival prognosis in the third trimester (30% death rate) compared to the first two trimesters (7%). Infection of women during pregnancy automatically resulted in infection of the foetus, which led to miscarriage or stillbirth in 87% of the studied cases. Interestingly, delivery, spontaneous

abortion or evacuation of the uterus, were shown to significantly improve women's chances of survival (Price et al. 1988).

As already presented for other VHF-causing agents, pathological investigations have revealed that lesions alone could not be accounted for death in cases of Lassa haemorrhagic fever. To study the link between high viremia and prognosis, circulatory levels of virus have been measured in 209 sequential specimens originating from 137 patients presenting Lassa fever (Johnson et al. 1987). In this study, an excess in mortality was clearly observed in patients with viremia levels superior to 10^3 $TCID_{50}$/ml (20.7% with AST<150 against 13.8% for low viremia/AST<150 and 12.5% for low viremia/AST>150), and dramatically increased with the presentation of high values (>150 U/l) of serum aspartate aminotransferase (AST), a marker for liver health (Fatality rate: 78%).

Impairment of the immune system is also a crucial aspect of Lassa virus pathogenicity, and was shown to be linked to Lassa virions structural components.

3.3. Therapeutic Strategies and Preventive Medicine

The classification of Lassa virus as a biosafety level 4 pathogen reflects the potential life-threatening character of the infection in humans and the current lack of vaccines and antiviral drugs approved for the efficient treatment of a resulting haemorrhagic fever.

The only antiviral compound that has been proven to display any beneficial therapeutic properties to combat Lassa virus is ribavirin, a guanosine analogue that interferes with the RNA metabolism required for the process of viral replication. The efficacy of ribavirin was first assessed in 1986 in acute cases of Lassa fever that occurred in the West African country of Sierra Leone (Geisbert & Jahrling 2004; FIELDS et al. 1993; Hastie 2011; McCormick & Fisher-Hoch 2002). Statistics established that, at the time of admission into hospital, a serological level of aspartate aminotransferase of equal or higher than 150 IU per litre was associated with a fatality rate of 55%, and this figure was used as a reference to demonstrate the therapeutic potential of ribavirin. The intravenous administration of ribavirin for a period of 10 days to patients showing similar clinical conditions was shown to decrease the fatality rate to 5% if the treatment was initiated within 6 days of the onset of fever. On the other hand, if a similar ribavirin treatment was started after seven or more days following the first onset of symptoms, the efficacy of this antiviral administration was shown to decrease with a case-fatality rate increased to 26%. In a similar way, it was established that patients admitted to hospitals with viremia levels superior or equal to $10^{3.6}$ $TCID_{50}$ per millilitre faced a fatality rate of 76%. Treatment with ribavirin begun within 6 days of illness had the effect of decreasing this fatality rate to 9%, whereas treatment started after seven days or more lowered the mortality rate to 47% only. These results showed that ribavirin could prevent death if administered in the early course of the disease, preferentially within 6 days of the initial onset of fever.

Considering the limited availability of ribavirin and the challenges in establishing an early diagnosis for Lassa fever, a preventive vaccine appears as a critical need to efficiently reduce the public health threats caused by the disease. The first efforts to create a vaccine that showed high-efficacy were those of Fisher-Hoch and collaborators (Fisher-Hoch et al. 2000). These investigators used recombinant vaccinia viruses expressing the NP protein, the full-length glycoprotein GP-C and the individual proteins GP-1 and GP-2 in different combinations to vaccinate 44 macaques vaccinated before being challenged with 10^3 to 10^4 PFU of live Lassa virus (Josiah strain). Whereas all non-vaccinated animals died following the challenge with live Lassa virus, macaques vaccinated with vaccinia viruses expressing all the structural proteins listed above displayed a significant survival rate of 90%. One particular aspect to notice is that, among macaques that were pre-treated with vaccinia viruses carrying GPC alone, 88% of the individuals survived although none displayed high titres of antibody (Fisher-Hoch et al. 2000) . On the other hand, macaques that received vaccines that carried only the NP protein displayed high titres of antibody but a poor survival rate of 20%. Similarly, all macaques vaccinated with a combination of GP-1 and GP-2 proteins injected at different sites survived, while all those vaccinated with either GP-1 or GP-2 died (Fisher-Hoch et al. 2000) . While this strategy could have paved the way to create a highly efficient preventive vaccine thanks to the immunogenic potential of epitopes located on GP-1 and GP-2, potential side-effects resulting from the use of vaccinia virus rendered the protocol unsuitable for the development of human vaccines.

More recent studies have involved the generation of a live attenuated, recombinant vesicular stomatitis virus that expresses the glycoprotein GP-C of Lassa virus Josiah strain (Geisbert et al. 2005). To test its efficacy, four macaques were vaccinated intramuscularly using the combination described above and two others received an equivalent injection of recombinant vesicular

stomatitis virus carrying the Ebola Zaire glycoprotein as a control. All six primates were then challenged with a single dose of 10^4 PFU of Lassa virus Josiah strain 28 days after vaccination. The two control animals started showing clinical signs of illness on day 3, and succumbed to Lassa fever on day 11 and 13. On the other hand, none of the vaccinated animals declared any signs of illness and were thus fully protected against Lassa virus infection. Blood analysis revealed that by the time of challenge, all four primates vaccinated with LASV GP-C had developed moderate- to high-levels of immonuglobulin G titres, and low-levels of neutralizing antibodies directed against Lassa virus. Potent cellular immunity was also demonstrated by the production of IFN-γ or TNF-α in CD8 cells in three of the four challenged GPC-vaccinated monkeys, including two who displayed =IFN-γ- or TNF-α-positive CD4 cells. These results indicated that live attenuated vaccines based on the vesicular stomatitis virus were a potent option for the future development of a Lassa virus vaccine.

As Lassa virus high pathogenicity has been shown to derive from the biochemical properties and interactions of its structural components with cellular partners (Mclay et al. 2013), a detailed description of the viral proteins' structure is crucial for the design of therapeutic strategies, both preventive and post-exposure.

3.4. Lassa virus

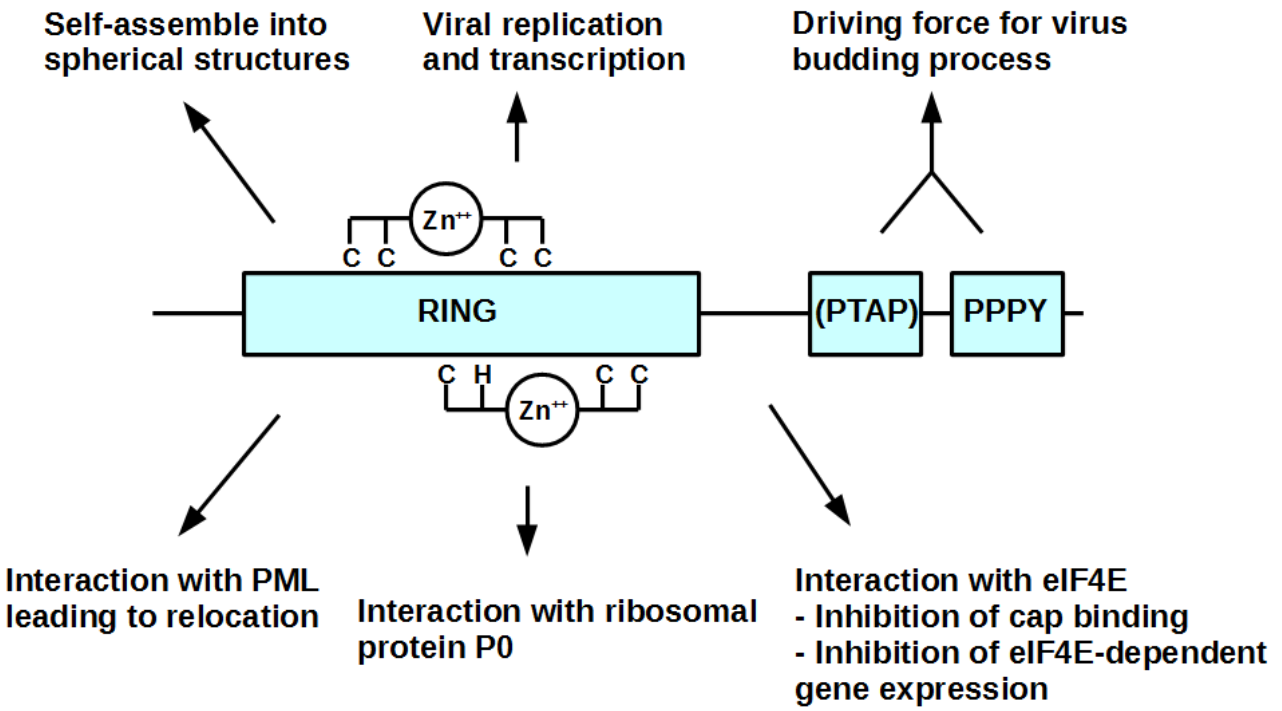
As an arenavirus, LASV possesses a bi-segmented ambisense RNA genome that encodes five different structural proteins: the surface envelope GP-1 and GP-2, the matrix protein Z, the RNA-binding nucleoprotein NP and the viral polymerase L (FIELDS et al. 1993). These structural proteins assemble into pleiomorphic virions that display a diameter ranging from 60 to 300 nm and morphology from spherical to oval. They acquire their envelope when budding from the host-cell membrane, a process that was shown to be mediated by the matrix protein Z (Salvato et al. 1992; Perez et al. 2003). Investigations into the structure and roles of LASV components have revealed key functions that may prove to be prime targets for the design of therapeutic molecules destined to block the viral replication process.

3.4.1. Matrix protein Z: Driving Force of arenavirus budding process

The matrix protein Z is a small RING finger protein (11 kDa) that was first suggested to play a role in the regulation of the viral transcription in the case of the New-World Tacaribe virus (Garcin et al. 1993).

The production of infectious virus-like particle by transfection of HEK-293T cells with combinations of plasmids encoding the NP, L, GP and Z proteins of LCMV has been used to study the requirements of structural proteins for the efficient production of VLPs (Perez et al. 2003). This has helped to establish the role of Z as the driving-force of arenavirus budding process. Perez and collaborators found that GP and NP were not required for the production of VLPs, with the absence of GP only resulted in a small decrease in their production by infected cells. On the other hand, the absence of expression of the Z protein dramatically reduced the production of VLPs, showing that Z acted as the main driving-force behind the process of LCMV budding. On the side, Z was also shown to inhibit the production of LCMV minigenome, an inhibition partially lifted by the

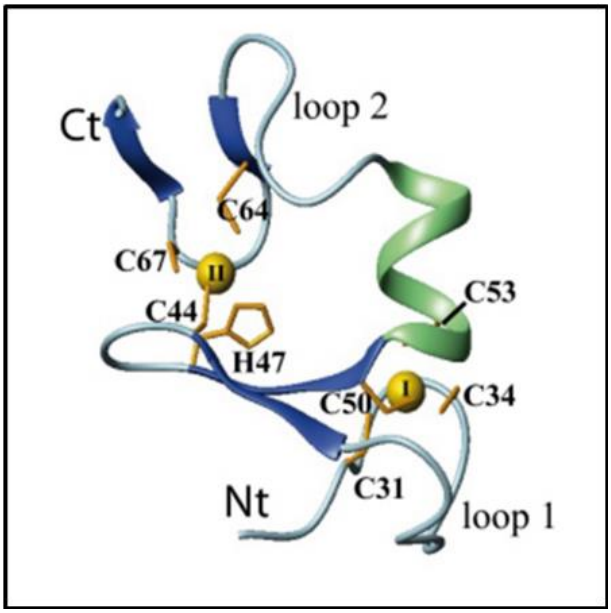
co-expression of GP. When looking at LASV Z, it was discovered that the protein possessed two L motifs, “late” domains known to be involved in the last stages of virus budding. These domains, PTAP and PPPY, were separated by eight amino-acids and tested to determine whether they contributed or not to the budding of LASV VLPs. Mutants of LASV Z proteins, which presented either one or both L domains altered, were expressed in transfected cells. Minigenome RNA associated with VLP production was amplified by PCR and showed that both proline-rich domains were required for the efficient budding. The entire study identified Z as being the driving-force of the arenavirus budding process, a protein that exhibited self-budding properties and proline-rich motifs in its C-terminus section. Considering that Z possesses a RING finger, it is logical to hypothesize that this domain could be involved in the recruitment of cellular proteins required for efficient budding. The study cited above showed in its final results that the cellular protein Tsg101 was implicated in arenavirus budding, suggesting that the Vacuolar Protein Sorting machinery could be a target of arenavirus to allow budding to occur. Identifying and characterizing these interactions would provide grounds for the design of an antiviral strategy preventing the release of newly-formed virions from cells (Figure 08).



<u>Figure 08:</u> Interaction map of Lassa virus and LCMV matrix protein Z

Summary of the Old-world arenavirus matrix protein Z interactions. Z is a 11 kDa RING protein involved in the viral replication, transcription and budding processes. The Cysteines and Histines residues involved in the coordination of Z two zinc atoms are shown, as well as the late domain motifs PPPY and PTAP, the latter being only present in Lassa virus (Günther & Lenz 2004).

The structure of the LASV Z protein has been studied by solution NMR to investigate its interaction with the eukaryotic translation initiation factor eIF4E, a cellular factor that modulates the expression of genes involved in both proliferation and cell survival (Figure 09) (Volpon et al. 2010; Culjkovic et al. 2007; Culjkovic et al. 2008). The RING domain of LASV Z, which extends from residue 30 to 70, folds into a helix (residues 50-58), two antiparallel β-sheets (β1: aa 42-44 and 47-49; β2: aa 63-64 and 69-70), and two loops (aa 35-39 and aa 58-62) (Volpon et al. 2010) . Two zinc atoms are coordinated in cross-brace fashion, which occurs through one histidine and seven cysteine residues that obey the C_3HC_4 finger consensus (Volpon et al. 2010). Z interacts with eIF4E via helix 1 resulting in the inhibition of eIF4E-dependent gene expression. This has been hypothesized to contribute to the inhibition of IFN production in plasmocytoid dendrictic cells (Volpon et al. 2010).



<u>Figure 09:</u> NMR solution structure of Lassa virus Matrix protein Z

Ribbon representation of LASV protein Z. The protein is composed of a RING motif between aa 30 and 70, helix 1 (aa 50-58 green), two anti-parallel β-sheets (β1: aa 42-44 and 47-49; β2: aa 63-64 and 69-70, blue), and two loops (aa 35-39 and aa 58-62). The two coordinated zinc atoms are indicated by yellow spheres (Volpon et al. 2010).

3.4.2. Nucleoprotein NP: Exonuclease and immune evasion

The best known function of the NP protein is to associate with the viral polymerase L to form the ribonucleoprotein, a complex in which it regulates the production of subgenomic mRNA, full-length genomic and antigenomic RNAs, in the steps of transcription and replication (Pinschewer et al. 2003). X-ray crystallography of the LASV NP has revealed the functional mechanisms of the protein in viral RNA synthesis and shed light on the role of NP in suppressing the host interferon immune response (Qi et al. 2010; Hastie et al. 2011b). The 569-amino acid full length NP of the LASV Josiah forms mainly trimers and some hexamers, with both species retaining their ability to bind random RNAs (Qi et al. 2010). When subjected to crystallisation trial experiments, only the trimeric complex has yielded crystals (Qi et al. 2010). These crystals formed in space group *P3* with three subunits in an asymmetric unit. The structure has been solved to a resolution of 1.80 Å with a final model displaying an R_{factor} of 0.18 and an R_{free} of 0.20 (Qi et al. 2010) (Figure 10).

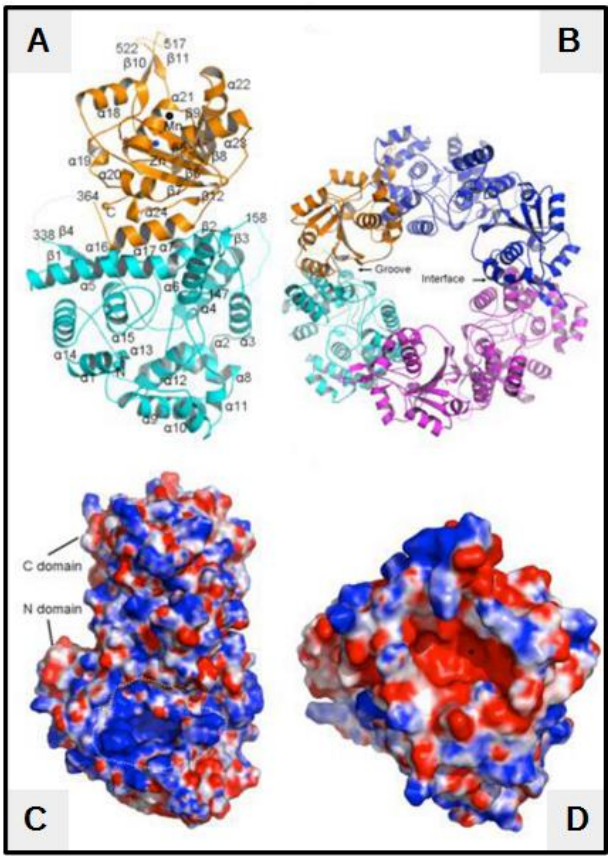


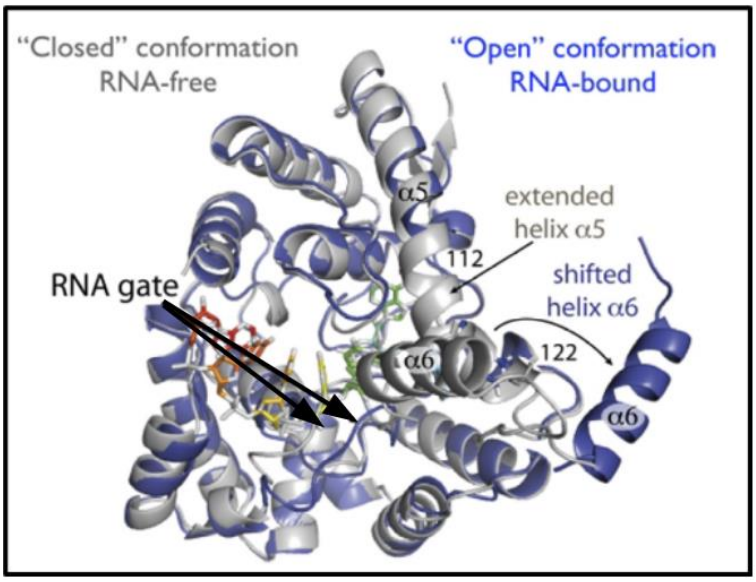
***Figure 10:* Crystal structure of LASV nucleoprotein**

(A) Diagram of the LASV protomer. The N-terminal domain is indicated in blue and folds mainly in α-helices. The C-terminal in yellow adopts a typical α/ß/α sandwich structure. A black sphere shows the location of a magnesium ion, and a blue sphere that of a zinc atom. Dotted lines are disordered loops. (B) LASV NP proteins arrange into a trimeric ring structure. The first protomer displays the same colours as (A), the second is coloured in blue and the third in purple. (C) and (D) are electrostatic surface potential maps of one NP protomer. The blue colour represents positive charges and the red colour negative charges. (D) a negatively-charged groove thought to be the 3'-5' exonuclease active site lies at the top of the C-terminal domain and comprises a Mn2+ ion (black sphere) (Qi et al. 2010).

The LASV NP is formed of a large N-terminal domain (residues 7-338) that mainly folds into α-helices and coils, while the smaller C-terminus domain (residues 364-561) arranges into a typical α/β/α sandwich (Figure 10-A). There are three disordered loops, ranging from residues 147 to 158; 338 to 364; and 517 to 522 (Qi et al. 2010). Three NP subunits assemble in a head-to-tail fashion to form a ring-shaped structure with a three-fold symmetry (Qi et al. 2010). An electrostatic surface potential map of one NP subunit shows a deep cavity at the bottom of the N-terminus domain believed to be the entrance of a cap-binding activity (Figure 10-C); another cavity lies at the top of the C-terminus section where the 3'-5' exoribonuclease active site was thought to be located (Figure 10-D), and a positively-charged groove is present between the two major domains where genomic RNAs were found to bind (Figure 10-C) (Qi et al. 2010).

A later study focusing on the Lassa virus nucleoprotein-ssRNA complex has revealed the structure of the RNA-binding domain and the mechanisms involved in the process (Figure 11) (Hastie et al. 2011). In this study, the recombinant N-terminal domain of LASV NP, that ranges from residue 1 to residue 340 was expressed and irreversibly bound to ssRNA before being subjected to crystallisation trials using the hanging drop vapour diffusion system. The crystals were found to form in the $P6_1$ space group with six molecules present per asymmetric unit. Treatment of NP1-340 with pepsin (NP_{pep}) was also used to remove the flexible 126-143 segment: when subjected to crystallisation, the sample arranged in a $P2_12_12_1$, with one molecule in the asymmetric subunit. Crystals of ssRNA/NP diffracted to 3Å and crystals of NP_{pep} to 1.8 Å resolution. In the RNA-free NP_{pep} structure, attention was drawn to the α5 helix (extended by 10 residues: 112-122 compared to the RNA-containing NP) and the α6 helix which, in the trimeric configuration of LASV NP, were found to lie across the RNA-binding groove, occulting its access. However, when looking at the same region within the RNA-containing $NP_{1\text{-}340,}$ the residues 112-122 were found to be flexible and

disordered, and the helix α5 had its length reduced to terminate before the RNA-binding grove. In the case of the α6 helix, the structure appeared to be rotated away from the pocket, leaving the α5-α6 connection loop (residues 232-243) to form an “open gate”. These results permitted to propose that the trimeric form of LASV NP, with its “closed gate” conformation must undergo structural changes before being able to bind RNA through its gating mechanism.



<u>Figure 11:</u> Comparison between the RNA-free and the RNA-containing N-terminal domains of LASV NP reveals a gating mechanism

In the RNA-free NP, the α5 helix and the α6 helix lie across the RNA-binding groove. In RNA-containing NP, the residues 112-122 are flexible and disordered, and the helix α5 ends before the RNA-binding grove. The α6 helix structure is rotated away from the pocket, leaving the α5-α6 connection loop to form an "open gate" (K. M. Hastie et al. 2011).

Another potential role of NP is putative 3'-5' exonuclease activity. This has been assessed by comparing the ability of wild-type trimeric and hexameric NPs to digest DNA and RNA, which have been found to be able to cleave both types of substrates (Qi et al. 2010). Divalent cations are necessary for optimal endonuclease activity, and the protein shows the following preference for them: $Mn^{2+} > Co^{2+} > Mg^{2+} > Ca^{2+} > Zn^{2+} > Fe^{2+} > Ni^{2+} > Cu^{2+}$(Qi et al. 2010).

An additional zinc ion is coordinated by C506, C529, H509 and E399 despite the absence of any zinc finger motif (Qi et al. 2010). Its location in the vicinity of the exonuclease active and of the putative RNA-binding suggests that it might help stabilise the C-terminal domain of NP, and/or help the process of specific substrate binding (Qi et al. 2010).

Studies on the Old-World prototypic LCMV have uncovered mechanisms by which arenaviruses disable the host innate immune response, effectively contributing to the establishment of persistent infections with the potential of developing into life-threatening haemorrhagic fevers (Martínez-Sobrido et al. 2006). It was shown that infection of Vero cells by LCMV effectively impaired the cells ability to transcriptionally activate the IFN-β promoter in response to DNA transfection or infection by Sendai the virus. This impairment is mediated by the NP protein, which blocks the activation of IRF-3, a transcriptional factor known to induce the production of type I IFN (Hastie et al. 2011b).

Additional studies have revealed that this property is shared among Old-World and New-World arenaviruses, with the exception of Tacaribe virus (Martínez-Sobrido et al. 2007).

Alanine substitution at the five putative catalytic sites of NP: D389, E391, D466, D533 and H528 resulted in the inability to suppress the production of IFN in 293T cells infected with Sendai virus. The abilities tosuppressi the production of IFN induced by the immunostimulatory RNAs poly(I:C),

and to suppress the production of IFN normally triggered by RNAs extracted from Pichinde virions were also found defective upon mutation (Qi et al. 2010). This shows that the ribonuclease activity of NP plays a central role in suppressing the production of IFN induced by viral infection and by the presence of immonustimulatory RNAs.

3.4.3. Glycoprotein complex: Prime target for antiviral therapy

As described above, the glycoprotein complex of Lassa virus is responsible for host-cell attachment and cell-entry, two processes for which correct conformation and correct state of oligomerization of GP-1 and GP-2 are crucial. While the two proteins were originally believed to assemble into hetero-tetramers, more recent studies by Schlie and collaborators have suggested that they might adopt a trimeric organisation (Schlie et al. 2010). This investigation into LASV GP oligomerization state was realized by cross-linking analysis and subsequent immunoblotting performed on sucrose-cushion-purified Vero-cell-produced LASV virions. Separation of cross-linked GP-1 proteins by SDS-PAGE revealed three different species with a calculated mass of 46 kDa for monomers, 92 kDa for dimers and 138 kDa for trimers. The quaternary structure of the GP-2 subunit was further researched by establishing a eukaryotic expression system in *D. melanogaster* Schneider II cells in which the full-length ectodomain of GP-2 was secreted in an N-glycosylated form. Separation by sucrose gradient revealed that the predominant oligomeric state of the protein was trimeric, and heat treatment in the presence of SDS under non-reducing conditions helped establish that no di-sulfide bonds were involved in the oligomerization process.

The X-ray structure of the LCMV GP-2 has recently been solved by (Igonet et al. 2011) in its postfusion conformation (Figure 12). In this study, the part of GP-C ranging from residue 312 to residue 418 and lacking the fusion peptide of GP2 to avoid aggregation was crystallised and the structure of the GP-2 ectomain was determined to 1.8 Å resolution.

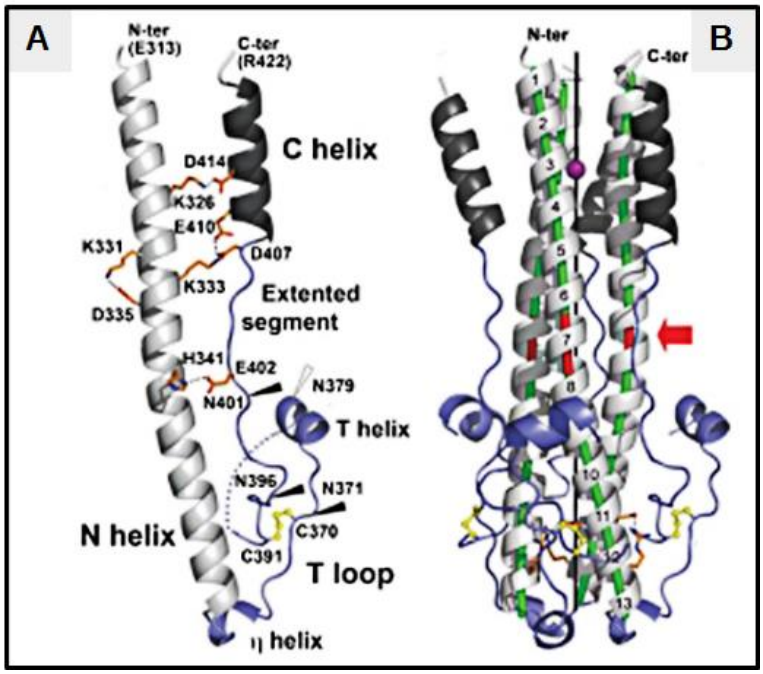


***Figure 12:* Crystal structure of the LCMV GP2 trimer in post-fusion conformation**

(A) Ribbon diagram of the LCMV GP2 protomer. The protein displays a "hairpin" conformation that includes a helix at the start and at the end of the protomer. Here, the N-terminal helix ranged from residue 315 to residue 360 (N-helix); and the C-terminal helix spans between residues 408 and 420 (C-helix). The conserved di-sulfide bond is indicated by a yellow ball-and-stick. Residues involved in salt-bridge are shown in orange (A and B). Full and empty arrowheads point at respectively conserved and non-conserved putative N-glycosylation sites. (B) LCMV GP2 protomers form a trimer. The coiled coil axis is marked by the central black line. The central Cl^- ion is shown by a magenta sphere, green tubes represent the N-helix axes with the "stutter" heptad interruption indicated in red and pointed at by a full red arrow (Igonet et al. 2011).

It was found that the recombinant LCMV GP-2 protein folds and arranges in the form of an elongated trimer about 75 Å in length. The protomer displayed a “hairpin” conformation that is typical of viral fusion proteins in a post-fusion state. This arrangement includes a helix at the start and at the end of each protomer: in LCMV GP-2, the N-terminal helix ranges from residue 315 to residue 360 (N-helix); and the C-terminal helix spans between residues 408 and 420 (C-helix) (Figure 12-A). The C-terminal and N-terminal regions of each protomer were found to be in a configuration where they interact with each other. This ensures that the two membrane-interacting region of the protein are located in close proximity in respect to each other after the occurrence of the fusogenic event. As for description, the GP-2 post fusion trimer can be divided into two halves based on their respective structural arrangements. The top half of the trimer was found to be constituted by a six-helix bundle with a putative central chloride ion, coordinated by a conserved Asn325. The bottom part of the trimer is formed by a trimeric core comprising the C-terminal half of the N-helices arranged in a coiled-coil configuration. This central coiled-coil core was found wrapped in a first inner “layer” composed of an extended polypeptide segment ranging from Cys391 to the C-helix. This layer was also described to be surrounded by a third “flap” outer layer made of a loop, the T loop, extending from the base of the N helix, continuing into a short two-turn alpha “T” helix, and finishing at Cys391. This loop was found to engage in a disulfide bond with Cys370. An interesting feature of the GP-2 coiled-coil region is the presence of a stutter at the seventh turn of the N-helix (Figure 12-B). A stutter is a perturbation of the heptad repeat motif caused by the omission of three residues which locally distorts the packing of hydrophobic layers in the coiled-coil (Geisbert et al. 2005). In the stutter, an additional “*defg*” motif is inserted into two “*abcdefg*” motifs. In the GP-2 coiled-coil, the stutter induces a curvature of the N-helices, resulting in the formation of an internal cavity delimited by two aromatic residues: Phe332 and Phe343. The presence of an x-layer type stutter was subsequently found to occur in the coiled-coil of the

majority of class I and class III viral protein fusions, with the exception of members from the lentivirus clade. It is suggested that this stutter might confer a functional advantage to proteins that allow membrane fusion to occur after changing their own structural conformation, a theory that has yet to be investigated.

Designing antiviral compounds targeting steps of the LASV life cycle is a potential strategy to develop novel therapies against arenavirus-caused haemorrhagic fevers. Viral entry being the first, critical step for viral infection, it therefore appears as a prime target for the development of such new ways to treat arenaviral fevers. Several attempts have already been made to discover molecules displaying the ability of disrupting the process of viral entry (Lupas et al. 1995).

The Old-World prototype virus LCMV has been used to evaluate the impact of amphipathic DNA polymers as a potential anti-arenaviral drug designed to block the first steps of virus infection (Mclay et al. 2013). In this study, the 40-base randomer REP 9 was shown to block cell infection by LCMV with IC_{50} in the low nanomolar range, and to prevent the direct cell-to-cell propagation of the virus. This anti-viral activity appeared to act by targeting the surface glycoprotein complex, and to be dependent on the size and hydrophobicity of the amphipathic polymer, but not on its sequence. This activity also appeared to solely target the early process of infection as no additional impact on the processes of virion assembly and budding was discovered. Further investigation utilising a virus overlay protein binding assay and an ELISA-format solid phase virus-binding assay demonstrated that REP 9 acted by disrupting the binding of GP to α-DG without affecting the structure of either of these proteins.

Similar efforts were engaged towards identifying synthetic combinatorial small molecules that displayed the potential to act as effective inhibitors of arenavirus infection (Lee et al. 2008). Using a cell-based high-throughput screening assay, Lee and collaborators identified three compounds

that, in addition to meeting selective criteria of potency, chemical tractability and drug-like properties, were shown to actively block infections caused by LASV, JUNV and MACV. A subsequent cellular fusion reporter assay was utilized to establish the mode of actions of these selected compounds. In the experiment, permissive Vero cells expressing JUNV or LASV GP were infected by a recombinant vaccinia virus encoding the bacteriophage T7 RNA polymerase. These cells were co-cultured with VeroE6 reporter cells previously infected with a recombinant vaccinia virus expressing the β-galactosidase protein under the control of the T7 promoter. The cells were incubated for 45 min with the candidate compounds before being pulsed with neutral or low pH (5.0) media. β-galactosidase expression, induced upon fusions between the reporter and effector cells, was detected by chemoluminescence. With this, the compounds were shown to block the pH-dependant membrane fusion mediated by arenaviruses to enter their target cell and two of them, named 16G8 and 17C8, showed a broad activity against arenaviruses, with inhibition of JUNV, MACV and GUNV.

In the process of identifying and/or designing therapeutic compounds with the ability to block the crucial entry step of the LASV infection cycle, invaluable information would be derived from the determination and analysis of the GP complex structure, especially, in here, of the LASV GP. However, despite its significant medical importance, LASV is still under-investigated and the molecular mechanisms involved in the infection and assembly processes remain unsolved. Here, this paucity of information will be addressed by using structural biology techniques contributing towards an atomic description of LASV architecture and assembly mechanisms.

4. *Aims of research*

The aim of this study is to perform sub-tomogram averaging on fixed Lassa virus to unravel the structure of the surface envelope glycoprotein complex. Subsequent fitting of X-ray crystallographic structures for GP-1 and GP-2 will provide atomic details of intermolecular interactions during LASV infection and assembly, which will in turn act as platforms for effective drug design.

Material and Methods

5. *Virus growth, purification and inactivation*

LASV growth, purification and inactivation were kindly performed in the laboratory of Dr. Thomas Strecker, University of Marburg, under Biosafety level 4 conditions. Lassa virus (Josiah strain) was propagated in Vero cells and purified using ultracentrifugation and 20% sucrose gradient. The inactivation of the virus was performed using a 4% paraformaldehyde solution.

6. *Cryo-electron microscopy and Sub-tomogram averaging*

Structural determination of viral components provides key information to further the understanding of viral mechanisms and functions by unravelling intermolecular interactions and protein functions. Challenges in studying the structure of a pleiomorphic, non-symmetrical enveloped virus can be bypassed by using a combination of electron cryo-microscopy, sub-tomogram averaging and X-ray crystallography.

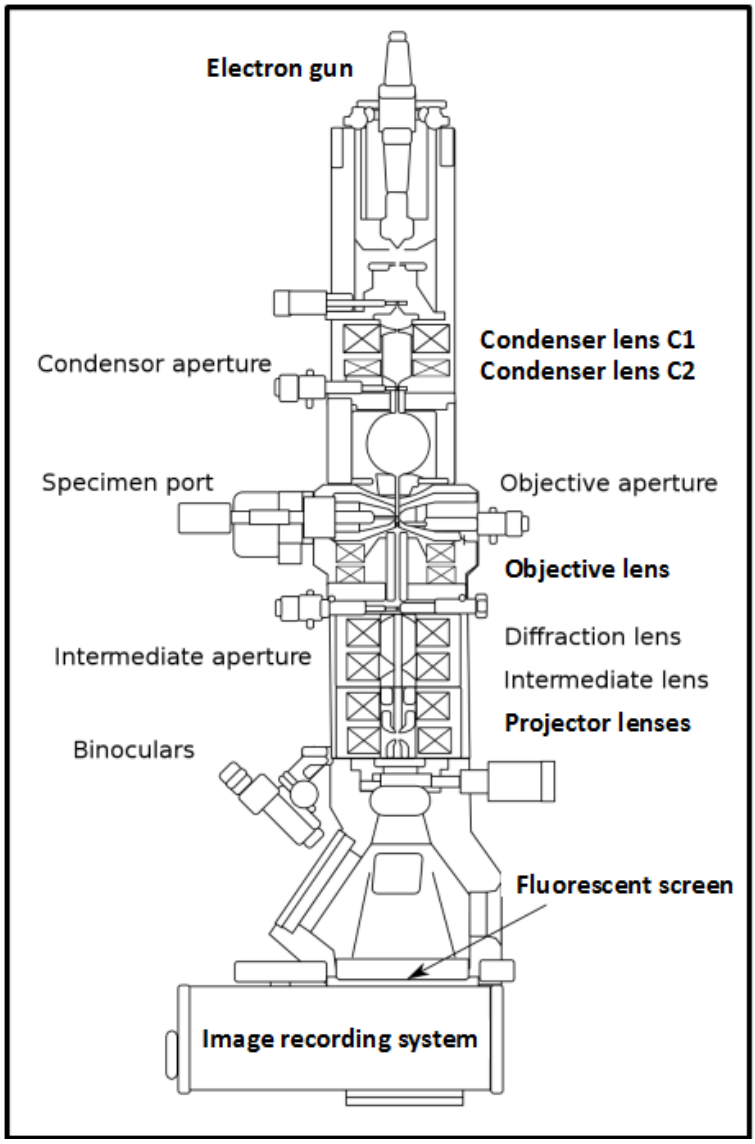
6.1. The Electron microscope: General design

The transmission electron microscope (TEM) utilises the charge and wavelike properties of accelerated electrons to visualize objects too small to seen by light-optical instruments. In a TEM, electrons are emitted by an electron source (eg. a tungsten filament), accelerated to a desired potential difference through high voltage (typically 100-300 kV), and focused using a set of electromagnets (Figure 12). The typical optical configuration for a TEM is composed of three stages of lenses: the condenser lenses, the objective lenses and the projector lenses.

The condenser system controls the electron beam primary state of condensation when it hits the sample to be imaged. It is typically formed by two lenses: C1 and C2. C1 is a strong magnetic lens that determines the “spot size”, while C2 is a weak magnetic lens that allows the diameter of the

illumination to be varied continuously over a wide range and contain a condenser aperture that changes the convergence angle of the illumination.

Located below the specimen stage, the objective lenses produce a magnified image or an electron-diffraction pattern of the object to be investigated. Finally, projector lenses are used to expand the beam and create an image or an electron-diffraction pattern across the TEM screen for direct visualization (Egerton 2005).



<u>Figure 13:</u> Schematic diagram of a Transmission Electron Microscope

Electrons emitted by the electron gun are condensed by the condenser lenses C1 and C2 before hitting the specimen. The objective lens creates a magnified image of the object and the projector lenses expand the beam to be visualized on a fluorescent screen. The TEM is usual couple to an Image recording system such as a CCD camera (www.wikipedia.org).

The utilisation of electrons requires imaging to be performed under high-vacuum conditions to prevent any unwanted interactions with matter from happening. As a consequence, biological specimens have to be prepared to sustain these high-vacuum conditions and be subsequently imaged. One method of fixation involves the rapid plunging of a specimen into a liquid cryogen and the formation of a vitrified layer of ice (Adrian et al. 1984). On this vitrified sample, electron tomography can be performed to unravel the 3D-structure of an object.

6.2. Cryo electron-microscopy: principles and sample preparation

Among the different techniques that have been developed to flash-freeze biological samples in a near-native state, rapid plunge-freezing was demonstrated to preserve macro-molecules in a native, “frozen-hydrated” state with considerably low impacts on atomic sub-structures. This cryo-fixation is achieved by quick freezing the sample in specific cryogens, which allows the water to adopt a non-crystalline vitreous form (Dobro et al. 2010).

The first step in preparing a biological specimen for cryo-electron microscopy is to choose and clean a suitable grid and support film (Dubochet & McDowall 1981; McDowall et al. 1983). The most commonly used support is copper meshed grids coated with a thin carbon film that displays arrays of holes, which in turn can be of various sizes. Molybdenum and gold grids are also available, and choices are made based on toxicity and thermal stability requirements.

Carbon films possess hydrophilic properties when freshly-made but progressively become hydrophobic as time passes. As high-resolution cryo-electron microscopy imaging requires the formation of a uniformly thin layer of vitreous ice embedding the sample of interest, the hydrophilic properties of the carbon film must be restored to allow for the sample to spread evenly over its surface. This is achieved by plasma cleaning, or “glow discharging”, a process in which a gas is ionized under low vacuum conditions. The resulted radicals are then allowed to bombard the surface of the carbon film removing contaminants and restoring hydrophilicity (Dobro et al. 2010).

Cryogens utilized for plunge-freezing are chosen based on their properties in terms of thermal conductivity, heat capacity, freezing point and boiling point (Table 02). More specifically, the formation of amorphous vitreous ice requires cooling rates of a minimum of 10^4 K/s, otherwise crystalline ice will be formed. As a result, alkanes such as propane and ethane are used preferentially over liquid nitrogen, which acts as a coolant to liquefy the two former gazes and

keep them at a required low temperature during the process of plunge-freezing. However, difficulties may arise during the procedure due to the freezing points of ethane and propane being higher than the temperature of liquid nitrogen. To prevent the solidification of the alkane cryogens when in thermal contact with their coolant, a mixture of 37% ethane and 63% propane can be considered (Dubochet et al. 1971).

Cryogen	Thermal Conductivity (W/(m.K))	Freezing point (K)	Boiling point (K)	Heat Capacity ($J.K^{-1}.mol^{-1}$)
Nitrogen	0.024	63.15	77.355	29.124
Ethane	0.018	90.4	184.6	68.5
Propane	0.015	85.5	230.90-231.11	73.60

<u>Table 02:</u> Properties of cryogens commonly used in cryo-electron microscopy

Thanks to the extremely benign nature of structural artefacts resulting from the successful vitrification of a biological sample, cryo-electron microscopy has become a technique of choice to study specimens otherwise difficult or impossible to investigate by other structural biology methods such as X-ray crystallography or nuclear magnetic resonance. Cryo-EM can be further divided into three major techniques for 3D structure determination: cryo-electron crystallography (Tivol et al. 2008), 'single-particle' analysis (Walz & Grigorieff 1998; Saibil 2000), and cryo-electron tomography (Ruprecht & Nield 2001).

6.3. Cryo-electron tomography

6.3.1 Principles of electron tomography

To provide insight into the structure of LASV, the technique of electron microscopic tomography (EMT) in cryo-conditions was used. Electron microscopic tomography permits the generation of a 3D model of the subject of interest by computationally reconstructing this object mass density from a set of 2D projections taken at different tilt angles when rotating the sample around a common tilt axis (Figure 13).

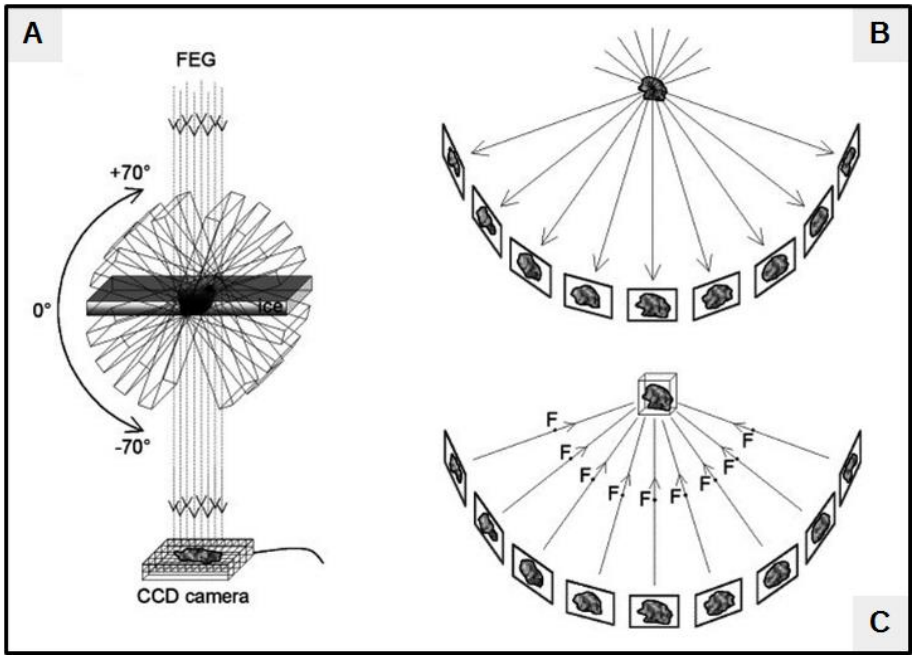
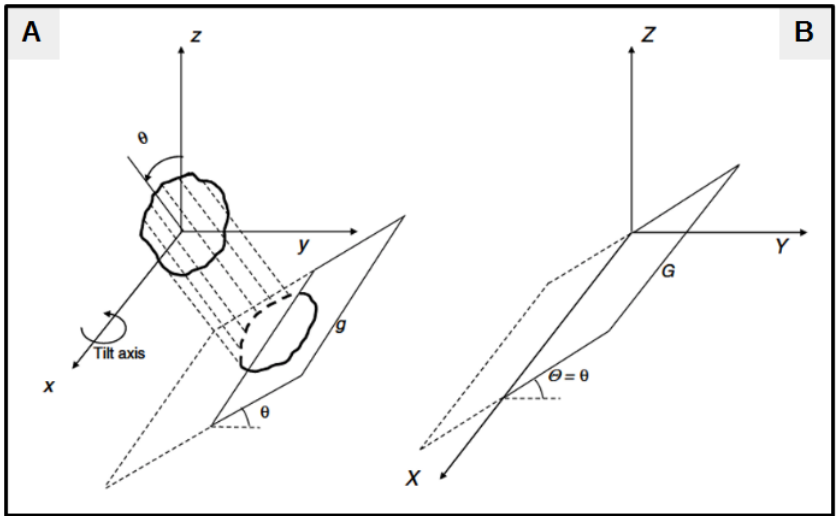


***Figure 14:* Principles of electron tomography**

Electron tomography is a technique based on the acquisition of a set of 2D-projections of a specimen at various tilt angles usually ranging from -70° to +70° (A). This set of 2D density projections (B) is subsequently mutually aligned before being used to reconstruct the specimen object in 3D using a weighted back-projection procedure (C) (Grünewald et al. 2003).

The method relies on the mathematical Central Section Theorem. This theorem states that the 2D Fourier transform of a projection image taken from an object is a central plane through the 3D Fourier transform of this object (Figure 15). In another words, a specific object can be represented in the reciprocal Fourier space using a Fourier transform, which is a breakdown of the object's density distribution into sine functions (Koster et al. 1997; Baumeister et al. 1999; McEwen & Marko 2001; Grünewald et al. 2003). By acquiring 2D projections of an object at different tilt angles, the Fourier Space is sampled and the object is described using the amplitudes and phase shifts of all sine functions recorded. If we describe these sine functions as having wavelengths equal to $d/2$, then the object structure is known down to a resolution of d. To sample the Fourier space in its entirety, the object must be tilted over a range of -90° to +90° around a single axis and images recorded at an angular spacing that is small enough to avoid the loss of information.



<u>Figure 15:</u> Overview of the Central Section Theorem

The 2D Fourier transform of an object's projection image is a central plane through the 3D Fourier transform of this object. Real space coordinates are marked (x,y,z) and reciprocal Fourier space coordinates (X,Y,Z). An object in real space in placed at the origin of a system and projection g of this object is formed at an angle of ϑ of the z axis by integrals along parallel lines that are all orthogonal to the x-axis (A). The Fourier transform of the projection g coincides with the Fourier transform of the original object restricted to a plane that extends through the origin of the reciprocal space system that contains the x-axis and is tilted by ϑ compared to the XY plane (B) (Carazo et al. 1992).

Unlike the single particle reconstruction technique that is based on a collected dataset of conformationally homogeneous randomly-oriented identical particles, a reconstructed tomographic volume contains all orientations of a particular specimen. This property makes EMT particularly suitable for the study of large (100-500nm) structurally heterogeneous samples such as the highly pleiomorphic Lassa virus, at moderate to high resolution (4-20 nm).

Automated data collection for EMT has been made possible by the increase availability of computer power, the introduction of computer-controlled electron microscopes and the development of charged-couple device (CCD) cameras to digitally record EM images. This development of these new technologies considerably eased the collection of tilt-series datasets and made the process operationally simpler and time-efficient (Frank 2006).

6.3.2. Electron microscopic tomography: Data collection

As described above, the strategy of EMT data collection involves tilting the specimen holder in incremental steps around a tilt-axis perpendicular to the electron beam. At each step, a 2D projection image of the sample is recorded and information concerning tilt angle, measured defocus and electron dose are stored.

The typical acquisition of a tilt series spans an angular range of 120 degrees, tilting the sample from -60° to +60°. During the course of acquisition, the calculation of eucentric errors helps minimize translational shifting in the *x*, *y*, and *z* directions. This is particularly important in the *z* direction where shifts in eucentric height translate into significant defocus changes that render the generation of a high-resolution 3D model challenging (Zheng et al. 2010).

To correct for *x* and *y* translational shifts, the automated acquisition of tomographic datasets involves a process of automated tracking where auxiliary images are recorded at each tilt change and cross-correlated to calculate changes taking place in the *xy* plane. On the other end, the autofocusing process is performed by the "beam tilt" method in which the defocus is determined by taking two images of the same region with the electron beam tilted in opposite directions. The method relies on the fact that an image of a specimen in the focal plane of the objective lens does not move when the beam is tilted. Therefore, from calculating the displacement between the two recorded images can be derived the change in focus and the objective lens is then subsequently adjusted to bring the specimen back to the desired focus value (Zheng et al. 2010).

6.3.3. Mechanical Hindrance and Missing Wedge

As described above, the process of electron tomography implies sampling an object's density at different tilt angle to efficiently fill the Fourier space, with a theoretical angular range spanning from -90° to +90° to prevent information loss. However, this tilt range is in fact limited to -70° to + 70°. This is due to the slab geometry of the sample holder, and to the increase path length of the electrons through the sample at high tilts that triggers a high proportion of inelastic and multiple elastic scattering events. This limited tilt range results in the formation of a “missing wedge” in Fourier space: a wedge-like region where information is missing (Figure 16).

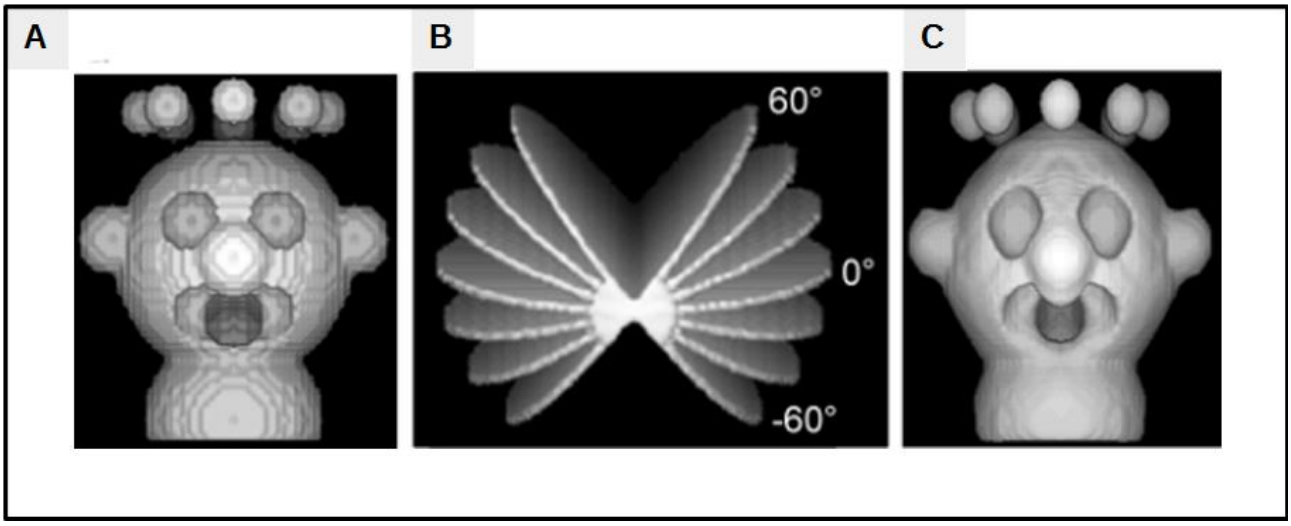


***Figure 16:* Structural distortion resulting from missing wedge data collection artefact**

Exemplification of the impact of the missing wedge loss of information on a structural model reconstructed by electron tomography. (A) presents the original object and (B) a set of planes picturing the principles of electron tomography and the resulting missing wedges of information. (C) Effects of the missing wedge on reconstructed structure: the features are distorted by an elongation factor acting in the z-direction (Orlova & Saibil 2011).

The loss of information translates into Real space by an elongation factor acting in the *z*-direction which distorts reconstructed 3D structures. The elongation factor is given by the equation:

$$\text{sqrt}(\text{sqrt}((\alpha_{max} + \sin\alpha_{max}\cos\alpha_{max})/(\alpha_{max} - \sin\alpha_{max}\cos\alpha_{max}))$$

where α_{max} is the maximum tilt angle (Koster et al. 1987; Koster & de Ruijter 1992). As an example, if α_{max}= 65°,the resulting elongation factor is approximatively 1.5. As a result, 3D models obtained by electron-tomography possess an anisotropic resolution that has to be taken into account when analysing tomographic volumes and performing structure refinement processes.

6.3.4. Low-dose acquisition

Vitrified biological specimens are highly sensitive to beam-induced damage. This leads to a conflict where maximising the number of projections acquired to enhance the quality of the generated 3D model (ie: the resolution achieved) clashes with having to minimize the cumulative electron dose received by the sample. The present method in use consists of a dose fractionation in which a pre-established fixed amount of allowable electron dose is partitioned over the number of tilt images required to achieve a target resolution. In other words, 2D projections imaging has to be performed under low-dose conditions (typically 1-3 e-/Å2 per image) with the electron-dose distributed over the tilt series (Radermacher 1992). As a consequence, electron cryo-tomographic data possess an inherent low signal-to-noise ratio (SNR), in addition to the anisotropic resolution that derives from mechanical hindrance in the microscope when tilting the sample. This vulnerability renders imaging at high tilt difficult as the effective thickness of the sample rapidly increases. With this change in thickness, a higher dose of electrons is required to penetrate through the sample and give a useful contribution to the image, a requirement that is difficult to meet due to the inherent electron sensitivity of cryo-samples.

This is also aggravated by the translational shifts occurring when tilting the sample, a process that requires the spending of additional doses to track *xy* shifts and defocus changes. To take this into account, strategies for automated data acquisition have been developed with additional implements allowing users to define the supporting auto-tracking and focusing areas outside the region of interest, preferentially on the tilt axis (Hong et al. 2014; McEwen et al. 1995). Also, more modern approaches to automated tomographic data acquisition rely on stage calibrations and on predictive tracking implementations. As an example, pre-calibration of the stage can be performed by measuring the movement of the stage in the *x-y* and *z* directions at low magnifications, and extrapolating the results to predict the overall image movements at higher magnifications

(Dierksen et al. 1992; Dierksen et al. 1993; Grimm et al. 1997; Rath et al. 1997; Zheng et al. 2010). However, these fall back to the original technique of cross-correlating auxiliary images if the new position of the imaged sample is predicted with significant statistical errors (Ziese et al. 2002).

In the same way, this inherently low SNR hinders the 3D-reconstruction of an object from low-dose acquired data tomographic data as the process requires the precise alignments of 2D projections images at low and high tilts. To facilitate the process of reconstruction, high-density fiducial markers in the form of 5-10nm in diameter gold beads are added to the sample before plunge-freezing and are used during the iterative stage of alignment.

6.4. Sub-tomogram averaging

Low SNR can be overcome by averaging several hundreds or even thousands of tomographic sub-volumes containing identical particles of interest together in order to generate high-resolution electron density maps.

Referred to as “sub-tomogram averaging”, the method implies the selection of a number of tomographic sub-volumes containing one particle of interest that have then to be aligned together to the same orientation before being subjected to 3D averaging. The process of alignment, and by extension of classification if the particles of interest are known to be conformationally heterogeneous, is particularly challenging due to the high level of noise in low SNR images arising from dose fractionation. This process is also hindered by the presence of the missing wedge in reciprocal space, that is, a loss of information requiring compensation by computerized routines (Mastronarde 2005).

6.4.1. Missing Wedge Correction applied to Sub-tomogram alignment

Most of the techniques developed for the alignment of tomographic sub-volumes are designed to screen rotation angles and translational shifts between sub-volumes using cross-correlation scores as a measure of similarity. Those approaches have been tailored to prevent the “missing wedge bias” in which areas of missing information have the tendency of correlating with each other, leading to erroneous sub-volume alignments (Schmid et al. 2006; Bartesaghi et al. 2008).

The process of “template matching” was introduced by (Bartesaghi et al. 2008) as a way of detecting and identifying macromolecules of interest in reconstructed tomograms in a fully automated manner (Figure 17). The strategy is based on a normalised cross-correlation algorithm that “scans” tomographic sub-boxes to low-pass filtered templates generated from (ideally) high-resolution homologous structures not affected by the missing wedge loss of information. During

the “scan”, the templates are rotated around the three Eulerian angles φ, ψ and ϑ in incremental steps that depend on the predicted order of symmetry of the researched complex. In their study of phantom cells, the calculation of cross-correlation coefficients was constrained to Fourier space regions unaffected by the missing wedge as a way to avoid bias. Alternative approaches have been used in later studies to generate template models including: rotational averaging of particles (Frangakis et al. 2002), the use of an object such as a cylinder (Beck et al. 2004; Zanetti et al. 2006; Zhu et al. 2006), or an arbitrarily-chosen reference in the dataset (Nickell et al. 2007).

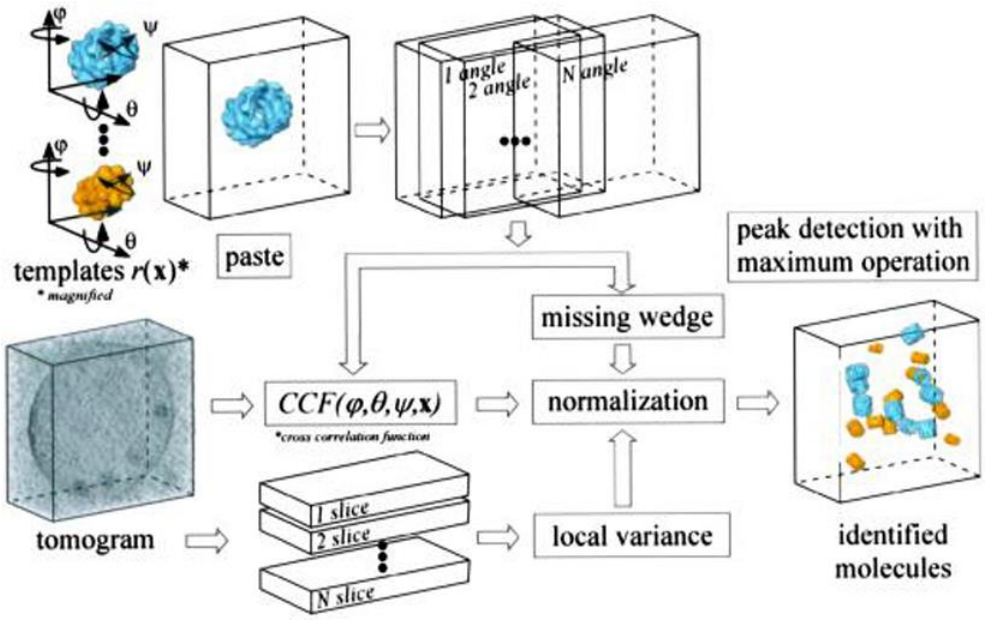
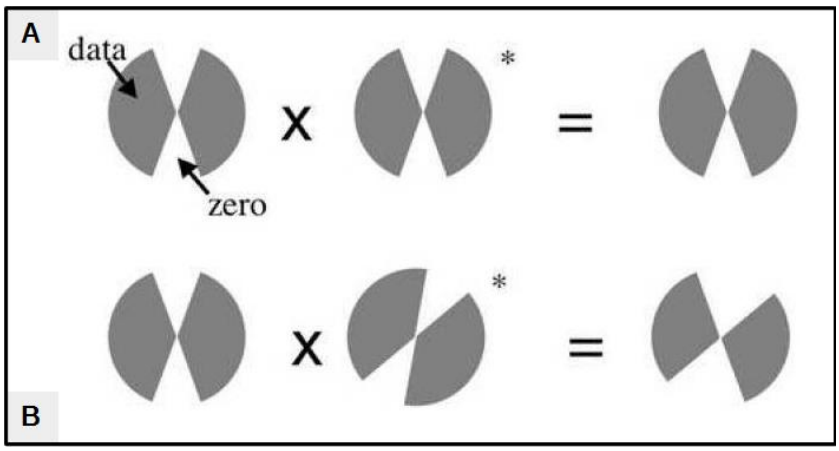


***Figure 17:* Schematic diagram presenting the concept of template-matching**

A fully automated search is performed within a tomographic volume. A parallelized 3D-scanning algorithm (Eg: based on cross-correlation and multivariate statistical analysis) is used to measure the similarity/difference between a template model derived from a high-resolution structure similar to that of interest, and objects within the volume to scan (Frangakis et al. 2002).

While the study cited above focuses on aligning missing-wedge affected sub-volumes to isotropic template references, the most common problem of aligning two volumes both affected by a missing wedge without the use of a pre-existing model was later addressed by (Murphy et al. 2006) (Figure 18). In this study, the alignment problem was corrected by scaling cross-correlation scores with a factor of 1/(volume fraction of non-zero data shared by both volumes). In other words, the amplitudes of cross-correlation peaks at each orientation of two particles with respect to each other were restored by normalising the peak height according to the size of the Fourier space region where information from both sub-volumes is available.



<u>Figure 18:</u> Diagram representing the relationship between the missing wedges of two particles in Fourier space and the effect of alignment on products

If missing wedges are aligned with maximum overlap (A), the product complex displays a missing wedge identical to those of its components. If the missing wedges of two particles are rotated with respect to each other (B), zeros in either of the Fourier trans forms will produce zeros in their product (Schmid & Booth 2008).

The same year, Bartesaghi and collaborators proposed an alternative strategy to by-pass this computationally expensive angular screening using Spherical Harmonics while accounting for the loss of information at high tilts by modelling the missing wedge as a multiplying mask in reciprocal space (Bartesaghi et al. 2008).

Some of these existing techniques were implemented into general electron microscopy and tomography software packages that include IMOD (Bartesaghi et al. 2008), *Bsoft* (Kremer et al. 1996; Nicastro et al. 2006) and XMIPP (Sorzano et al. 2004; Scheres et al. 2008).

6.4.2. Contrast Transfer Function and Sub-tomogram averaging

Biological samples being weak electron scatterers, the primary source of contrast when imaging unstained specimens by cryo-EM is phase contrast. Phase contrast is enhanced by acquiring images in under-focus conditions. In these images, amplitude contrast has been shown to contribute by only 7-14% to the process of image formation (Heymann & Belnap 2007; Heymann et al. 2008). Assessing image quality to aim at high-resolution structural reconstruction requires the understanding of three fundamental notions: the Contrast Transfer Function (CTF) (Toyoshima & Unwin 1988), the effective envelope function, and the background noise (Wade 1992).

6.5. Contrast Transfer Function

This process of image formation in cryo-electron microscopy of thin vitrified biological samples can be described by the phase contrast transfer function, an oscillatory mathematical function of the spatial frequency that modulates the amplitudes and phases of the electron diffraction pattern formed at the back focal plane of the objective lens (Figure 19). This function varies with the applied defocus and the specifications of the microscope, such as the spherical aberration and chromatic aberration functions, and the temporal and spatial coherence of the electron source (Glaeser & Downing 1992; Zhu et al. 1997).

As described before (Frank 2006):

$$\mathrm{CTF}(f)=A(\sin(\pi\lambda f^2(\Delta z-0.5\lambda^2 f^2 c_s))+B\cos(\pi\lambda f^2(\Delta z-0.5\lambda^2 f^2 c_s)))$$

With: - f: the spacial frequency

- Δz: the defocus

- c_s: the spherical aberration

- λ: the electron wavelength

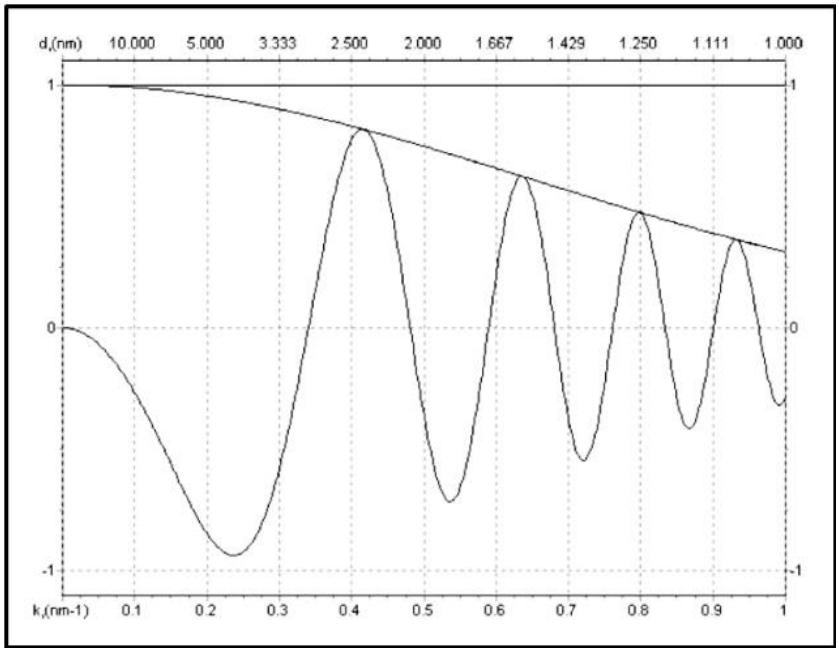
- A: is a defocus-dependent envelope function that describes the decay of the signal for an

instrument with a defined value of angular aperture (Erickson & Klug 1971; Wade 1992)

- *B*: the fraction of amplitude contrast.

It is to note that, in cryo-ET, the CTF is variant across the recorded 2D projections, as tilted specimens exhibit a lateral focus gradient that spatially depends on the distance from the tilt axis (Wade 1978; Zanetti et al. 2009).

The resulting CTF is a sinusoidal function that decays with increases in spatial frequency.



<u>Figure 19:</u> Example of Contrast Transfer Function - FEG 200 kV

Theoretical CTF for a Field Emission Gun microscope operating at 200 kV with a spatial coherence of 0.1 mrad, and an energy width of the incident beam of $\Delta E = 0.7$ eV. A defocus of 3.4 µm was chosen for the first zero to be located at (2.8 nm)-1 (Frank 2006).

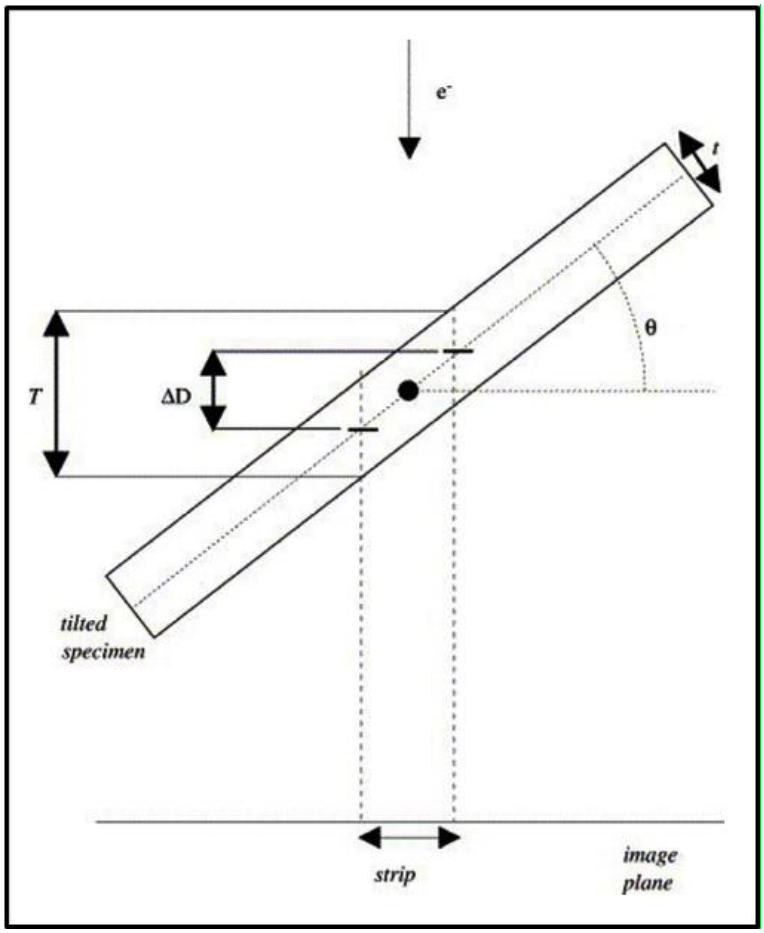
Due to its oscillatory nature, the CTF presents a succession of contrast reversal events separated by spatial frequency values at which the intensity of the signal equals zero. Precise determination of the zero-crossing frequencies (ie: precise determination of the value of the defocus) is essential to restore contrast at high spatial frequencies, and is the key to achieving high resolution. Otherwise, in the absence of correction, the defocus at which a tilt series is recorded is chosen according to the position of the corresponding CTF first zero, which determines the highest resolution achievable.

CTF-correction by phase flipping is rendered very challenging by the inherent low SNR of vitrified biological samples, which hinders the detection of CTF oscillations from the power spectra of 2D projections. In addition to this, mechanical instability leads to changes in the value of the defocus over the course of a tilt series acquisition making it necessary to determine the shape of the CTF for each 2D projection involved in a tomographic reconstruction to perform accurate phase flipping.

6.5.1. CTF determination: Strip-based periodogram averaging

Several methods have been investigated to determine the image defocus in the case of low-dose cryo-electron tomography. In 2006, Fernandez and Crowther proposed a solution based on periodogram averaging (Winkler & Taylor 2003; Fernández et al. 2006). This was later implemented into CTFPLOTTER, a program developed by Xiong and collaborators able to compute 1D power spectra from periodogram and rotational averagings, and to estimate noise backgrounds (Xiong et al. 2009). Periodogram averaging is a technique that subdivides images into tiles from which power spectra are computed from the square magnitude of their discrete Fourier transform (ie: their "periodogram"). These periodograms are averaged together and an estimate of the image power spectrum is derived from this average. Averaging periodograms significantly reduces noise, but a trade-off has to be reached between the number and the size of the tiles computed, that is, a trade-off between the achieved noise reduction and the resolution of the resulting averaged power spectrum (Xiong et al. 2009). From the estimated power spectrum, background subtraction is performed to extract its sinusoidal components, which is achieved by fitting a predicted theoretical curve to local minima (Fernández et al. 1997; Mindell & Grigorieff 2003).

Fernandez and Crowther's approach to CTF determination involves the Welch version of periodogram averaging (Fernández et al. 2006; Mallick et al. 2005; Zhu et al. 1997; Huang et al. 2003; Fernández et al. 1997). Their strip-based method relies on extracting for each image a strip where the CTF is considered similar to that of the untilted plane (Figure 20). The width of this strip is calculated using the equation: $\Theta = \Delta D / w$ with ΔD: the maximum difference in defocus for which the CTF is assumed invariant, Θ: the tilt angle, and w: the strip width. From this strip, overlapping tiles are computed with an optimal overlap of half the tile size. Power spectra are computed for each tile and averaged, before subjecting the result to rotational averaging.



<u>Figure 20:</u> Computation of a strip for CTF-determination in tilted specimens

To determine the parameters of the CTF, a strip is extracted from each image to be subjected to correction. The strip is chosen around the tilt axis, running parallel to it, with a width based on the equation Θ=ΔD/w. From this strip, overlapping tiles are computed with an optimal overlap of half the tile size. Power spectra are computed for each tile and averaged, before subjecting the result to rotational averaging (Fernández et al. 2006).

6.5.2. CTF determination: Background subtraction and Defocus determination

Determination of the CTF in 2D projections acquired under low-dose conditions requires a step of "background subtraction", where background noise is subtracted from the particle signal. Background noise was originally modelled as an exponential function of frequency. In their presentation of CTFPLOTTER, Xiong and collaborators estimated that, due to the low electron count per image, most of the background noise can be attributed to quantum noise, an assumption on which they derived their noise floor subtraction method (Fernández et al. 2006). The method is based on recording quantum noise images over a carbon film support without specimen at low (0.1 $e^-/Å^2$) to high (10 $e^-/Å^2$) electron doses to obtain their 1D power spectra from peridogram and rotational averaging. The background noise power spectrum for a particular set of projections is then interpolated from the power spectra of two images recorded at a dose close to that utilized for data acquisition.

CTF fitting and defocus determination rely on maximizing the cross-correlation between a generated theoretical CTF curve and an image's power spectrum obtained after background subtraction with the primary goal of locating the first zero-crossing frequency with accuracy.

6.5.3. CTF correction methods for tilted specimens

CTF correction performed in the context of electron tomography has been geared to tackle the defocus gradient that affects images recorded at a tilt angle. To fulfil this goal, two types of methods prevail: a localized line-by-line strategy (Xiong et al. 2009), and a global approach (Fernández et al. 2006).

Fernández, Li and Crowther described in their 2006 study a CTF correction approach exploiting a concept of "strip" similar to the one previously described for CTF determination. Any strip along the *x*-line of the image can be considered for the purpose of CTF correction instead of the previous restriction to areas surrounding the tilt axis. For this strip, the defocus is computed by the equation:

$$D(x) = D_0 + d(x) \tan \vartheta$$

with :

- *D(x): the defocus applied to pixels located along the x-line*
- *D_0 : the estimated defocus at the untilted plane*
- *x: the coordinate along the X-axis*
- *d(x): the distance to the tilt axis*
- *ϑ: the tilt angle*

With this method, the image CTF correction is broken down to a series of local corrections performed by phase flipping, each correction constrained to one area of the image that displays a spatially invariant CTF. Computationally tasking, the process can be made faster by dividing the projection into overlapping strips from which Fast Fourier Transforms are computed (Winkler & Taylor 2003; Philippsen et al. 2006).

Assuming a thin specimen where the relationship between the image contrast and projected density is linear, Winkler and Taylor designed a defocus gradient determination method in which

2D projections are divided into small patches. For each patch along a line parallel to the tilt axis, the squared magnitudes of their Fourier Transforms were added together and a 1D power spectrum for each row was obtained from rotational averaging to calculate the defocus. CTF correction was then performed by solving a matrix of equations.

Finally, Philippsen and collaborators described in 2006 an analytical expression to simulate image formations under tilted conditions for thin specimens. This expression was proposed as a linear transform in the frequency domain, and deconvolution of the CTF subsequently performed by solving a matrix of linear equations.

7. *Jsubtomo: A new tool for sub-tomogram averaging*

7.1. General overview of Jsubtomo

Jsubtomo is a computational package developed in the Oxford Particle Imaging Centre, Division of Structural Biology, University of Oxford by Dr. Juha Huiskonen. Based on the *Bsoft* package (Heymann & Belnap 2007), and implemented in the C and Python programming languages, Jsubtomo is designed to detect, align and average volumetric data extracted from tomographic reconstructions. Allowing for the study of irregular pleiomorphic entities, the programme has been successfully used in a number of projects aiming to unravel viral architectures (Arranz et al. 2012; Bowden et al. 2013; Huiskonen et al. 2010; Liljeroos et al. 2011; Maurer et al. 2013; Pietilä et al. 2012).

Jsubtomo is composed of three different programmes named *jsubtomo*, *jview* and *jave*.

- ***jsubtomo*** is used to perform template matching inside reconstructed tomographic volumes to detect particles of interest; to select these particles based on a set constrained cross-correlation cut-off value; and to refine the location and orientations of these particles using an iterative procedure.

- ***jview*** is used to incorporate *a priori* information concerning the particles of interest's locations and orientation to facilitate their coordinates refinement. This *a priori* set of information may be estimated from the geometry of the studied entity: for example, the orientations of glycoproteins present on the surface of a spherical virus may be estimated from the spherical geometry of the virus by determining the centre of the virus volume.

- ***jave*** is the programme used to average particles previously extracted from tomographic subvolumes.

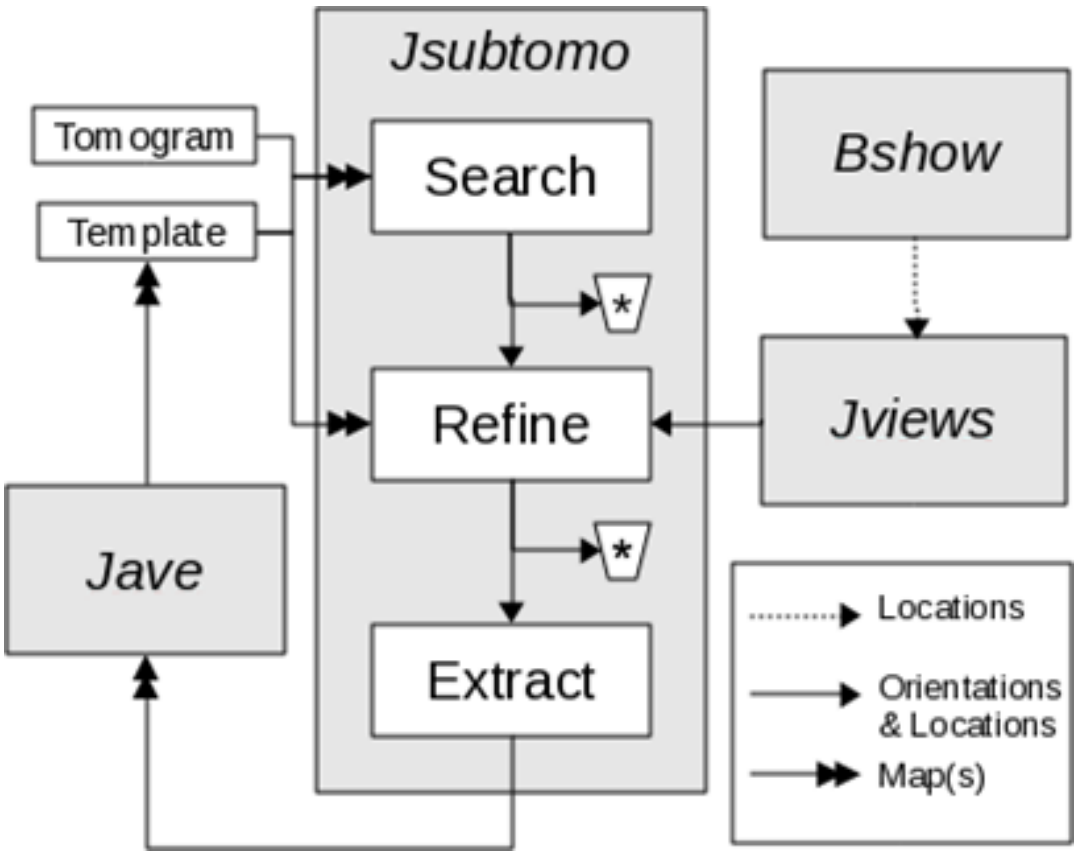


Figure 21: ***Jsubtomo workflow***

The Jsubtomo package is designed to detect, align and average volumetric data extracted from tomographic reconstructions. Composed of Jsubtomo, Jave and Jviews, the package allows for iterative refinements of electron-microscopy determined structures (Courtesy of Dr. Juha Huiskonen).

7.1.1. Jsubtomo: Implementation for the search, refinement and extraction of particles

The *jsubtomo* programme comprises three different modes of action: Search, Refine and Extract.

- **Search** is used to find particles in an input image. This mode implies the utilization of a pre-defined template that can be generated by either filtering high-resolution homologous structures obtained by X-ray crystallography, or by the rotational averaging of manually selected identical particles. The Search mode calculates a cross-correlate map between the template and the tomographic volume to search. For this purpose, rotational and translational increments of search are set and the loss of information resulting from the missing wedge artefact can be weighted for by including in the process a mask restricting calculations to parts of the Fourier Space that were sampled. A cross-correlation (CCC) filter can be added to the search to select the best correlation hits defined by a CCC threshold, and a distance filter can also be used to remove overlapping hits or hits that are defined as too close to each other in space.

- **Refine** performs the local refinement steps to determine the location and rotation of selected particles. This is achieved by performing a cross-correlation search around a grid of neighbouring views defined by an angular sampling, angular limits, and a maximal shift limit. At each successive iteration, the angular step size and range are halved and the best correlating rotation and shift are saved in an output .star file. This is achieved using cross-correlation and template-matching.

- **Extract** is used to extract sub-volumes from their initial tomographic reconstruction. During this step, a transform option can be added to bring the extracted particle to the reference standard view of the template to facilitate the subsequent step of particle averaging.

7.1.1.1. Creation strategy of a starting model for sub-tomogram averaging

The lack of pre-existing complete homologous glycoprotein structures for LASV GP complex restricted the possibilities for the creation of a starting model from which to perform the initial iterative refinement steps. In this study, an initial template was created by manually picking a small set of glycoprotein particles. The lack of alignment around the alpha angle, the angle around the vertical axis of the glycoprotein spike, results in a blurring of the starting model structure. However, the presence of noise in the tomographic volumes utilized for this step results in the appearance of random features in the average that will help the refinement process to converge (White et al. 2010; Zanetti et al. 2006).

7.1.1.2. Jview: Definition of initial particle views from a priori knowledge

A priori knowledge on the location and orientation of particles can be derived from associated biological geometry if these particles are not randomly located nor randomly oriented within a tomographic volume. In other words, the shape of a biological object can be used as a reference to define initial view vectors associated with related particles of interest as it restricts their locations and orientations in space.

As an example, this is the case for glycoproteins located on the surface of a viral particle. In the case of a spherical enveloped virion, the glycoproteins' locations and orientations can be easily estimated providing the virus's radius is known: They will be expected to sit on the outer leaflet of the envelope and extend radially from its surface. After defining initial seeds that evenly cover the surface of the virion, the angular and translational search to refine the particle's starting locations and orientations are restricted to comply with the virus geometry.

In the investigation of the LASV glycoprotein's structure, the highly pleiomorphic shape of Lassa virions proved a hindrance to the application of this technique. As a consequence, glycoprotein particles were manually picked in boxed sub-tomographic volumes. However, the particles' orientations could still be estimated from nearly-spherical envelopes, which eased the angular refinement. From the spherical-shape of the virus, two Euler angles theta and phi could be estimated, while the alpha remained at this stage unsolved.

7.1.1.3. Jave: sub-tomogram averaging and missing wedge

When performing the task of sub-volume averaging, a reciprocal space missing wedge mask can be added to the program parameters to account for the loss of information that results from single-axis tomography data collection.

7.1.2. Procedure for LASV glycoprotein sub-tomogram averaging using Jsubtomo

For LASV surface complex GP refinement, the following procedure was established:

A) Extraction of sub-volumes that contain LASV particles from full-size tomograms

A.1) Manual picking of virus particles in tomographic volumes

A.1.1) The procedure for subtomogram averaging of LASV surface glycoprotein starts with the manual picking of LASV virions. To do so, a tomogram file containing LASV particles is first opened in the programme *bshow*. To define the center coordinate of a virus particle and by doing so select the particle for future use, the particle-picking tool is utilised. The viral particle is selected by defi-ing the volumetric centre point of the virus, in other words, the centre of the viral volume in x,y, and z. This centre point of the virus volume will be used to estimate the glycoproteins' orientation from the viral particles' geometry. This first selection step is then repeated until all virions have been processed. The viruses' coordinates are saved in a STAR file (Hall 1991).

A.1.2) The previous step is repeated until all viral particles located in all collected tomograms have been processed. At this point, the virions' diameters are recorded, in pixels, to be used in the ex-traction step.

A.2) Extraction of selected virions

To extract the selected virions, *jsubtomo.py* is run in the **extract mode** (parameter "--mode ex-tract"). This mode will output the virion sub-volumes into individual volume files using the STAR files saved in step A.1.2 as input files.

A.2.1) To extract all relevant information from the original tomographic volumes, virions boxes to be output are given a size (parameter "--size") of approximately 25% larger than the largest viri-

on's size in the data set. Here, the virions are approximately 200 pixels in diameter, so a size of 250 x 250 x 250 pixels is used.

A.2.2) The output files of the virions' sub-volumes are in the MAP-format. At this stage, accompanying STAR-files are created for each virion sub-volume extracted (parameter "--output").

A.3) Data floatation

Last of the LASV sub-volumes extraction procedure is a data floatation step using *bimg*, a *Bsoft* programme for image manipulation. *Bimg* is used to normalize the virion sub-volumes so that their average value is 0 and one standard deviation corresponds to value 1. The datatype of the voxels needs to be "float" to normalize all operations to be performed on the sub-volumes. For instance, multiplication with real-space and reciprocal space-masks, and data averaging of unnormalised data parts would lead to differential selection of information, aberrant differences in pixel values and uneven contribution towards the final averaged output. Moreover, the glycoproteins should appear white against dark background as per convention in electron microscopy.

B. Generation of two independent initial models

B.1) Manual picking of a set of spikes in the virion sub-volumes

B.1.1) The next step in the process is the manual picking of a set of spikes from the previously extracted virion sub-volumes. To proceed, a virion sub-volume MAP file is opened in *bshow*. The center coordinate of a spike is defined using the particle-picking tool and this step is repeated until all clearly distinct spikes have been selected. As above, the spikes' coordinates are saved in a STAR file. (Note: In here, to generate initial starting models, only approximately 200 spikes are necessary. This figure will vary according to the state of the signal-to-noise ratio of the recorded tomograms and the general structure of the investigated protein).

B.1.2) The spike dimensions in pixels are measured for future use.

B.2) Assignment of view vectors

This step focuses on assigning initial view vectors to the spikes by running the *jviews.py* script. These view vectors approximate the direction of the spikes based on the spherical geometry of the virion. To start the assignment process with a first approximation of the spikes locations, the STAR files generated in step B.1.1 are used as input files.

B.2.1) To assign the vector views, the central coordinate of the virion are used in addition to its spherical geometry properties. Here, the virion is centered (after step A.2) in a box with a size of 800 x 800 x 800 pixels, the central coordinate is thus 400,400,400 pixels (option "--Radial 400,400,400").

Assign view vectors: jviews_par.py --Radial 400,400,400 --shift -30 --output --suffix _init *part???.star

B.3) Generation of a real-space mask

This step generates a real space mask using *beditimg*. In this mask, values of 1 denote regions inside of the mask and values of 0 denote regions outside of the mask. This mask is designed to be large enough to contain both the spike and a patch of the underlying membrane.

B.3.1) The mask file is generated with the same size as the expected averaged structure: 120x120x120 pixels.

B.3.2) To be utilized in the generation of initial starting models that in turn will be used for structure refinement, this real-space mask needs to contain a large spherical mask of a size only slightly smaller than the size of the averaged structure.

B.3.3) To avoid introducing spurious correlation between the model and the data, a Gaussian-shaped fall-off is used when generating the mask. In here, a sigma-width of 4 pixels was chosen.

Real space mask: beditimg -create 120,120,120 -sphere 60,60,60,48 -edgewidth 4 -fill 1 mask.map

B.4) Generation of a missing-wedge mask

A reciprocal space mask is generated using *jsubtomo_create_wedgemask.py* to exclude the 'missing wedge' region that results from mechanical hindrance and incomplete sampling of the Fourier space during data collection. As before, values of 1 denote regions inside of the mask and values of 0 denote regions outside of the mask. The mask file should be of the same dimensions as will be used for the averaged structure or larger, in here at least 120 pixels.

Reciproqual space mask: jsubtomo_create_wedgemask.py --verbose 7 --size 120 --wedge 90,-60,60 wedge.spi

B.5) Generation of selection files

A selection file (SEL file) is generated to define which virion particles belong to a set "1", and which virion particles belong to a set "2". The general refinement process is indeed performed in two independent sub-datasets and in a parallelized fashion to assess the validity of the resulting refined structures and calculate the attained resolution.

This selection file can be generated either in a generic text editor or by using commands ls and awk. Each line should contain the name of an input STAR file generated in step B.2, followed by space and number 1 or 2, each referring to either independent set.

B.6) Sub-set averaging and initial models

In this step, an average is independently generated from each sub-set of particles, 1 and 2. To perform this task, the *jsubtomo_create_averages.py* script is used with the SEL file generated at step B.5 as input file.

B.6.1) A size (parameter "--size") at least 32 pixels larger than the largest dimension of the spike is given to the area of information to be averaged into initial models. Here, the glycoprotein spike is approximatively 40 pixels long, so a box size of 72 x 72 x 72 pixels is used.

B.6.2) A high point group symmetry is applied on the averages (parameter --symmetry c100"). This high point symmetry is used to approximate a cylindrical averaging around the vertical axis of the spike, and in addition helps with noise reduction.

B.6.3) To name the output files of the project the option (parameter –suffix") can be used.

B.6.4) The real-space mask generated in step B.3 is used to mask away potential surrounding spikes/structures that would interfere with the refinement process (parameter "--mask"). In addition, the reciprocal-space missing-wedge mask generated in step B.4 is used to account for the loss of information resulting from the mechanical hindrance within the microscope and the resulting anisotropic sampling of the Fourier space (parameter "--Mask").

Generation of two averages: jsubtomo_create_averages.py --verbose 7 --sampling 3.1 --origin 60,60,60 --symmetry c100 --size 120,120,120 --prefix aves --mask mask.map --Mask wedge.spi --proc 24 lv_init.sel

Creation of: *aves_average.map*

aves_average_even.map

aves_average_odd.map

C. Iterative alignment and averaging of the two initial spike models

The iterative alignment of spike densities to the two initial models, and their subsequent averaging are carried out by the script *jsubtomo_iterate_gold.py*.

C.1) Refinement of the view vectors' direction: Determination of ϑ and φ

In this stage, the theta (θ) and phi (ϕ) angles defining the direction of spikes' view vectors are being refined. To do so, the SEL file generated at step B.5 is used as input file.

C.1.1) The refinement process is started using an angular sampling of 8 degrees (parameter "--angles 8,8,8").

C.1.2) 16-degree changes in the direction of the spike view vector (parameter"--thetaphilimit 16") are being allowed to refine the spikes' orientation within their respective tomographic subvolumes. In this stage, the alpha angle, the angle around the vertical axis of the spike, is kept fixed (parameter "--alphalimit 0"). The refinement of the alpha angle and hence the determination of the spikes' orientation around their vertical axis will be addressed once the theta and phi angles have been solved.

C.1.3) Small translational shifts are allowed to happen during the alignment step (here: 5 pixels). This is introduced to account for inaccuracies in the manual picking of the spikes (parameter "--shiftlimit 5").

C.1.4) Once again, a high point group symmetry is applied on the averages (parameter "--symmetry c100"). As described above, the symmetrization process helps reduce noise in the average structure.

C.1.5) In this step, the real space mask and the reciprocal space mask generated in steps B.3 and B.4 are used respectively to mask away surrounding spikes and to account for the missing wedge (parameters “--mask” to define the real-space mask and “--Mask” to define the reciproqual-space mask).

C.1.6) The two independent templates used in this step are the two averages generated in step 2.6 and output as MAP files. They are identified by the tags “even” and “odd” in their filename. To utilize these two MAP files as templates in the refinement process, the parameters “--Template1” and “--Template2”, are used respectively.

C.1.7) To ensure an accurate centering of the structures, a low pass filter at 30-Å resolution (parameter “--resolution”) is incorporated in the process.

C.1.8) To speed up the refinement, a bin factor can be used. When refining the structure of LASV GP, a bin factor of two was applied (parameter “--bin 2”).

C.1.9) This step of refinement of the theta and phi angles by iterative alignment, and subsequent averaging, is run through 10 iterations (parameters “--firstiter 1 --lastiter 10”). The changes in the spikes' angles and locations are monitored over the course of the iterative process so that it can be stopped when no significant changes are observed any further.

Refine direction of view vectors: jsubtomo_iterate_gold.py --verbose 7 --firstiter 1 --lastiter 10 --resolution 30.0 --mask mask.map --Template1 aves_average_even.map --Template2 aves_average_odd.map --Mask wedge.spi --origin 60,60,60 --angles 8,8,8 --symmetry C360 --alphalimit 0 --thetaphilimit 16 --size 120,120,120 --sampling 3.1 --shiftlimit 15 --prefix lv --proc 8 --bin 2 lv_init.sel

C.1.10) Before starting the refinement of the last alpha angle, the SEL file must be updated: the filenames will now refer to the ones generated in the last iteration in step C.1.9 and will thus take into account the refined positions and directions of the GP spikes. However, the assignment into groups “1” and “2” must remain the same.

C.2) Refinement of alpha: orientation around the spikes' vertical axis

The goal of this stage is to refine the angle around the view vector, the alpha angle, hence determining the orientation of the spike along its vertical axis and unravelling the symmetry of the complex. In this step, the input file is the SEL file generated at step C.1.10.

C.2.1) Firstly, the accurate dimensions of the spike need to be measured. To do so, the low-pass filtered MAP file generated in the last iteration in step C.1 (identified by the tag "_lp" in the filename) is opened in *bshow*. A new real space mask is then generated similarly to step B.3. This new real-space is designed to include the spike but exclude the membrane and any neighboring background, selecting only the ectodomain units of the GP complex protruding outwards from the viral membrane

Dimensions of the spike: 33x35 pixels which equates to 10.23 x 10.85 nm

Real-space mask: beditimg -create 120,120,120 -sphere 60,60,80,28 -edge 3 -fill 1 mask2.map

C.2.2) To start the refinement of the alpha angle, an 8-degree angular sampling (parameter "--angles 8,8,8") is used, as for the previous refinement of θ and ϕ.

C.2.3) This time, a 180-degree change in the angle around the spike view vector (parameter "--alphalimit 180") is allowed but the theta and phi angles are kept fixed (parameter "--thetaphilimit 0"): this step will scan the orientations around the vertical axis of the spike and therefore any other parameter must remain unchanged.

C.2.4) As a consequence, no translational shifts are allowed to happen during the refinement of the spikes' orientation around their vertical axis (parameter "--shiftlimit 0").

C.2.5) Similarly, no symmetrisation on the averages is carried out at this point (omit parameter "--symmetry").

C.2.6) To account for the missing wedge, the reciprocal space mask generated in step 2.4 is used.

C.2.7) At this stage, the two MAP files generated in the last iteration of step 3.1 without symmetry (identified by the tags "even_nosym" and "odd_nosym" in their filename) are used as templates for the iterative process (parameters "--Template1" and "--Template2", respectively).

C.2.8) A low pass filter at 50-Å resolution (parameter "--resolution") is applied to the first iteration to keep from biasing the refinement process. The filter parameters are then adjusted in the subsequent iterations in an automatic fashion by using the option (parameter "--adaptivefilter"). This automation is based on the gold-standard Fourier Shell Correlation (FSC) between the two independent averages. In this study, the 0.143 criterion (parameter "--fsccrit 0.143") was used to set the FSC correlation cut-off that defines the resolution achieved in the refined averages. In order to prevent the process from over-refining the average and introducing potential errors, the refinement procedure is not carried out past the resolution defined by the first zero of the Constrast Transfer Function (CTF) (parameter "--maxhires"), unless CTF correction has been previously applied to the tomograms (Refer to Chapter 3).

C.2.9) As before, 10 iterations of the alpha angle-focused sequence of alignment and subsequent averaging of spike densities are let to run (parameters "--firstiter 11 --lastiter 20"). The changes in the spike angles and locations are monitored in the same way as in 3.1.9. When no significant changes are observed, the process can be stopped.

Refine angle around view vectors: jsubtomo_iterate_gold.py --verbose 7 --firstiter 11 --lastiter 20 --resolution 50.0 --adaptivefilter --fsccrit 0.143 --minhires 38 --hiresfact 1.2 --lowresfact 10 --mask mask2.map --Template1 lv_iter10_average_even_nosym.map --Template2 lv_iter10_average_odd_nosym.map --Mask wedge.spi --origin 60,60,60 --angles 8,8,8 --alphalimit 180 --thetaphilimit 0 --size 120,120,120 --sampling 3.1 --shiftlimit 0 --prefix lv --proc 8 --bin 2 lv_init10.sel

C.3) Symmetry assessment

The degree of symmetry in the structure can be assessed by examining the low-pass filtered MAP file generated in the last iteration in step 3.2 (identified by the tag "_lp" in the filename) in *bshow* or *chimera*.

C.4) Accurate determination of alpha

This stage focuses of the accurate refinement of the alpha angle around the view vector of the spikes by use of symmetry. To do so, the parameters used in the process are the same as in step C.2 except:

C.4.1) This step of accurate refinement of alpha is made by allowing appropriate changes in the angle around the spike view vector. In here, the glycoprotein spike is expected to display a 3-fold symmetry, so changes of 60 degrees in the alpha angle (parameter"--alphalimit 60") were allowed. Since focusing on the alpha angle only, the theta and phi angles were kept fixed (parameter "--thetaphilimit 0").

C.4.2) During the process, the established correct symmetry is applied to the averages (parameter --symmetry c3").

C.4.3) The two MAP files generated in the last iteration in step 3.2 are used as input template files (parameters “--Template1” and --Template2).

C.4.4) A low pass filter is applied at the resolution of the last iteration from step 3.2 (parameter “--resolution”) in the first iteration. The filter parameters are then adjusted in the subsequent iterations automatically (parameter “--adaptivefilter”) based on gold-standard FSC between the two independent averages. As before, the criterion of 0.143 is assigned to the FSC (parameter “--fsccrit”).

C.4.5) This time, 5 iterations of alignment and averaging are run (parameters “--firstiter 21 --lastiter 25”). At the same time, the changes in the spike angles and locations are monitored. As before, when no significant changes are observed, the process can be stopped.

Accurate determination: jsubtomo_iterate_gold.py --verbose 7 --firstiter 21 --lastiter 25 --resolution 40.0 --adaptivefilter --fsccrit 0.143 --minhires 38 --hiresfact 1.2 --lowresfact 10 --mask mask2.map --Template1 lv_iter20_average_even.map --Template2 lv_iter20_average_odd.map --Mask wedge.spi --origin 60,60,60 --angles 3,3,3 --alphalimit 60 --symmetry c3 --thetaphilimit 0 --size 120,120,120 --sampling 3.1 --shiftlimit 0 --prefix lv --proc 8 --bin 2 lv_init20.sel

C.4.6) Before proceeding to the last sequence of complete refinement of the GP spikes parameters, the input SEL file is updated as to refer to the filenames generated in the last iteration of step 3.4.5. The assignment into groups “1” and “2” must stay the same.

C.5) Complete refinement of spikes' parameters

The goal of this stage is to accurately refine all three angles simultaneously. The input file is the SEL file generated at step 3.4.6. The parameters are the same as in step C.4 except that:

C.5.1) A fine angular sampling (4-degrees) (parameter "--angles 4,4,4") is used.

C.5.2) A limit of 8 degrees of change is set for the alpha angle, the theta and the phi angles (parameters "-- alphalimit 8" and "--thetaphilimit 8").

C.5.3) Only small shifts are allowed to account for inaccuracies in previous alignment steps (parameter "--shiftlimit 5") (in pixels).

C.5.4) The two MAP files generated in the last iteration in step 3.4 are used as input template files (parameters "--Template1" and --Template2).

C.5.5) A low pass filter is applied at the resolution of the last iteration from step 3.4 (parameter "--resolution") in the first iteration. The filter parameters are adjusted in the subsequent iterations automatically (parameter "--adaptivefilter") based on gold-standard FSC between the two independent averages. The same criterion as C.2.8 is applied to the FSC (parameter "--fsccrit").

C.5.6) 10 iterations of alignment and averaging are let to run(parameters "--firstiter 26 --lastiter 35"). During these 10 runs of iteration, the changes in the spike angles and locations are monitored. When no significant changes are observed, the process is stopped.

Refine angle around view vectors: jsubtomo_iterate_gold.py --verbose 7 --firstiter 26 --lastiter 35 --resolution 40.0 --adaptivefilter --fsccrit 0.143 --minhires 38 --hiresfact 1.2 --lowresfact 10 --mask mask2.map --Template1 lv_iter25_average_even.map --Template2 lv_iter25_average_odd.map --

Mask wedge.spi --origin 60,60,60 --angles 2,2,2 --alphalimit 8 --thetaphilimit 8 --size 120,120,120 --sampling 3.1 --shiftlimit 5 --prefix lv --proc 8 --bin 2 lv_init25.sel

8. New- and Old-World Arenavirus Glycoprotein Expression and Crystallization Trials

8.1. Cloning strategy for the expression of Arenaviruses GP proteins

To complement information obtained using cryo-electron tomography and sub-tomogram averaging, the expression and purification of Old-World and New-World Arenavirus glycoproteins with the intention of determining their atomic structures using X-ray crystallography were needed.

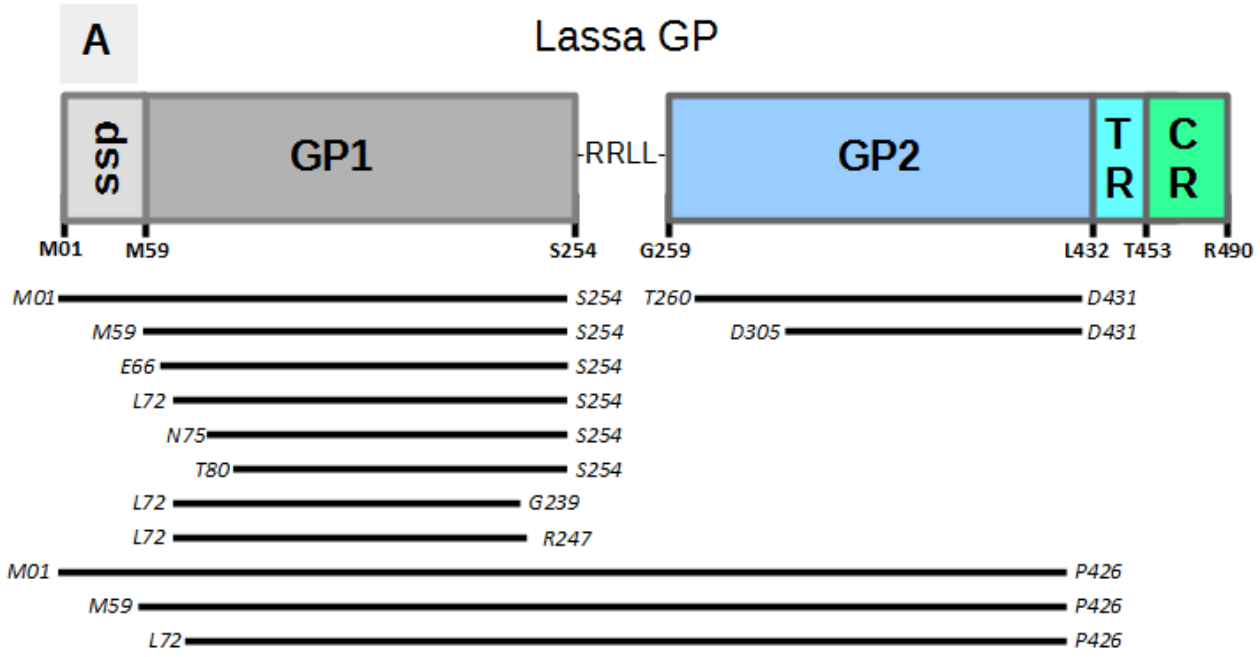
The ultimate goal of protein expression intended for crystallisation trials is the production of a homogeneous sample of correctly-fold secreted glycoprotein in regards to oligomeric and glycosylated state. Homogeneity in the protein sample leads to the production of homogeneous and well-ordered protein crystals, a pre-requisite for the generation of high resolution diffraction when crystals are exposed to monochromatic X-rays. Arenavirus glycoproteins being subjected to the oxidative extracellular environment, two cysteine residues adjacent in a three-dimensional structure are likely to oxidise into a disulfide bride. Intramolecular disulfide bridges stabilise the folding of the polypeptide chain while interchains disulfide bonds are likely to be involved in protein oligomerisation.

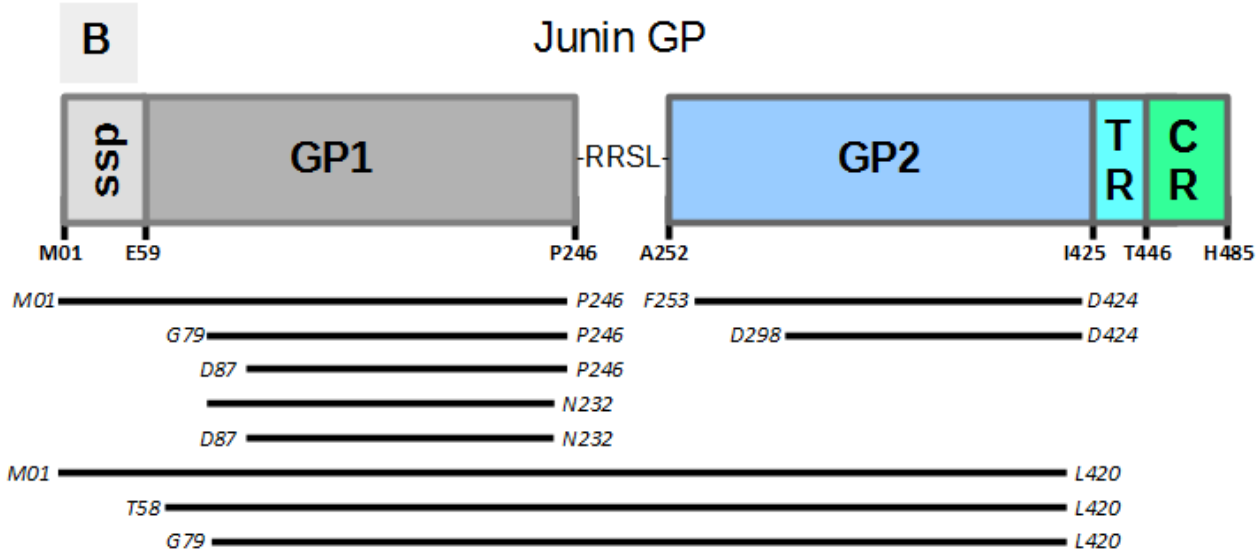
N-linked glycosylation, predicted by the NXSX/NXTX sequon (Mellquist et al. 1998), strongly influences protein structure by the hydrophilic nature and the molecular weight of N-linked glycans (Imperiali & O'Connor 1999) and has been shown to increase global protein stability, possibly by influencing entropic effects, and to play a role in assisting the formation of oligomeric complexes (Rudd et al. 1997). On the other hand, N-linked glycosylation increase heterogeneity and surface disorder and might preclude the formation of crystals.

The construct design for the Old-World Lassa virus, and the New-World Junin and Guanarito virus glycoprotein expression trials was based on the X-ray crystallographic structures of the New-World

Arenavirus Machupo GP1 (pdb:2WFO) and the Old-Word Arenavirus LCMV GP2 (pdb:3MKO) glycoproteins (Figure 22, Table 03). The two protein structures were respectively solved by (Bowden et al. 2009) and (Igonet et al. 2011). During construct design, the number of cysteines and the predicted O- and N-linked glycolysation sites in the putative constructs were taken into account to define boundaries. Construct design of Lassa and Junin glycoprotein precursor GP included wild-type sequences and mutated sequences with altered cellular convertase cleavage sites.

Vectors were constructed using the pOPIn suite of very high efficiency expression vectors, all containing cleavable His-tags and suitable for expression in mammalian HEK293 cells.


A
Lassa GP
ssp
GP1
-RRLL-
GP2
T
R
C
R
M01
M59
S254
G259
L432
T453
R490
M01 S254
M59 S254
E66 S254
L72 S254
N75 S254
T80 S254
L72 G239
L72 R247
M01 P426
M59 P426
L72 P426
T260 D431
D305 D431


B
Junin GP
ssp
GP1
-RRSL-
GP2
T
R
C
R
M01
E59
P246
A252
I425
T446
H485
M01 P246
G79 P246
D87 P246
N232
D87 N232
M01 L420
T58 L420
G79 L420
F253 D424
D298 D424

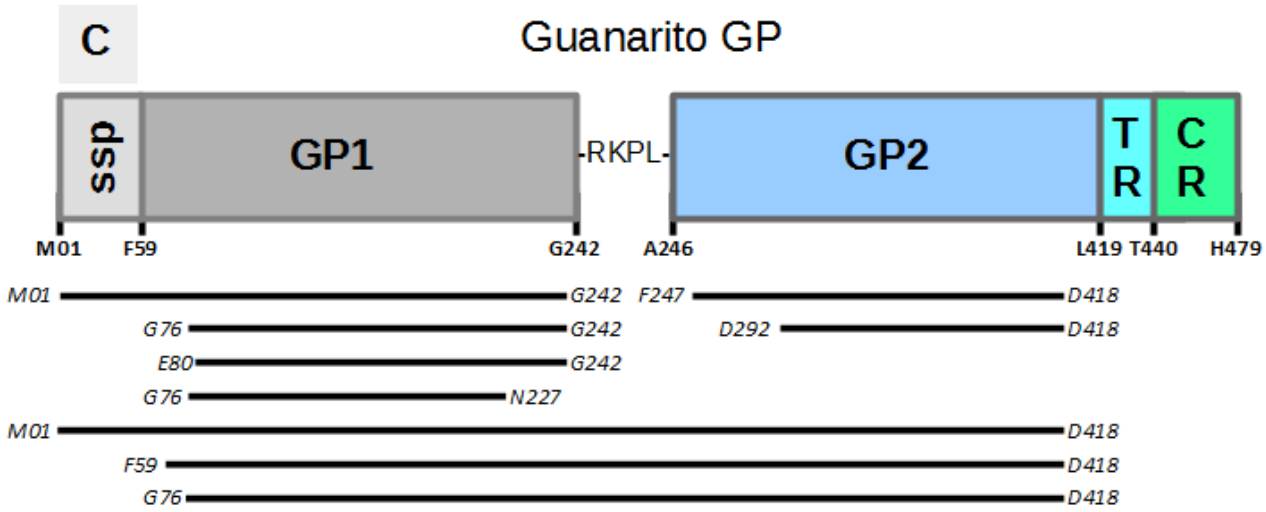


***Figure 22:* Schematic diagram of LASV, JUNV and GUNV construct design for glycoprotein expression**

The construct design for expression trials of Lassa (A), Junín (B), and Guanarito (C) glycoproteins is presented. The short segment S of the arenavirus genome encodes the glycoprotein precursor GP-C that is proteolytically cleaved by the cellular proprotein convertase site 1 protease (SK-1/S1P) to yield the stable signal peptide (ssp), GP-1 and GP-2 (TR: transmembrane region; CR: cytoplasmic region).

Gene name	N-terminal boundary	C-terminal boundary	Protein mass	Number of Cysteines	Number of N	Vector
LASV GP1	L72	S254	20.7	6	6	pOPINTTGNeo
LASV GP1	T80	S254	19.8	6	5	pOPINTTGNeo
LASV GP1	M59	S254	22.3	6	6	pOPINTTGNeo
LASV GP1	E66	S254	21.4	6	6	pOPINTTGNeo
LASV GP1	N75	S254	20.4	6	6	pOPINTTGNeo
LASV GP1	L72	G239	18.9	6	6	pOPINTTGNeo
LASV GP1	L72	R247	19.9	6	6	pOPINTTGNeo
LASV GP1	G02	S254	28.5	8	6	pOPINENeo
LASV GP2	T260	D431	19.9	6	4	pOPINTTGNeo-3C-CD4
LASV GP2	T260	D431	19.9	6	4	pOPINTTGNeo
LASV GP2	D305	D431	14.9	3	4	pOPINTTGNeo
LASV GP2	D305	D431	14.9	3	4	pOPINH
LASV GP	G02	D431	48.5	14	9	pOPINENeo
LASV GP	M59	D431	42.3	12	9	pOPINTTGNeo
LASV GP	M59	D431	42.3	12	9	pOPINTTGNeo
LASV GP	L72	D431	40.7	12	9	pOPINTTGNeo
LASV GP-mut	G02	D431	48.5	14	9	pOPINENeo
LASV GP-mut	M59	D431	42.3	12	9	pOPINTTGNeo
LASV GP-mut	M59	D431	42.3	12	9	pOPINTTGNeo
LASV GP-mut	L72	D431	40.7	12	9	pOPINTTGNeo
GUNV GP1	G76	G241	18.9	6	3	pOPINTTGNeo
GUNV GP1	E80	G241	18.5	6	3	pOPINTTGNeo
GUNV GP1	G02	G241	27.2	7	3	pOPINENeo?
GUNV GP1	G76	N227	17.4	6	3	pOPINTTGNeo
GUNV GP2	F247	D418	20.0	6	5	pOPINTTGNeo
GUNV GP2	F247	D418	20.0	6	5	pOPINTTGNeo-3C-CD4
GUNV GP2	D292	D418	16.8	5	5	pOPINTTGNeo
GUNV GP2	D292	D418	16.8	5	5	pOPINH
GUNV GP	G02	D418	47.7	13	7	pOPINENeo
GUNV GP	F59	D418	41.3	12	9	pOPINTTGNeo
GUNV GP	G76	D418	39.4	12	9	pOPINTTGNeo
JUNV GP1	G79	P246	19.3	6	4	pOPINTTGNeo
JUNV GP1	D87	P246	18.4	6	4	pOPINTTGNeo

JUNV GP1	G79	N232	17.7	6	4	pOPINTTGNeo
JUNV GP1	D87	N232	16.8	6	4	pOPINTTGNeo
JUNV GP2	F253	D424	19.9	6	3	pOPINTTGNeo
JUNV GP2	F253	D424	19.9	6	3	pOPINTTGNeo-3C-CD4
JUNV GP2	D298	D424	15.0	3	3	pOPINTTGNeo
JUNV GP2	D298	D424	15.0	3	3	pOPINH
JUNV GP	G02	D424	48.2	13	7	pOPINENeo
JUNV GP	T58	D424	41.9	12	7	pOPINTTGNeo
JUNV GP	T58	D424	41.9	12	7	pOPINTTGNeo
JUNV GP	G79	D424	40.0	12	7	pOPINTTGNeo
JUNV GP-mut	G02	D424	48.9	13	7	pOPINENeo
JUNV GP-mut	T58	D424	42.3	12	7	pOPINTTGNeo
JUNV GP-mut	G79	D424	40.0	12	7	pOPINTTGNeo

<u>Table 03:</u> Construct design for Lassa, Junin and Guanarito glycoproteins

Constructs were designed for Old-world Lassa virus and New-World Junin and Guanarito viruses to investigate the structure of glycoproteins GP1, GP2 and full-length GP. Mutations in the full length glycoprotein GP were targeted to the cellular convertase SKI-1/S1P cleavage site to prevent any proteolysis.

8.2. Small-scale expression trials

This work was kindly supervised by Dr. Louise Bird (OPPF).

The expression trials were performed at the Oxford Protein Purification Facility UK at the Research Complex at Harwell using pOPIN vectors as described below and oligoes normalised to 100 µM. Constructs designs was done in collaboration with Dr. Thomas Bowden, OPIC, Division of Structural Biology, University of Oxford and cDNA synthetically produced from a commercial source (GeneArt).

High-Throughput PCR Reactions

A Master Mix is prepared as followed:

Number of 50ul Reactions	1	50
2 X KOD Hot Start Buffer (µl)	25	1250
DNTP mix (2mM)	10	500
KOD Hot Start (1 U/µl)	1	50
Sterile Water (µl)	6	300
Total Volume (µl)	42	2100

KOD Xtreme^{TM} Hot Start Polymerase (Novagen 71975-3) Master Mix

42 µl Master mix were dispensed per well of the PCR plate. 3 µl of 10 µM forward primer and 3 µl of 10 µM reverse primer were added in addition to 2 µl of template plasmid (10-20 ng/µl) to the appropriate PCR wells.

Thermal cycling was performed using the following parameters:

98°C	10 seconds
98°C	0 or 1 second
60°C	5 seconds
72°C	X minutes (based on the largest PCR product expected: 15sec/kbp)
Go to step 2	29 times
72°C	2 minutes
4°C	hold

Analysis of the PCR products was done on a 1.25% TBE (Tris-borate-EDTA) Agarose gel run at 100 V for 30 minutes using 5ul aliquots of the PCR products, 2.0 ul 5x DNA Loading buffer (0.25 % w/v Bromophenol Blue in 30% v/v Glycerol), 5ul Hyperladder I (BioLine BIO-33025/BIO-33026) and SYBRSafe stain (InVitrogen S33110/S33100).

Small-scale expression trials showed high level expression of Lassa virus GP2, Guanarito GP precursor, Guanarito GP2, Junin GP1, with proteins displaying the expected size of respectively 15 kDa, 40 kDa, 13 kDa and 20 kDa (not including the molecular weight added by the possible occupancy of N-linked glycosylation sites: +/- 10 kDa).

Low-level expression of Junin GP2 is also detected with protein sizes of 15 kDa (+/- 10 kDa). To understand which factors precluded the expression of Lassa virus GP1, the putative relative three-dimensional positions of the LASV GP1 cysteine residues were analysed using the crystallographic atomic structure of Machupo glycoprotein 1 (2WFO). Assuming similar folding of the two proteins, the distance between LASV C230-C35 (7 Å) and LASV C179-154 (9 Å) based on the Machupo X-ray

data would suggest the cysteines engage in intramolecular disulphide bonds to stabilise the three-dimensional structure. However, the large distance between LASV C117 and LASV211 (20 Å) clearly indicate the absence of an additional intramolecular disulphide bond, suggesting that these two cysteines residues might be involved in the oligomerisation of the LASV glycoprotein complex and destabilise the monomers LASV GP1. Protein expression trials using LASV GP1 constructs with C117 and C211 mutated into Methionine (C117 mutated alone, C211 mutated alone and co-mutation of C117 and C211) will help test this hypothesis.

In-house expression of Guanarito GP2 and purification using SEC 200 10/30 column indicates the presence of a single oligomeric species in a glycosylated Guanarito GP2 sample. Assessment of sample purity by SDS-page using glysosylated Guanarito GP2 proteins shows solution homogeneity, suitable for crystallisation trials.

Results and Discussion

9. *Cryo-electron tomography of chemically-fixed Lassa virions*

9.1. Lassa virions are highly pleomorphic structures whose host cell-derived envelope is covered by surface glycoprotein spikes to a varying extent.

Low-dose imaging of a suspension of LASV chemically fixed with 4% paraformaldehyde reveals a highly pleomorphic morphology for Lassa virions (Figure 23). LASV appeared as an enveloped virus whose form varied from spherical to oval (Figure 23 A-D).

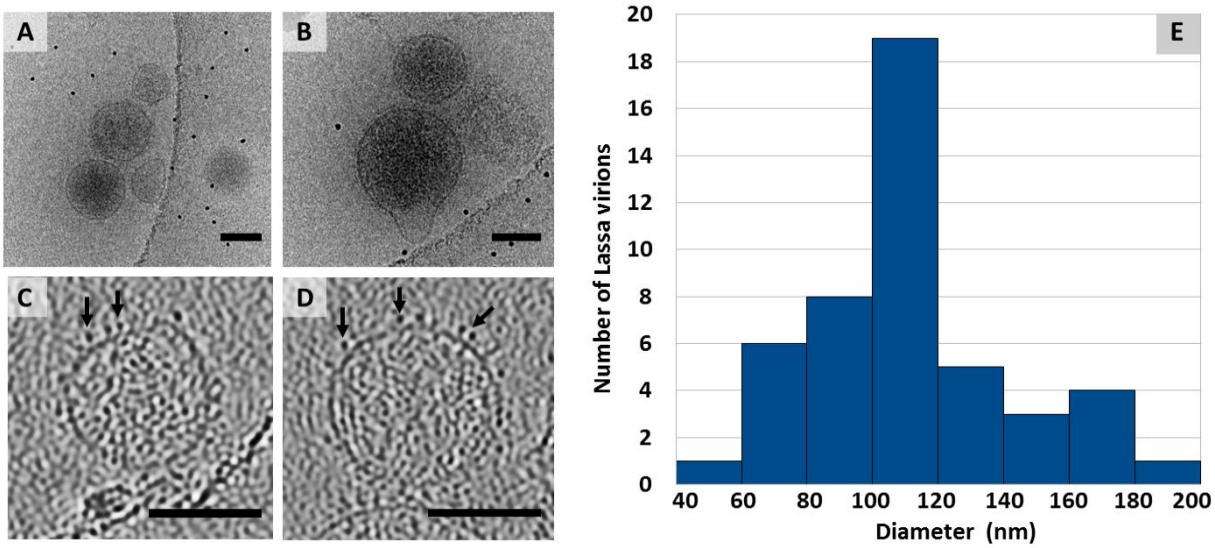


***Figure 23:* Lassa virus pleomorphic structure observed by cryo-electron tomography**

Two-dimensional projections and tomographic reconstructions acquired under low-dose conditions show the general morphology of Lassa virus. The enveloped virions are pleomorphic with varying diameters from 45 to over 190 nm, and varying shapes (A-B-E) (scale bars = 100 nm). Some virions displayed a distortion of their content and membrane possibly resulting from the budding process (B, arrow). Tomographic slices of LASV allow the observation of the Lassa glycoprotein complex located at the surface at the viral envelope and where they cover the virion to a varying extent (C-D, arrows)

The diameter of the observed virions showed great variations: the smallest particles imaged presented a diameter of 45 nm, while le largest particle observed was 198 nm in diameter. The population of LASV observed under cryo-conditions unevenly distributed itself into several size classes: 40% of the imaged LASV population displayed a diameter from 100 to 120 nm, while 30% and 26 % respectively showed a diameter from 60 to 100 nm and from 120 to 180 nm. Only 4% of the observed population of virions displayed a diameter below 60 nm or above 180 nm (Figure 21-E). This lack of uniformity in the virus global architecture hints at the arrangement of its structural components. Lassa virions are expected to display no global symmetrical lattice arrangement or, in other words, structural components are expected to be locally ordered: e.g.: symmetrical arrangement of surface homopolymeric complexes, local protein-protein interactions (GP-Z; Z-NP).

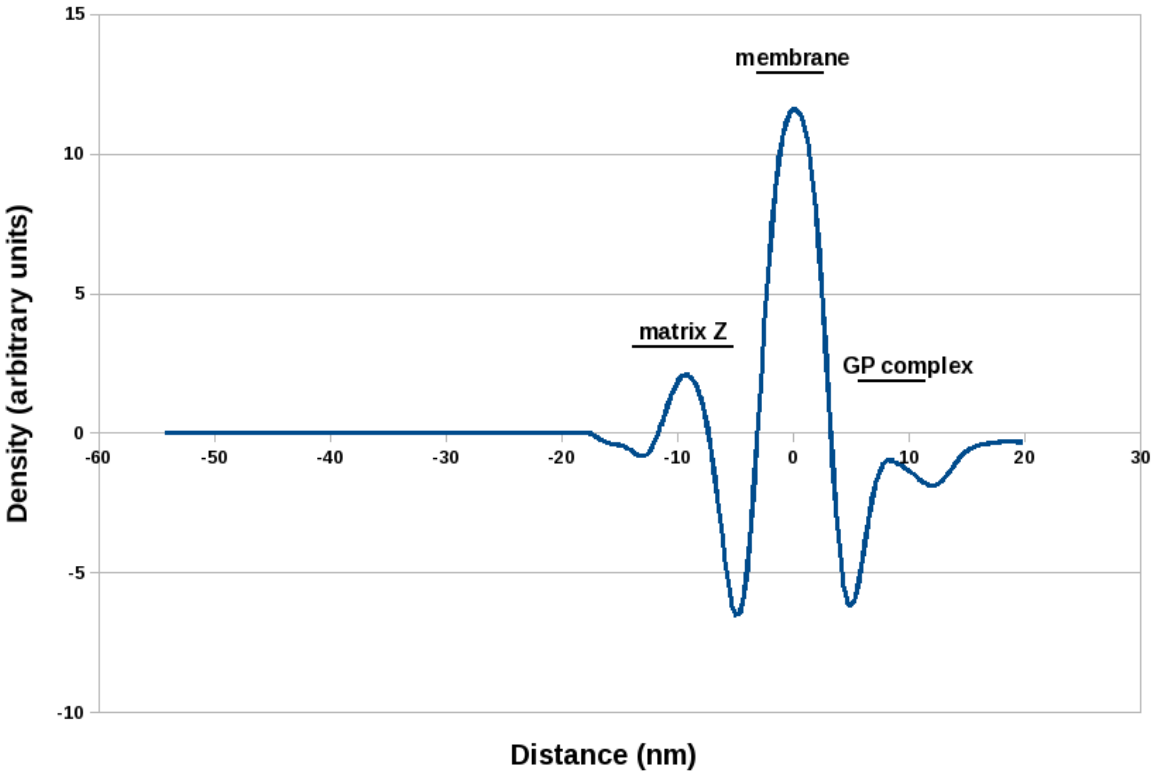
In addition to this broad variety in the virus shape and size, it was observed that a number of LASV virons possessed a distortion in the virion content and envelope. Similar structures had already been observed in electron-microscopy studies of haemorrhagic fever-associated viruses. During their investigation of the mechanisms involved in the budding process of filoviruses using electron tomography of plastic-embedded infected cells, Welsch and collaborators described the occurrence of membrane distortions localized at the end of filament-like Marburg virus particles (Welsch et al. 2010). Tomograms of 300 nm thick sections revealed membrane bulges or “hooks” at one end of filovirus particles while the other extremity formed a round, intact hemisphere. When occurring, this distortion only affected one pole of the filamentous viral particle, and was never found on both ends of the virus structure. It was proposed for this distortion to be the result of a local instability in the virus membrane that resulted from the scission of burgeoning Marburg virions off their parent cells. Based on these observations, membrane bulges distorting the shape

of Lassa virus particles can be hypothesized to be the result of the pinching mechanism involved when budding off the host-cell membrane (Figure 23-B).

The advantage of electron microscopy performed under cryo-conditions with a low-dose imaging strategy when dealing with biological samples, which are prone to beam-induced damage, is the retention of high-resolution information. Tomographic volumes reconstructed from tilt-series acquired at relatively high defoci (-7 and -6 um) to help generate contrast shed light on the ultrastructure of Lassa virions and the spatial distribution of major structural proteins (Figure 23 C-D).

The glycoprotein complex GP appeared as a randomly-distributed electron density located on the surface of the virions. Electron tomograms showed globular spikes extending radially outwards from the outer bilayer leaflet (Figure 23 C-D arrows). The spikes appeared as composed by an approximately 10 nm in diameter globular head and a stalk region. They seemed to cover the envelope with different degree of density that varied from one virion to another among the population observed so no mean spacing was established.

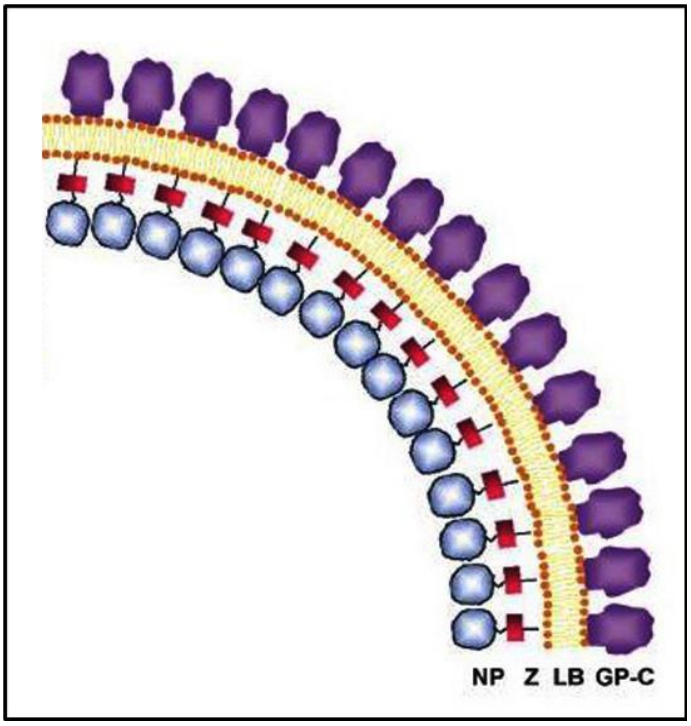
Analysis of the density distribution across the viral envelope reveals the presence of protein density proximal to the inner membrane of the envelope, assumed to be the matrix protein Z (Figure 24).



<u>Figure 24:</u> Membrane density profile reveals the location of the Z and GP proteins

A profile of density distribution across the viral membrane is shown for an averaged structure of the Lassa glycoprotein complex obtained by sub-tomogram averaging using the Jsubtomo package. The plot of the density distribution as a function of distance from the centre of the membrane indicates the locations of the matrix protein Z and of the glycoprotein complex (bars).

It is worth noting that previous studies investigating the supramolecular assembly of arenaviruses uncovered two layers of density apposed to the inner leaflet of the viral membrane, layers that were assigned to the matrix protein Z and the nucleoprotein NP (Neuman et al. 2005). Interestingly, the analysis of *en face* views of New-World Pichinde, New-World Tacaribe and Old-World LCMV virions that lacked the GP-C surface complexes revealed diffraction patterns for the NP-associated density suggesting a paracrystalline formation. A class average of side views showed a grid-like arrangement of about 50-Å dot-like densities. Based on these observations, it was proposed that the number of packaged NP proteins might regulate the size of the virion. According to earlier quantitative studies, this model would be consistent with the reported L/NP/GP-1/GP-2/Z molar ratio of Lassa virus 1: 160:60:60:20 (Strecker et al. 2003) (Figure 25).



<u>Figure 25:</u> Schematic representation of the arenavirus organization

The putative organization of arenaviruses is depicted with the nucleoprotein NP in blue, the matrix protein Z in red, the lipid bilayer LB in brown and yellow, and the surface glycoprotein complex GP-C in purple. This model would be consistent with previous quantitative studies reporting the L/NP/GP-1/GP-2/Z molar ratio of LASV to be 1:160:60:60:20 (Neuman et al. 2005).

To investigate this possibility and the structure of the LASV glycoprotein spikes, sub-tomogram was performed using the in-house developed package *Jsubtomo*.

10.1. Sub-tomogram processing of LASV surface glycoprotein complex

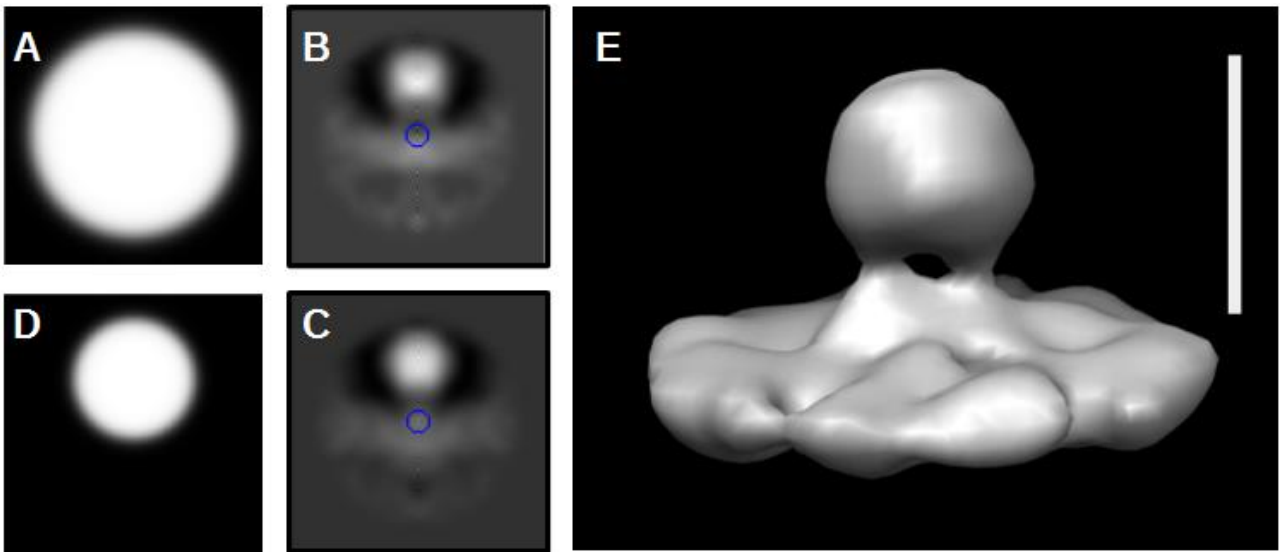
To sum up the procedure used to generate the refined electron-density map of GP:

Sub-tomogram averaging of the complex spike was performed by first dividing the dataset into two sub-datasets of equal size and generating two independent initial models by extracting, rotating and averaging the manually picked GP. Refinement of the spikes locations and orientations was performed independently in each sub-dataset spikes to assess the validity of the obtained glycoprotein averaged structure. The achieved resolution was then calculated using a three-dimensional Fourier Shell Correlation (FSC), which measured the normalised cross-correlation coefficient between two different Fourier transforms of two different volumes over different shells in Fourier Space (Figure 26).

The averaged structure displayed a 90-100 Å globular ectodomain supported by two stalks, hinting at a dimeric oligomerisation of the glycoprotein complex. Analysis of the FSC determined the resolution at 35 Å, coherent with the first zero-crossing of the Contrast Transfer Function for data acquired at -7 and -6 um defocus.

This putative two-fold oligomerisation of the GP complex is in disagreement with previous studies that predicted the glycoprotein to adopt a trimeric or maybe a quadrimeric arrangement on the viral surface (Schlie et al. 2010). Furthermore, if considering the class I fusion protein character of the transmembrane GP2, a trimeric arrangement would be further in accordance with the well-researched trimer of hairpins conformation specific to this class of proteins. Possible explanations for this disagreement may arise from the GP structure observed above: as described, the protein exhibits a globular head structure measuring around 95 Å. To compute this refined electron-

density map, a dataset of tomograms reconstructed from tilt-series acquired at -7 and -6 µm underfocus were used. As the present protocol excluded any CTF-correction implementation, the highest resolution achievable was limited to the first zero-crossing of the contrast transfer function located around 35 Å. Thin structural elements that would permit the observation of a three-fold symmetry within the globular head might be too small to be detected by the current protocol. To overcome this difficulty, additional data recorded closer to focus would be necessary and/or CTF correction to retrieve high-resolution spatial frequencies suppressed during data acquisition.



<u>Figure 26:</u> Sub-tomogram averaging using Jsubtomo: process tools and structure

Sub-tomogram averaging in Jsubtomo was carried out using a paralleled template-matching refinement process performed in two independent sub-datasets containing nearly-identical numbers of GP particles. This process first required the generation of models to which particles of interest were cross-correlated. The manually selected particles were masked away from any surrounding structures by applying a real-space spherical mask that displayed a Gaussian fall-off (A) before being averaged in each sub-dataset into initial models (B: sub-dataset "even",C: sub-dataset "odd"; Blue circles indicate the volume origin). The refinement of each spike view-vector's direction was performed using the same real-space mask, while refining the angle around the vertical axis of the spike required a smaller spherical mask including GP globular head but excluding the membrane and neighbouring background (D). Subtomogram averaging using two parallel and independent sub-dataset shows a globular ectodomain supported by two proteic trans-membrane stalks at a resolution of 35 Å according to Fourier Shell Correlation (FSC criterion = 0.143)(bar: 11nm)(E).

10.2. CTF-correction implementation scheme

Discussion with the Boulder Laboratory for 3-D Electron Microscopy of Cells, University of Colorado, Boulder, USA contributed to the implementation of an electron tomography data collection scheme acquiring additional images for the purpose of subsequent CTF correction of the tilt-series. This new acquisition scheme was developed by Matthias Eibauer and first utilized to unravel the structure of membrane proteins *in situ*, including the mycobacterial outer membrane protein MspA (Figure 27) (Eibauer et al. 2012).

The low-dose imaging conditions required to collect tomographic data on specimens sensitive to beam-induced damage precludes the determination of the CTF parameters directly from 2D projections because of their inherent low SNR. To by-pass these difficulties, Eibauer and collaborators designed a data-collection strategy that allowed a projection-based determination and hence correction of the CTF by phase-flipping. During the tilt-series acquisition, two additional images T1 and T2 are recorded in addition to the recorded area R. Those two additional images are recorded at high electron dose, and areas to be images are chosen on the tilt axis to avoid deviation from the eucentric height of R, and hence a deviation from the defocus value in R. From T1 and T2, power spectra are calculated and two defocus values d1 and d2 respectively, are calculated for each image. The defocus of the recorded image R is then retrieved using the equation:

$$\mathbf{Defocus(R) = ½ \times (Defocus(T^1) + Defocus(T^2))}$$

and used to retrieve high-frequency information by phase-reversal.

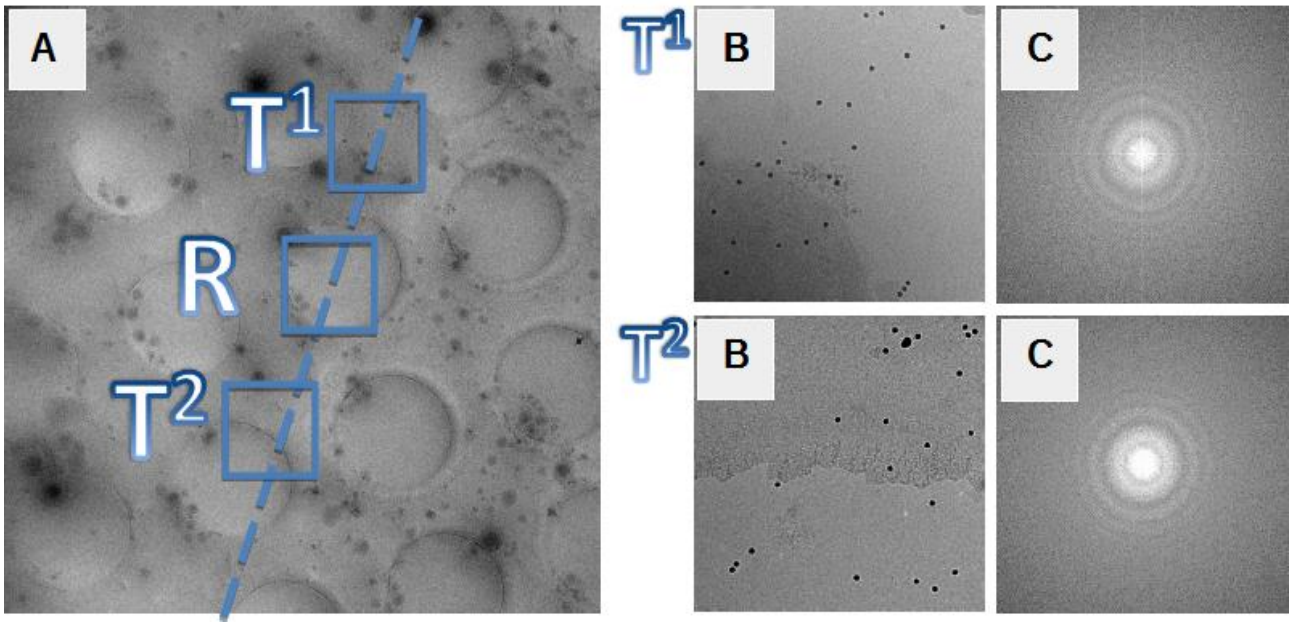


***Figure 27:* Low-dose acquisition scheme for CTF-corrected cryo-electron microscopy**

To retrieve high-resolution frequencies lost during data-acquisition due to the effects of the CTF, we need to precisely determine the defocus under which the tilt-series was recorded. To do so, a new acquisition scheme was implemented within SerialEM as the low-dose conditions under which the area of interest is imaged precludes the determination of the defocus due to the low signal-to-noise ratio (A). In addition to the record area (R) containing the object of interest, two additional trial images T1 and T2 are acquired (B). These two additional regions, located on the tilt-axis, are imaged under high-dose conditions. From T1 and T2, a power spectrum is calculated and a defocus value is determined for each image according to the position of the TON rings (C). The defocus of the record image R is retrieved by Defocus(R) = ½ x (Defocus(T1) + Defocus(T2)) and subsequently used to correct the image by phase reversal (Eibauer et al. 2012).

An important feature of this strategy is the possibility of applying a projection-based CTF correction to the tomographic data as opposed to performing one global correction to the tilt-series by utilising one single value of defocus. Defocus determination using the strategy described above highlighted the limitations of global CTF corrections (Figure 28). Deviation from the 0° tilt value of defocus increases at high tilt until reaching a 30% change in value in some recorded cases. Utilising one single defocus value to CTF correct tilt-series would lead to inaccurate determination of the first zero-crossing frequency in most 2D projections, and hence would preclude the retrieval of high-frequencies.

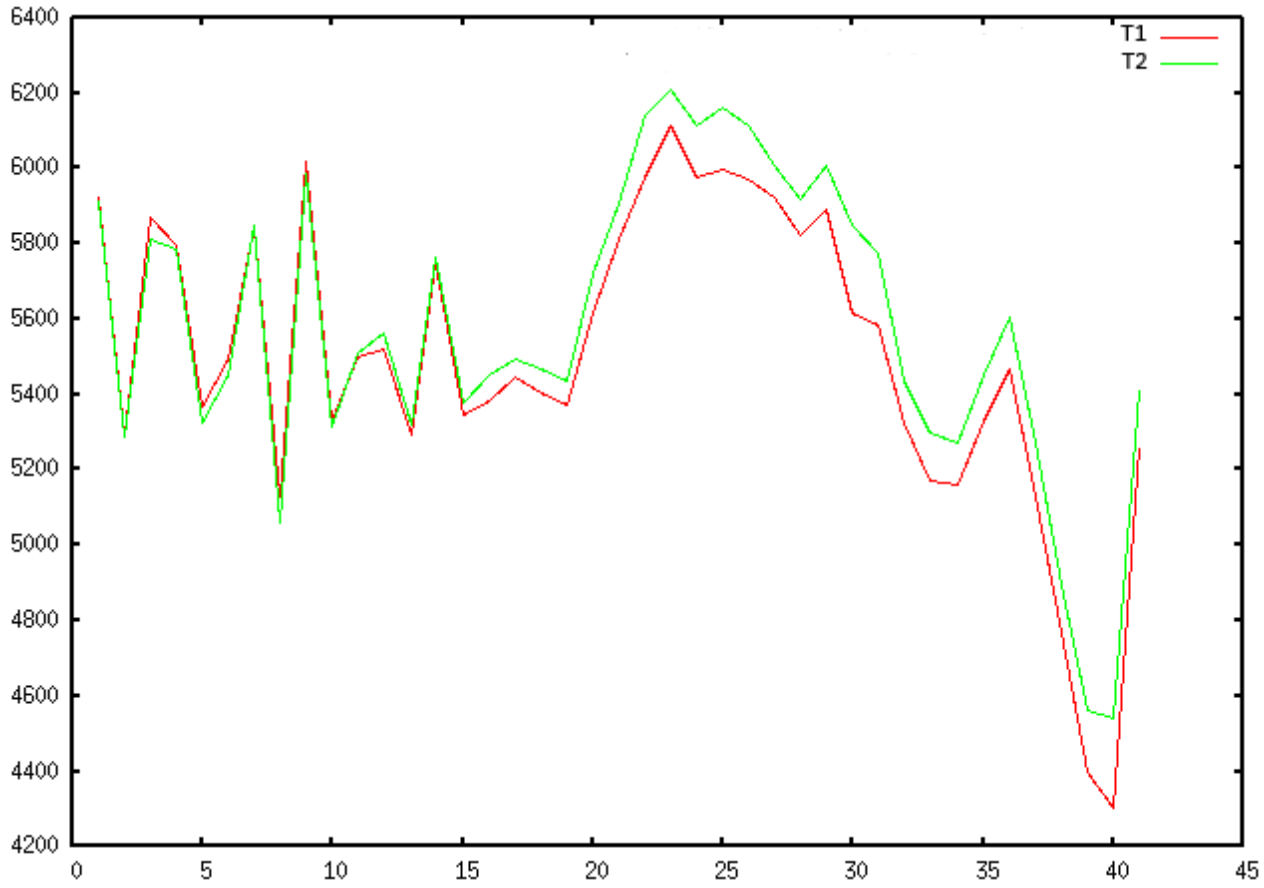


***Figure 28:* Projection-based defocus determination in tomographic tilt series**

Defocus calculation using the method developed by (Eibauer et al. 2012) highlights significant variations of defocus over the course of tilt-series data acquisition. For a pre-set defocus value of 6 UM, automated acquisition records projections at defoci ranging from 6.2 µm to 4.4 µm at high tilt. By averaging the defocus values obtained by calculating the power spectrum of the additional image T1 (red) and T2 (green), an accurate defocus value can be retrieved for every projection image in a tilt-series.

In here, CTF correction prior to tomogram computation and sub-tomogram averaging did not result in the expected improvement of the refined spike structure whose symmetry and hence oligomerisation status remains unclear. Lack of improvement in the resulted averaged structure raised the question of conformational heterogeneity of the selected sample and possible contamination of manually picked virus particles with glycoprotein-bearing vesicles. The selection of unwanted protein particles could be avoided by utilising a specimen sample of virus-like particles displaying the LASV GP purified by ultra-centrifugation and sucrose-cushion.

To provide insight into the high-resolution architecture of arenavirus glycoproteins, protein expressions, purifications and crystallisations were carried out in view of determining atomic structures by X-ray diffraction.

11. Crystal structure of Venezuelan haemorrhagic Fever virus GP2

11.1. Crystal structure of the Guanarito virus GP-2 subunit

The GP-2 subunit of the Venezualan Guanarito virus (GenBank accession number AAS55656.1, residues 292 to 418) was cloned into a pOPINTTGNeo mammalian expression vector using high-throughput In-Fusion PCR cloning (REF). To prevent the aggregation of recombinant GP-2 proteins, the last 50 amino-acids located at the N-terminal end of the GP-2 ectodomain, which contain the the N-terminal hydrophobic fusion loop, were excluded from the construct. Expression was performed in GlcNAc transferase I-deficient human embry- onic kidney 293S (HEK 293S) cells, to help crystallisation by assuring a homogenous Man5GlcNAc2 glycosylation process (Chang et al. 2007; Crispin et al. 2006; Reeves et al. 2002). Crystallisation trials were conducted using the sitting-drop vapor diffusion method by utilizing 100 nl protein plus 100 nl precipitant (0.15MLi2SO4, 0.1 M citric acid, pH 3.5, 18% [wt/vol] polyethylene glycol 6000 [PEG 6000]) equilibrated against 95 µl reservoirs at 22°C (Walter et al. 2005).

The immersion of crystals in solution containing 25% (vol/vol) ethylene glycol before, the subsequent cooling to 100 K, the recording of diffraction data at beamline I04, Diamond Light Source down to 4.1-Å resolution, and the final structure refinement using molecular replacement with Phaser were performed by Dr. Thomas Bowden, STRUBI, University of Oxford (Table 04) (Parsy et al. 2013).

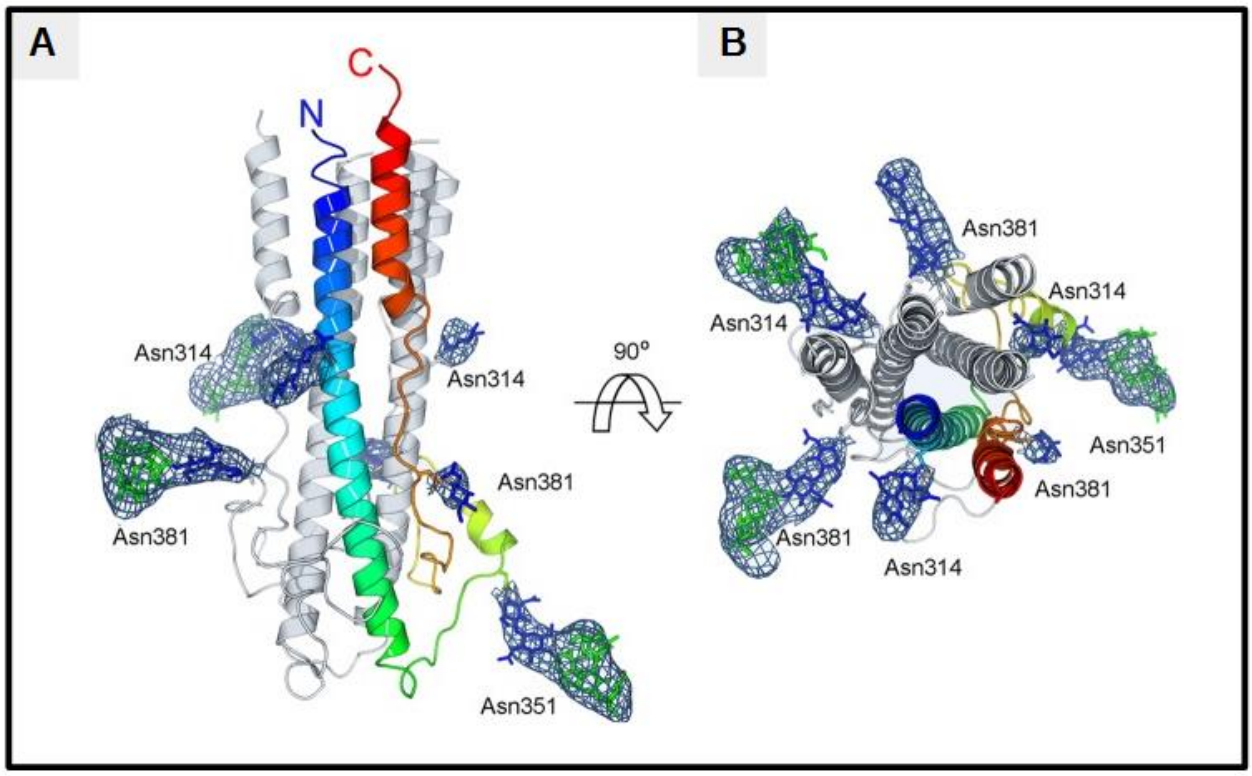
The asymmetric unit of the studied crystal contained three GP-2 subunits that were found to be composed of three different regions: a N-terminal α-helix located between residues 301 and 346; a “T region” (Igonet et al. 2011) spanning from residue 347 to residue 398; and a C-terminal α-helix located between residue 399 and residue 408 (Figure 29). The 50-amino-acid-long N-terminal α-helix runs along the 85-Å length of the GP-2 monomer. This N-terminal helix is followed by a series

of loops and a short α-helix (aa 358-364) that together form the “T region”. Through this T region, the N-terminal helix then connects to the short C-terminal helix. In the trimeric GP-2, the C-terminal and the N-terminal helices are packed in an antiparallel fashion and arrange into a trimeric coiled-coil structure. A “stutter” is present in the N-terminal region of the coiled-coil, which, together with the overall architecture of the trimer, confirms that the New-World Guanarito GP-2 sub-unit is a class I fusion glycoprotein whose structure is here presented in a post-fusion conformation (Harrison 2008; White et al. 2008)

PARAMETER	RESULTS
Data collection statistics	
Resolution range in Å	50.0–4.14 (4.29–4.14: outer shell)
Space group	$P4_1$
Cell dimensions	
a, *b*, *c* in Å	99.2, 99.2, 79.9
α, β, γ in degrees	90.0, 90.0, 90.0
Wavelength in Å	0.953
Number of unique reflections	5,894 (576)
Completeness (%)	100.0 (99.6)
R_{merge} (%)	9.9 (>100.0)
$I/\sigma I$	26.6 (1.9)
Average redundancy	5.1 (5.1)
CC $_{½}$ (Karplus & Diederichs 2012)	0.998 (0.362)
Refinement statistics	
Resolution range in Å	38.4–4.14 (4.63–4.14)
Number of reflections	5,561 (1,563)
R_{work} (%)	25.5 (29.1)
R_{free} (%)	27.6 (33.7)
RMSD (Davis et al. 2007)	
Bonds in Å	0.010
Angle in degrees	1.4
Between NCS-related Cα atoms	0.8
Number if molecules per asymmetric unit	3
Number of atoms per asymmetric unit (protein/carbohydrate)	2,649/358
Average *B* factors ($Å^{2)}$ (protein/carbohydrate)	214/290
Model quality (Ramachandran plot)f	
Favoured region (%)	91.2
Allowed region (%)	95.5
Disallowed region (%)	4.5

<u>Table 04:</u> Crystallographic data and refinement statistics

(Parsy et al. 2013) (Courtesy of Dr. Thomas Bowden).



<u>Figure 21:</u> Structure of glycosylated GUNV GP-2

(A) Cartoon diagram of trimericGTOVGP2in the postfusion conformation. One protomer in the asymmetric unit is colored as a rainbow, with the N-terminus shown in blue and the C-terminus in red. N-linked carbohydrates are shown as sticks, with GlcNAc residues colored blue and mannose residues colored green. A maximum-likelihood weighted $2F_o$-F_c electron density map is plotted around each glycan at 1?. (B) View of panel of A rotated by 90°. The fusion loop N-terminal region colocalises with the C-terminal trans-membrane region, consistent with a post-fusion complex conformation (Parsy et al. 2013) (Courtesy of Dr. Thomas Bowden).

Acknowledgements

I would like to thank the Wellcome Trust for founding the project and the staff of the Division of Structural Biology for their help and continuous support. I would like to express special thanks to my supervisor Dr. Juha Huiskonen for scientific and personal guidance throughout my four years in STRUBI, and to Dr. Thomas Bowden and Dr. Ray Owens for supervising the X-ray crystallography part of the project.

Special thanks to Dr. Pascale Schellenberger, David Bitto, Nayab Malik, Steinar Halldórsson, and Margareth Jones for their presence and advice.

Thanks to Christ Church GCR for being a port-of-call in the storm.

Many thanks to my mentor, Julia Hill.

References

Adrian, M. et al., 1984. Cryo-electron microscopy of viruses. *Nature*, 308(5954), pp.32–6.

Agnihothram, S.S., York, J. & Nunberg, J.H., 2006. Role of the stable signal peptide and cytoplasmic domain of G2 in regulating intracellular transport of the Junín virus envelope glycoprotein complex. *Journal of virology*, 80(11), pp.5189–98.

Arranz, R. et al., 2012. The structure of native influenza virion ribonucleoproteins. *Science (New York, N.Y.)*, 338(6114), pp.1634–7.

Auperin, D.D. et al., 1984. Sequencing studies of pichinde arenavirus S RNA indicate a novel coding strategy, an ambisense viral S RNA. *Journal of virology*, 52, pp.897–904.

Baize, S. et al., 1999. Defective humoral responses and extensive intravascular apoptosis are associated with fatal outcome in Ebola virus-infected patients. *Nature medicine*, 5, pp.423–426.

Baize, S., Kaplon, J. & Faure, C., 2004. Lassa virus infection of human dendritic cells and macrophages is productive but fails to activate cells. *The Journal of Immunology,* 2004 Mar 1;172(5):2861-9.

Bartesaghi, A. et al., 2008. Classification and 3D averaging with missing wedge correction in biological electron tomography. *Journal of structural biology*, 162(3), pp.436–50.

Baumeister, W., Grimm, R. & Walz, J., 1999. Electron tomography of molecules and cells. *Trends in cell biology*, 9(2), pp.81–5.

Beck, M. et al., 2004. Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science (New York, N.Y.)*, 306(5700), pp.1387–90.

Beyer, W.R. et al., 2003. Endoproteolytic processing of the lymphocytic choriomeningitis virus glycoprotein by the subtilase SKI-1/S1P. *Journal of virology*, 77(5), pp.2866–72.

Borden, K.L., CampbellDwyer, E.J. & Salvato, M.S., 1997. The promyelocytic leukemia protein PML has a pro-apoptotic activity mediated through its RING domain. *FEBS letters*, 418(1-2), pp.30–4.

Borio, L. et al., 2002. Hemorrhagic fever viruses as biological weapons: medical and public health management. *JAMA : the journal of the American Medical Association*, 287(18), pp.2391–405.

Borrow, P. & Oldstone, M.B., 1992. Characterization of lymphocytic choriomeningitis virus-binding protein(s): a candidate cellular receptor for the virus. *Journal of virology*, 66(12), pp.7270–81.

Borrow, P. & Oldstone, M.B., 1994. Mechanism of lymphocytic choriomeningitis virus entry into cells. *Virology*, 198(1), pp.1–9.

Bowden, T. a et al., 2009. Unusual molecular architecture of the machupo virus attachment glycoprotein. *Journal of virology*, 83(16), pp.8259–65.

Bowden, T.A. et al., 2013. Orthobunyavirus ultrastructure and the curious tripodal glycoprotein spike. *PLoS pathogens*, 9(5), p.e1003374.

Bowen, M.D. et al., 2000. Genetic diversity among Lassa virus strains. *Journal of virology*, 74(15), pp.6992–7004.

Bowen, M.D., Peters, C.J. & Nichol, S.T., 1997. Phylogenetic analysis of the Arenaviridae: patterns of virus evolution and evidence for cospeciation between arenaviruses and their rodent hosts. *Molecular phylogenetics and evolution*, 8, pp.301–316.

Briese, T. et al., 2009. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS pathogens*, 5(5), p.e1000455.

Buchmeier, M. J., et al., 1995. *Family Arenaviridae.*, Elsevier/Academic Press, London, UK.

Buckley, S.M. & Casals, J., 1970. Lassa fever, a new virus disease of man from West Africa. 3. Isolation and characterization of the virus. *The American journal of tropical medicine and hygiene*, 19(4), pp.680–91.

Buckley, S.M., Casals, J. & Downs, W.G., 1970. Isolation and antigenic characterization of Lassa virus. *Nature*, 227(5254), p.174.

Burri, D.J. et al., 2013. The role of proteolytic processing and the stable signal peptide in expression of the Old World arenavirus envelope glycoprotein ectodomain. *Virology*, 436, pp.127–33.

Candurra, N.A. & Damonte, E.B., 1997. Effect of inhibitors of the intracellular exocytic pathway on glycoprotein processing and maturation of Junin virus. *Archives of virology*, 142(11), pp.2179–93.

Carazo, J.M. et al., 1992. Detection, classification and 3D reconstruction of biological macromolecules on hypercube computers. *Ultramicroscopy*, 40(1), pp.13–32.

Chang, V.T. et al., 2007. Glycoprotein structural genomics: solving the glycosylation problem. *Structure (London, England : 1993)*, 15(3), pp.267–73.

Charrel, R.N., de Lamballerie, X. & Emonet, S., 2008. Phylogeny of the genus Arenavirus. *Current opinion in microbiology*, 11, pp.362–368.

Choe, H. et al., 2011. Transferrin receptor 1 in the zoonosis and pathogenesis of New World hemorrhagic fever arenaviruses. *Current opinion in microbiology*, 14(4), pp.476–82.

Clegg, J.C.S., 2002. Molecular phylogeny of the arenaviruses. *Current topics in microbiology and immunology*, 262, pp.1–24.

Crispin, M. et al., 2006. Inhibition of hybrid- and complex-type glycosylation reveals the presence of the GlcNAc transferase I-independent fucosylation pathway. *Glycobiology*, 16(8), pp.748–56.

Culjkovic, B. et al., 2008. The eIF4E RNA regulon promotes the Akt signaling pathway. *The Journal of cell biology*, 181(1), pp.51–63.

Culjkovic, B., Topisirovic, I. & Borden, K.L.B., 2007. Controlling gene expression through RNA regulons: the role of the eukaryotic translation initiation factor eIF4E. *Cell cycle (Georgetown, Tex.)*, 6(1), pp.65–9.

Daniel, M.T. et al., 1993. PML protein expression in hematopoietic and acute promyelocytic leukemia cells. *Blood*, 82(6), pp.1858–67.

Davis, I.W. et al., 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic acids research*, 35(Web Server issue), pp.W375–83.

Debbie, J.G. & Abelseth, M.K., 1971. Pathogenesis of epizootic hemorrhagic disease. I. Blood coagulation during viral infection. *The Journal of infectious diseases*, 124(2), pp.217–22.

Delgado, S. et al., 2008. Chapare virus, a newly discovered arenavirus isolated from a fatal hemorrhagic fever case in Bolivia., *PLoS Pathog.* 2008 Apr 18;4(4).

Demby, A.H. et al., 1994. Early diagnosis of Lassa fever by reverse transcription-PCR. *Journal of clinical microbiology*, 32(12), pp.2898–903.

Dierksen, K. et al., 1992. Towards automatic electron tomography. *Ultramicroscopy*, 40(1), pp.71–87.

Dierksen, K. et al., 1993. Towards automatic electron tomography II. Implementation of autofocus and low-dose procedures. *Ultramicroscopy*, 49(1-4), pp.109–120.

Dobro, M.J. et al., 2010. Plunge freezing for electron cryomicroscopy. *Methods in enzymology*, 481, pp.63–82.

Dubochet, J. et al., 1971. A new preparation method for dark-field electron microscopy of biomacromolecules. *Journal of ultrastructure research*, 35(1), pp.147–67.

Dubochet, J. & McDowall, A.W., 1981. Vitrification of pure water for electron microscopy. *Journal of Microscopy*, 124(3), pp.3–4.

Dyck, J.A. et al., 1994. A novel macromolecular structure is a target of the promyelocyte-retinoic acid receptor oncoprotein. *Cell*, 76(2), pp.333–43.

Edington, G.M. & White, H.A., 1972. The pathology of Lassa fever. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 66(3), pp.381–9.

Egerton, R., 2005. Physical principles of electron microscopy. *Materials Today*, 8(12), p.49.

Eibauer, M. et al., 2012. Unraveling the structure of membrane proteins in situ by transfer function corrected cryo-electron tomography. *Journal of structural biology*, 180(3), pp.488–96.

Eichler, R., Lenz, O., Strecker, T., Eickmann, M., et al., 2003. Identification of Lassa virus glycoprotein signal peptide as a trans-acting maturation factor. *EMBO reports*, 4, pp.1084–1088.

Eichler, R. et al., 2004. Lassa virus glycoprotein signal peptide displays a novel topology with an extended endoplasmic reticulum luminal region. *The Journal of biological chemistry*, 279(13), pp.12293–9.

Eichler, R., Lenz, O., Strecker, T. & Garten, W., 2003. Signal peptide of Lassa virus glycoprotein GP-C exhibits an unusual length. *FEBS letters*, 538, pp.203–206.

Eschli, B. et al., 2006. Identification of an N-terminal trimeric coiled-coil core within arenavirus glycoprotein 2 permits assignment to class I viral fusion proteins. *Journal of virology*, 80(12), pp.5897–907.

Fernández, J.J., Li, S. & Crowther, R.A., 2006. CTF determination and correction in electron cryotomography. *Ultramicroscopy*, 106(7), pp.587–96.

Fernández, J.-J., Sanjurjo, J. & Carazo, J.-M., 1997. A spectral estimation approach to contrast transfer function detection in electron microscopy. *Ultramicroscopy*, 68(4), pp.267–295.

Fichet-Calvet, E. et al., 2007. Fluctuation of abundance and Lassa virus prevalence in Mastomys natalensis in Guinea, West Africa. *Vector borne and zoonotic diseases (Larchmont, N.Y.)*, 7(2), pp.119–28.

Fichet-Calvet, E. & Rogers, D.J., 2009. Risk maps of Lassa fever in West Africa. *PLoS neglected tropical diseases*, 3(3), p.e388.

FIELDS, B.N. et al., 1993. Fields virology. *Revista do Instituto de Medicina Tropical de São Paulo*, 35, pp.218–218.

Fisher-Hoch, S. et al., 1988. Hematologic dysfunction in Lassa fever. *Journal of medical virology*, 26(2), pp.127–35.

Fisher-Hoch, S.P. et al., 2000. Effective vaccine for lassa fever. *Journal of virology*, 74(15), pp.6777–83.

Fisher-Hoch, S.P. et al., 1985. Pathophysiology of shock and hemorrhage in a fulminating viral infection (Ebola). *The Journal of infectious diseases*, 152(5), pp.887–94.

Fisher-Hoch, S.P. et al., 1987. Physiological and immunologic disturbances associated with shock in a primate model of Lassa fever. *The Journal of infectious diseases*, 155, pp.465–474.

Frame, J.D. et al., 1970. Lassa fever, a new virus disease of man from West Africa. I. Clinical description and pathological findings. *The American journal of tropical medicine and hygiene*, 19(4), pp.670–6.

Frangakis, A.S. et al., 2002. Identification of macromolecular complexes in cryoelectron tomograms of phantom cells. *Proceedings of the National Academy of Sciences of the United States of America*, 99(22), pp.14153–8.

Frank, J., 2006. Electron tomography.. methods for three-dimensional visualization of structures in the cell. *2ed., Springer ISBN 038731234X*.

Froeschke, M. et al., 2003. Long-lived signal peptide of lymphocytic choriomeningitis virus glycoprotein pGP-C. *The Journal of biological chemistry*, 278, pp.41914–41920.

Gallaher, W.R., DiSimone, C. & Buchmeier, M.J., 2001. The viral transmembrane superfamily: possible divergence of Arenavirus and Filovirus glycoproteins from a common RNA virus ancestor. *BMC microbiology*, 1, p.1.

Ganz, T., 2009. Iron in innate immunity: starve the invaders. *Current opinion in immunology*, 21, pp.63–67.

Garcin, D., Rochat, S. & Kolakofsky, D., 1993. The Tacaribe arenavirus small zinc finger protein is required for both mRNA synthesis and genome replication. *Journal of virology*, 67(2), pp.807–12.

Geisbert, T.W. et al., 2005. Development of a new vaccine for the prevention of Lassa fever. *PLoS medicine*, 2(6), p.e183.

Geisbert, T.W., Hensley, L.E., et al., 2003. Pathogenesis of Ebola hemorrhagic fever in cynomolgus macaques: evidence that dendritic cells are early and sustained targets of infection. *The American journal of pathology*, 163(6), pp.2347–70.

Geisbert, T.W., Young, H.A., et al., 2003. Pathogenesis of Ebola hemorrhagic fever in primate models: evidence that hemorrhage is not a direct effect of virus-induced cytolysis of endothelial cells. *The American journal of pathology*, 163(6), pp.2371–82.

Geisbert, T.W. & Jahrling, P.B., 2004. Exotic emerging viral diseases: progress and challenges. *Nature medicine*, 10, pp.S110–S121.

Glaeser, R.M. & Downing, K.H., 1992. Assessment of resolution in biological electron crystallography. *Ultramicroscopy*, 47(1-3), pp.256–65.

Glushakova, S.E., Lukashevich, I.S. & Baratova, L.A., 1990. Prediction of arenavirus fusion peptides on the basis of computer analysis of envelope protein sequences. *FEBS letters*, 269(1), pp.145–7.

Goncalves, A.-R. et al., 2013. Role of DC-SIGN in Lassa virus entry into human dendritic cells. *Journal of virology*, 87(21), pp.11504–15.

González, P.H. et al., 1980. Lymphatic tissue in Argentine hemorrhagic fever. Pathologic features. *Archives of pathology & laboratory medicine*, 104, pp.250–254.

Grimm, R. et al., 1997. Energy filtered electron tomography of ice-embedded actin and vesicles. *Biophysical journal*, 72(1), pp.482–9.

Grünewald, K. et al., 2003. Prospects of electron cryotomography to visualize macromolecular complexes inside cellular compartments: implications of crowding. *Biophysical chemistry*, 100(1-3), pp.577–91.

Günther, S. & Lenz, O., 2004. Lassa virus*.*, *Crit Rev Clin Lab Sci.* 2004;41(4):339-90.

Hall, S.R., 1991. The STAR file: a new format for electronic data transfer and archiving. *Journal of Chemical Information and Modeling*, 31(2), pp.326–333.

Harrison, S.C., 2008. Viral membrane fusion. *Nature structural & molecular biology*, 15(7), pp.690–8.

Hastie, K., 2011. Structure of the Lassa virus nucleoprotein reveals a dsRNA-specific 3’ to 5’ exonuclease activity essential for immune suppression. *Proceedings of the National Academy of Sciences*, 108(6), pp.2396-401.

Hastie, K.M. et al., 2011. Crystal structure of the Lassa virus nucleoprotein-RNA complex reveals a gating mechanism for RNA binding. *Proceedings of the National Academy of Sciences*, 108, pp.19365–19370.

Hastie, K.M. et al., 2011. Structure of the Lassa virus nucleoprotein reveals a dsRNA-specific 3’ to 5' exonuclease activity essential for immune suppression. *Proceedings of the National Academy of Sciences of the United States of America*, 108(6), pp.2396–401.

Heymann, J.B. et al., 2008. Computational resources for cryo-electron tomography in Bsoft. *Journal of structural biology*, 161(3), pp.232–42.

Heymann, J.B. & Belnap, D.M., 2007. Bsoft: image processing and molecular modeling for electron microscopy. *Journal of structural biology*, 157(1), pp.3–18.

Hong, Y.P. et al., 2014. Alignment of low-dose X-ray fluorescence tomography images using differential phase contrast. *Journal of synchrotron radiation*, 21(Pt 1), pp.229–34.

Huang, Z. et al., 2003. Automated determination of parameters describing power spectra of micrograph images in electron microscopy. *Journal of structural biology*, 144(1-2), pp.79–94.

Huiskonen, J.T. et al., 2010. Electron cryotomography of Tula hantavirus suggests a unique assembly paradigm for enveloped viruses. *Journal of virology*, 84(10), pp.4889–97.

Igonet, S. et al., 2011. X-ray structure of the arenavirus glycoprotein GP2 in its postfusion hairpin conformation. *Proceedings of the National Academy of Sciences of the United States of America*, 108(50), pp.19967–72.

Imperiali, B. & O'Connor, S.E., 1999. Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Current opinion in chemical biology*, 3(6), pp.643–9.

Isaäcson, M., 1975. The ecology of Praomys (Mastomys) natalensis in southern Africa. *Bulletin of the World Health Organization*, 52, pp.629–636.

Johnson, K.M. et al., 1987. Clinical virology of Lassa fever in hospitalized patients. *The Journal of infectious diseases*, 155(3), pp.456–64.

Karotki, L. et al., 2011. Eisosome proteins assemble into a membrane scaffold. *The Journal of cell biology*, 195(5), pp.889–902.

Karplus, P.A. & Diederichs, K., 2012. Linking crystallographic model and data quality. *Science (New York, N.Y.)*, 336(6084), pp.1030–3.

Koken, M.H. et al., 1994. The t(15;17) translocation alters a nuclear body in a retinoic acid-reversible fashion. *The EMBO journal*, 13(5), pp.1073–83.

Koster, A.J. et al., 1997. Perspectives of molecular and cellular electron tomography. *Journal of structural biology*, 120(3), pp.276–308.

Kremer, J.R., Mastronarde, D.N. & McIntosh, J.R., 1996. Computer visualization of three-dimensional image data using IMOD. *Journal of structural biology*, 116(1), pp.71–6.

Kunz, S. et al., 2003. Mechanisms for lymphocytic choriomeningitis virus glycoprotein cleavage, transport, and incorporation into virions. *Virology*, 314(1), pp.168–78.

Kunz, S., 2009. Receptor binding and cell entry of Old World arenaviruses reveal novel aspects of virus-host interaction. *Virology*, 387, pp.245–249.

Kyte, J. & Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology*, 157(1), pp.105–32.

De la Torre, J.C., 2009. Molecular and cell biology of the prototypic arenavirus LCMV: implications for understanding and combating hemorrhagic fever arenaviruses. *Annals of the New York Academy of Sciences*, 1171 Suppl, pp.E57–E64.

Lavau, C. et al., 1995. The acute promyelocytic leukaemia-associated PML gene is induced by interferon. *Oncogene*, 11(5), pp.871–6.

Lecompte, E. et al., 2006. Mastomys natalensis and Lassa fever, West Africa. *Emerging infectious diseases*, 12(12), pp.1971–4.

Lee, A.M. et al., 2008. Unique small molecule entry inhibitors of hemorrhagic fever arenaviruses. *The Journal of biological chemistry*, 283(27), pp.18734–42.

Leifer, E., Gocke, D.J. & Bourne, H., 1970. Lassa fever, a new virus disease of man from West Africa. II. Report of a laboratory-acquired infection treated with plasma from a person recently recovered from the disease. *The American journal of tropical medicine and hygiene*, 19(4), pp.677–9.

Lenz, O. et al., 2001. The Lassa virus glycoprotein precursor GP-C is proteolytically processed by subtilase SKI-1/S1P. *Proceedings of the National Academy of Sciences of the United States of America*, 98, pp.12701–12705.

Li, W. et al., 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, 426(6965), pp.450–4.

Li, W. et al., 2006. Animal Origins of the Severe Acute Respiratory Syndrome Coronavirus : Insight from ACE2-S-Protein Interactions MINIREVIEW Animal Origins of the Severe Acute Respiratory Syndrome Coronavirus : Insight from ACE2 – S-Protein Interactions. *Journal of virology.* 2006 May;80(9):4211-9.

Liljeroos, L. et al., 2011. Electron cryotomography of measles virus reveals how matrix protein coats the ribonucleocapsid within intact virions. *Proceedings of the National Academy of Sciences of the United States of America*, 108(44), pp.18085–90.

Lupas, A. et al., 1995. Model structure of the Omp alpha rod, a parallel four-stranded coiled coil from the hyperthermophilic eubacterium Thermotoga maritima. *Journal of molecular biology*, 248(1), pp.180–9.

Mahanty, S. et al., 2003. Cutting edge: impairment of dendritic cells and adaptive immunity by Ebola and Lassa viruses. *Journal of immunology (Baltimore, Md. : 1950)*, 170(6), pp.2797–801.

Mahanty, S. et al., 2001. Low levels of interleukin-8 and interferon-inducible protein-10 in serum are associated with fatal infections in acute Lassa fever. *The Journal of infectious diseases*, 183(12), pp.1713–21.

Mallick, S.P. et al., 2005. ACE: automated CTF estimation. *Ultramicroscopy*, 104(1), pp.8–29.

Marsh, M. & Helenius, A., 2006. Virus entry: open sesame. *Cell*, 124, pp.729–740.

Martinez, M.G., Cordo, S.M. & Candurra, N. a, 2007. Characterization of Junin arenavirus cell entry. *The Journal of general virology*, 88(Pt 6), pp.1776–84.

Martínez-Sobrido, L. et al., 2007. Differential inhibition of type I interferon induction by arenavirus nucleoproteins. *Journal of virology*, 81(22), pp.12696–703.

Martínez-Sobrido, L. et al., 2006. Inhibition of the type I interferon response by the nucleoprotein of the prototypic arenavirus lymphocytic choriomeningitis virus. *Journal of virology*, 80(18), pp.9192–9.

Mastronarde, D.N., 2005. Automated electron microscope tomography using robust prediction of specimen movements. *Journal of structural biology*, 152(1), pp.36–51.

Maurer, U.E. et al., 2013. The structure of herpesvirus fusion glycoprotein B-bilayer complex reveals the protein-membrane and lateral protein-protein interaction. *Structure (London, England : 1993)*, 21(8), pp.1396–405.

Maxfield, F.R., 1982. Weak bases and ionophores rapidly and reversibly raise the pH of endocytic vesicles in cultured mouse fibroblasts. *The Journal of cell biology*, 95(2 Pt 1), pp.676–81.

McCormick, J.B. et al., 1987. A case-control study of the clinical diagnosis and course of Lassa fever. *The Journal of infectious diseases*, 155(3), pp.445–55.

McCormick, J.B. et al., 1986. Lassa virus hepatitis: a study of fatal Lassa fever in humans. *The American journal of tropical medicine and hygiene*, 35(2), pp.401–7.

McCormick, J.B. & Fisher-Hoch, S.P., 2002. Lassa fever. *Current topics in microbiology and immunology*, 262, pp.75–109.

McDowall, A.W. et al., 1983. Electron microscopy of frozen hydrated sections of vitreous ice and vitrified biological samples. *Journal of microscopy*, 131, pp.1–9.

McEwen, B.F., Downing, K.H. & Glaeser, R.M., 1995. The relevance of dose-fractionation in tomography of radiation-sensitive specimens. *Ultramicroscopy*, 60(3), pp.357–73.

McEwen, B.F. & Marko, M., 2001. The emergence of electron tomography as an important tool for investigating cellular ultrastructure. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 49(5), pp.553–64.

Mclay, L. et al., 2013. Targeting virulence mechanisms for the prevention and therapy of arenaviral hemorrhagic fever. *Antiviral Research*, 97, pp.81–92.

Mellquist, J.L. et al., 1998. The amino acid following an asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency. *Biochemistry*, 37(19), pp.6833–7.

Mertens, P.E. et al., 1973. Clinical presentation of Lassa fever cases during the hospital epidemic at Zorzor, Liberia, March-April 1972. *The American journal of tropical medicine and hygiene*, 22(6), pp.780–4..

Mindell, J.A. & Grigorieff, N., 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. *Journal of structural biology*, 142(3), pp.334–47.

Mitrakul, C. et al., 1977. Hemostatic and platelet kinetic studies in dengue hemorrhagic fever. *The American journal of tropical medicine and hygiene*, 26(5 Pt 1), pp.975–84.

Monath, T.P. et al., 1974. Lassa virus isolation from Mastomys natalensis rodents during an epidemic in Sierra Leone. *Science (New York, N.Y.)*, 185(4147), pp.263–5.

Murphy, G.E., Leadbetter, J.R. & Jensen, G.J., 2006. In situ structure of the complete Treponema primitia flagellar motor. *Nature*, 442(7106), pp.1062–4.

Nemeth, E. et al., 2004. IL-6 mediates hypoferremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin. *The Journal of clinical investigation*, 113, pp.1271–1276.

Neuman, B.W. et al., 2005. Complementarity in the supramolecular design of arenaviruses and retroviruses revealed by electron cryomicroscopy and image analysis. *Journal of virology*, 79(6), pp.3822–30.

Nicastro, D. et al., 2006. The molecular architecture of axonemes revealed by cryoelectron tomography. *Science (New York, N.Y.)*, 313(5789), pp.944–8.

Nickell, S. et al., 2007. Structural analysis of the 26S proteasome by cryoelectron tomography. *Biochemical and biophysical research communications*, 353(1), pp.115–20.

Ohkuma, S. & Poole, B., 1978. Fluorescence probe measurement of the intralysosomal pH in living cells and the perturbation of pH by various agents. *Proceedings of the National Academy of Sciences of the United States of America*, 75(7), pp.3327–31.

Orlova, E. V & Saibil, H.R., 2011. Structural analysis of macromolecular assemblies by electron microscopy. *Chemical reviews*, 111(12), pp.7710–48.

Parsy, M.-L. et al., 2013. Crystal structure of venezuelan hemorrhagic Fever virus fusion glycoprotein reveals a class 1 postfusion architecture with extensive glycosylation. *Journal of virology*, 87, pp.13070–5.

Perez, M., Craven, R.C. & de la Torre, J.C., 2003. The small RING finger protein Z drives arenavirus budding: implications for antiviral strategies. *Proceedings of the National Academy of Sciences of the United States of America*, 100(22), pp.12978–83.

Peters, C.J., 2002. Human infection with arenaviruses in the Americas. *Current topics in microbiology and immunology*, 262, pp.65–74.

Philippsen, A., Engel, H.-A. & Engel, A., 2006. The contrast-imaging function for tilted specimens. *Ultramicroscopy*, 107(2-3), pp.202–12.

Pietilä, M.K. et al., 2012. Virion architecture unifies globally distributed pleolipoviruses infecting halophilic archaea. *Journal of virology*, 86(9), pp.5067–79.

Pinschewer, D.D., Perez, M. & de la Torre, J.C., 2003. Role of the virus nucleoprotein in the regulation of lymphocytic choriomeningitis virus transcription and RNA replication. *Journal of virology*, 77(6), pp.3882–7.

Price, M.E. et al., 1988. A prospective study of maternal and fetal outcome in acute Lassa fever infection during pregnancy. *BMJ (Clinical research ed.)*, 297(6648), pp.584–7.

Qi, X. et al., 2010. Cap binding and immune evasion revealed by Lassa nucleoprotein structure. *Nature*, 468, pp.779–783.

Radermacher, M., 1992. Weighted Back-Projection Methods. In J. Frank, ed. *Electron Tomography*. Boston, MA: Springer US, p. pp 91–115.

Radoshitzky, S.R. et al., 2007. Transferrin receptor 1 is a cellular receptor for New World haemorrhagic fever arenaviruses. *Nature*, 446, pp.92–96.

Rath, B.K. et al., 1997. Low-Dose Automated Electron Tomography: A Recent Implementation. *Journal of Structural Biology*, 120(3), pp.210–218.

Reeves, P.J. et al., 2002. Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous N-glycosylation by a tetracycline-inducible N-acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proceedings of the National Academy of Sciences of the United States of America*, 99(21), pp.13419–24.

Rojek, J.M., Sanchez, A.B., et al., 2008. Different mechanisms of cell entry by human-pathogenic Old World and New World arenaviruses. *Journal of virology*, 82(15), pp.7677–87.

Rojek, J.M. & Kunz, S., 2008. Cell entry by human pathogenic arenaviruses. *Cellular microbiology*, 10, pp.828–835.

Rojek, J.M., Perez, M. & Kunz, S., 2008. Cellular entry of lymphocytic choriomeningitis virus. *Journal of virology*, 82(3), pp.1505–17.

Rudd, P.M. et al., 1997. The glycosylation of the complement regulatory protein, human erythrocyte CD59. *The Journal of biological chemistry*, 272(11), pp.7229–44.

Ruprecht, J. & Nield, J., 2001. Determining the structure of biological macromolecules by transmission electron microscopy, single particle analysis and 3D reconstruction. *Progress in biophysics and molecular biology*, 75(3), pp.121–64.

Salazar-Bravo, J., Ruedas, L.A. & Yates, T.L., 2002. Mammalian reservoirs of arenaviruses. *Current topics in microbiology and immunology*, 262, pp.25–63.

Salvato, M.S. et al., 1992. Biochemical and immunological evidence that the 11 kDa zinc-binding protein of lymphocytic choriomeningitis virus is a structural component of the virus. *Virus research*, 22(3), pp.185–98.

Scheres, S.H.W. et al., 2008. Image processing for electron microscopy single-particle analysis using XMIPP. *Nature protocols*, 3(6), pp.977–90.

Schlie, K. et al., 2010. Characterization of Lassa virus glycoprotein oligomerization and influence of cholesterol on virus replication. *Journal of virology*, 84(2), pp.983–92.

Schmid, M.F. et al., 2006. Structure of Halothiobacillus neapolitanus carboxysomes by cryo-electron tomography. *Journal of molecular biology*, 364(3), pp.526–35.

Schmid, M.F. & Booth, C.R., 2008. Methods for aligning and for averaging 3D volumes with missing data. *Journal of structural biology*, 161(3), pp.243–8.

Sevilla, N. et al., 2002. Biology and pathogenesis of lymphocytic choriomeningitis virus infection. *Current Topics in Microbiology and Immunology*, 276, pp.125–144.

Sieczkarski, S.B. & Whittaker, G.R., 2002. Dissecting virus entry via endocytosis. *The Journal of general virology*, 83(Pt 7), pp.1535–45.

Di Simone, C. & Buchmeier, M.J., 1995. Kinetics and pH dependence of acid-induced structural changes in the lymphocytic choriomeningitis virus glycoprotein complex. *Virology*, 209(1), pp.3–9.

Di Simone, C., Zandonatti, M.A. & Buchmeier, M.J., 1994. Acidic pH triggers LCMV membrane fusion activity and conformational change in the glycoprotein spike. *Virology*, 198, pp.455–465.

Sorzano, C.O.S. et al., 2004. XMIPP: a new generation of an open-source image processing package for electron microscopy. *Journal of structural biology*, 148(2), pp.194–204.

Speir, R.W. et al., 1970. Lassa fever, a new virus disease of man from West Africa. IV. Electron microscopy of Vero cell cultures infected with Lassa virus. *The American journal of tropical medicine and hygiene*, 19(4), pp.692–4.

Strecker, T. et al., 2003. Lassa virus Z protein is a matrix protein and sufficient for the release of virus-like particles [corrected]. *Journal of virology*, 77, pp.10700–10705.

Sullivan, B.M. et al., 2011. Point mutation in the glycoprotein of lymphocytic choriomeningitis virus is necessary for receptor binding, dendritic cell infection, and long-term persistence. *Proceedings of the National Academy of Sciences of the United States of America*, 108(7), pp.2969–74.

Takada, A. et al., 2004. Human macrophage C-type lectin specific for galactose and N-acetylgalactosamine promotes filovirus entry. *Journal of virology*, 78, pp.2943–2947.

Terrell, T.G. et al., 1973. Pathology of Bolivian hemorrhagic fever in the rhesus monkey. *The American journal of pathology*, 73, pp.477–494.

Toyoshima, C. & Unwin, N., 1988. Contrast transfer for frozen-hydrated specimens: determination from pairs of defocused images. *Ultramicroscopy*, 25(4), pp.279–91.

Trappier, S.G. et al., 1993. Evaluation of the polymerase chain reaction for diagnosis of Lassa virus infection. *The American journal of tropical medicine and hygiene*, 49(2), pp.214–21.

Volpon, L. et al., 2010. Structural characterization of the Z RING-eIF4E complex reveals a distinct mode of control for eIF4E. *Proceedings of the National Academy of Sciences of the United States of America*, 107(12), pp.5441–6.

Wade, R.H., 1992. A brief look at imaging and contrast transfer. *Ultramicroscopy*, 46(1-4), pp.145–156.

Wade, R.H., 1978. The phase contrast characteristics in bright field electron microscopy. *Ultramicroscopy*, 3(3), pp.329–34.

Walker, D.H. et al., 1982. Pathologic and virologic study of fatal Lassa fever in man. *The American journal of pathology*, 107(3), pp.349–56.

Walter, T.S. et al., 2005. A procedure for setting up high-throughput nanolitre crystallization experiments. Crystallization workflow for initial screening, automated storage, imaging and optimization. *Acta crystallographica. Section D, Biological crystallography*, 61(Pt 6), pp.651–7.

Wang, L.H., Rothberg, K.G. & Anderson, R.G., 1993. Mis-assembly of clathrin lattices on endosomes reveals a regulatory switch for coated pit formation. *The Journal of cell biology*, 123(5), pp.1107–17.

Weis, K. et al., 1994. Retinoic acid regulates aberrant nuclear localization of PML-RAR alpha in acute promyelocytic leukemia cells. *Cell*, 76(2), pp.345–56.

Welsch, S. et al., 2010. Electron tomography reveals the steps in filovirus budding. *PLoS pathogens*, 6(4), p.e1000875.

White, J.M. et al., 2008. Structures and Mechanisms of Viral Membrane Fusion Proteins: Multiple Variations on a Common Theme. *Crit Rev Biochem Mol Biol.* 2008 May-Jun;43(3):189-219.

White, T.A. et al., 2010. Molecular architectures of trimeric SIV and HIV-1 envelope glycoproteins on intact viruses: strain-dependent variation in quaternary structure. *PLoS pathogens*, 6(12), p.e1001249.

Winkler, H. & Taylor, K.A., 2003. Focus gradient correction applied to tilt series image data used in electron tomography. *Journal of structural biology*, 143(1), pp.24–32.

Winn, W.C. et al., 1975. Lassa virus hepatitis. Observations on a fatal case from the 1972 Sierra Leone epidemic. *Archives of pathology*, 99(11), pp.599–604.

Wulff, H., Fabiyi, A. & Monath, T.P., 1975. Recent isolations of Lassa virus from Nigerian rodents. *Bulletin of the World Health Organization*, 52(4-6), pp.609–13.

Xiong, Q. et al., 2009. CTF determination and correction for low dose tomographic tilt series. *Journal of structural biology*, 168, pp.378–387.

York, J. et al., 2005. Genetic analysis of heptad-repeat regions in the G2 fusion subunit of the Junín arenavirus envelope glycoprotein. *Virology*, 343(2), pp.267–74.

York, J. et al., 2004. The signal peptide of the Junín arenavirus envelope glycoprotein is myristoylated and forms an essential subunit of the mature G1-G2 complex. *Journal of virology*, 78, pp.10783–10792.

Yun, N.E. & Walker, D.H., 2012. Pathogenesis of Lassa Fever. *Viruses*, 4, pp.2031–2048.

Zanetti, G. et al., 2009. Contrast transfer function correction applied to cryo-electron tomography and sub-tomogram averaging. *Journal of structural biology*, 168, pp.305–312.

Zanetti, G. et al., 2006. Cryo-electron tomographic structure of an immunodeficiency virus envelope complex in situ. *PLoS pathogens*, 2(8), p.e83.

Zapata, J.C. & Salvato, M.S., 2013. Arenavirus variations due to host-specific adaptation. *Viruses*, 5(1), pp.241–78.

Zheng, S.Q., Sedat, J.W. & Agard, D.A., 2010. Automated data collection for electron microscopic tomography. *Methods in enzymology*, 481, pp.283–315.

Zhu, J. et al., 1997. Three-dimensional reconstruction with contrast transfer function correction from energy-filtered cryoelectron micrographs: procedure and application to the 70S Escherichia coli ribosome. *Journal of structural biology*, 118(3), pp.197–219.

Zhu, P. et al., 2006. Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature*, 441(7095), pp.847–52.

Ziese, U. et al., 2002. Automated high-throughput electron tomography by pre-calibration of image shifts. *Journal of microscopy*, 205(Pt 2), pp.187–200.

Zinkernagel, R.M., 2002. Lymphocytic choriomeningitis virus and immunology. *Current topics in microbiology and immunology*, 263, pp.1–5. .