Molecular determinants of cellular attachment-mediated zoonosis of emerging paramyxo- and arenaviruses.

Rhys Pryce

New College
Division of Structural Biology
Nuffield Department of Medicine
University of Oxford



A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Trinity Term 2018

Molecular determinants of cellular attachment-mediated zoonosis of emerging paramyxo- and arenaviruses.

Rhys Pryce
New College & Division of Structural Biology, University of Oxford
A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy
Trinity Term 2018

Abstract

The spillover of viral pathogens from wild animal reservoirs into human populations is the predominant cause of emerging disease events worldwide and represents an acute and persistent threat to global health, economy, and biodefence. The ability of a virus to specifically attach to cell-surface receptors during host-cell entry is a critical determinant of cross-species transmission.

Through crystallographic and biochemical analysis, this thesis presents a structural rationale for the determinants of differential host-cell receptor recognition, for paramyxo- and arenaviruses, two groups of important emerging viral pathogens. Comparative study of the protein architectures of cellular attachment-mediating glycoproteins from two non-pathogenic viral species—Whitewater Arroyo arenavirus (WWAV) and Cedar henipavirus (CedPV)—with those of highly pathogenic members within the same genera, provides a template for understanding the molecular basis of distinct receptor tropism characteristics.

We demonstrate that the attachment glycoprotein of WWAV, whilst possessing a conserved overall fold, is highly diversified relative to pathogenic arenaviruses, despite the ability to recognise an overlapping repertoire of host-cell entry receptors. Whilst as yet poorly defined auxiliary receptors permit arenaviral entry into human cells *in vitro*, therefore not unequivocally precluding zoonosis of WWAV at the point of cellular attachment, primary receptor utilisation is a correlate of pathogenicity and therefore can account for the non-pathogenic nature of WWAV. In contrast, analysis of the attachment glycoprotein of CedPV reveals no such structural incompatibilities with established henipaviral entry receptors. Thus, these data support the hypothesis that zoonosis of CedPV is not restricted by cellular-attachment and, by extension, host cell attachment alone is insufficient to account for the severe pathogenesis of henipaviruses.

The data and analyses presented herein provide a blueprint for the prediction of receptor tropism characteristics and therein assessment of the zoonotic potential of emerging viruses. Furthermore, in presenting a structure-based classification of the discrete structural states adopted by the conformationally labile arenaviral attachment glycoproteins, we provide a template for improved vaccine design.

Acknowledgements

Acknowledgements

I would firstly like to thank my supervisors, Thomas and Juha, who have been attentive and available throughout my time in STRUBI. It has been a thoroughly rewarding experience, which is in no small part due to you both. I would also like to thank Antra and Ilona for their help, guidance and tuition throughout my project. Thanks also to Karl, Margaret, Ollie, Sai, Tom Walter, Weixian, Weng, and all of our external collaborators for invaluable technical assistance, guidance, and collaboration. It has been great to be a part of the Bowden/Huiskonen groups, both scientifically and socially (**Fig. 1**). STRUBI has been a fantastic place to work, which makes the prospect of moving elsewhere quite scary. It has a lot to live up to.

I would like to thank my STRUBI ~~colleagues~~ friends past and present (Alice, Christiane, Christina, Eamon, Ellie, the Helens, Joy, Lea, Liz, Max, Nayab, Rachel, Rob, Sam, Serban, Steinar, Vicky, and Yasa) for the stimulating, funny, and sometimes pottery related chats. The Wellcome cohort, particularly Jemma, for four years of mutual moaning, fun, fact checking and administration faffing. Finally, I need to thank my long-term housemates Anita and Harshmeena for putting up with me.



Figure 1: The Bowden group. Photograph depicting representative members of the Bowden group. From left to right: Yasa, Liz, Alice, Weng, Thomas, Vicky, Me, Ilona, and Rob.

I feel obliged to also thank my home and uni friends (Alex, Amber, Blair, Brady, Bry, **Carrie**, Charlotte, Dave, Demelza, Jed, Laura B, **Tom**, and Worthy) who, though having absolutely zero impact upon this thesis, have managed to keep me entertained throughout. It would be unfair and impossible to pick favourites. Also, my family, ta.

Table of Contents

Chapter 1: Introduction *8*

1.1 General perspectives: zoonotic viral diseases as a global health threat **8**

1.1.1 Viruses as a global health threat 8
1.1.2 Viral zoonoses 9
1.1.3 General principles of viral attachment and fusion 11
1.1.4 Therapeutic targeting of the viral surface 14
1.1.5 Aims 15

1.2 General henipavirus biology **15**

1.2.1 Henipaviral emergence and disease 15
1.2.2 Newly identified henipaviruses 17
1.2.3 Genome organisation and protein coding strategy 19
1.2.4 Viral architecture and structural proteins 20
1.2.5 Cellular attachment and membrane fusion 21
1.2.6 B-class ephrins as viral receptors 23

1.3 General arenavirus biology **25**

1.3.1 Arenavirus classification, distribution, and disease 25
1.3.2 Notable emergent and newly discovered arenaviruses 28
1.3.3 Disease prevention and treatment strategies 30
1.3.4 Genome organisation and viral architecture 31
1.3.5 Cellular attachment and fusion 32

Chapter 2: Materials & methods *38*

2.1 Molecular biology **38**

2.1.1 Cloning strategy and expression vectors 38
2.1.2 cDNA templates 39
2.1.3 Polymerase chain reaction 40
2.1.4 Primer specifications 40
2.1.5 Agarose gel electrophoresis 41
2.1.6 Enzymatic restriction digests 41
2.1.7 DNA ligation 41
2.1.8 Transformation 42
2.1.9 Plasmid amplification and purification 42

2.2 Protein expression and purification **43**

2.2.1 Rationale and outline workflow 43
2.2.2 Mammalian cell culture 44
2.2.3 Small-scale transient transfection 44
2.2.4 Large-scale transient transfection 45
2.2.5 Western Blot analysis of small-scale expression tests 46
2.2.6 Cell harvesting and diafiltration 47
2.2.7 Nickel-affinity chromatography 47
2.2.8 Protein concentration determination 48
2.2.9 Enzymatic deglycosylation 48
2.2.10 Size-exclusion chromatography 49

2.3 Protein crystallisation **49**

2.4 Structure determination **50**

2.4.1 Crystal harvesting and cryo-protection 50
2.4.2 X-ray data collection and reduction 50
2.4.3 Phase estimation 51
2.4.4 Model building and refinement 52
2.5 Structural analysis 54
2.5.1 Molecular visualisation and presentation 54
2.5.2 Solvent accessibility determination 54
2.6 Structural phylogenetic analysis 54
2.7 X-ray footprinting mass-spectrometry 55
2.7.1 Sample irradiation 55
2.7.2 Mass spectrometry 55
2.8 Circular dichroism spectroscopy 55
2.9 Enzyme-linked immunosorbent assays 56
2.10 Sample preparation for glycan composition analysis 57
2.11 Numerical data representation and statistical analyses 58
Chapter 3: Crystal structure of the Cedar virus attachment glycoprotein 60
3.1 Preface 60
3.2 Summary of findings 60
3.3 Introduction 61
3.4 Expression, purification, and crystallisation of CedPV G 63
3.4.1 Construct design and expression testing 63
3.4.2 Large-scale transient expression and purification 64
3.4.3 Crystallisation 65
3.5 Expression, purification, and crystallisation of CedPV G-receptor complexes 65
3.5.1 Construct design, large-scale transient expression, and purification 65
3.5.2 Crystallisation 69
3.6 CedPV G structure determination 69
3.6.1 Crystal harvesting and X-ray diffraction data collection 69
3.6.2 Crystal structure determination 70
3.7 CedPV G structure analysis 72
3.7.1 Structure overview 72
3.7.2 Comparison with known paramyxoviral attachment proteins 74
3.7.3 Glycosylation of CedPV G and a model for self-association 76
3.7.4 Mapping sequence conservation within CedPV G 79
3.7.5 Structural features of the putative receptor binding site of CedPV G 82
3.7.6 Molecular determinants of differential ephrin tropism 84
3.8 Discussion 88
3.9 Conclusion 91
Chapter 4: Crystal structure of the Whitewater Arroyo virus GP1 93
4.1 Introduction 93
4.2 Summary of findings 94
4.3 Expression, purification, and crystallisation of WWAV GP1 95
4.3.1 Construct design and expression testing 95
4.3.2 Large-scale expression and purification 96

4.3.3 Crystallisation 98

4.4 Expression and purification of transferrin receptor 1 98
4.4.1 Construct design and expression testing 98
4.4.2 Large-scale expression, purification, and complexation 100
4.4.3 Crystallisation trials 101

4.5 WWAV GP1 structure determination 103
4.5.1 Crystal harvesting and X-ray diffraction data collection 103
4.5.2 Crystal structure determination 104

4.6 WWAV GP1 structure analysis 105
4.6.1 Structure overview 105
4.6.2 Disulphide bonding within GP1 108
4.6.3 Comparison with the recently reported WWAV GP1 109
4.6.4 Glycosylation and a model for trimerisation of WWAV GP1 112
4.6.5 The receptor binding surface of WWAV GP1 115
4.6.6 Sequence conservation of reservoir species' TfR1 orthologues 122
4.6.7 Molecular determinants of differential TfR1 tropism amongst WWAV strains 123

4.7 Discussion 128

4.8 Conclusion 130

Chapter 5: 132

Discrete structural and functional states of arenavirus glycoproteins 132

5.1 Preface 132

5.2 Summary of findings 133

5.3 Introduction 134

5.4 Structural transitions and LAMP1 binding of LASV GPC 140
5.4.1 Construct design and small-scale expression testing of LAMP1 140
5.4.2 Large-scale expression and purification of LAMP1 141
5.4.3 The pH and glycan dependence of LAMP1 binding to LASV VLPs 142
5.4.4 pH-induced structural rearrangements of LASV GPC accommodate LAMP1 145

5.5 The pH dependence of isolated WWAV and LORV GP1s 147
5.5.1 Crystallisation of WWAV and LORV GP1 147
5.5.2 The crystal structures of WWAV and LORV GP1 at neutral and acidic pH 148
5.5.3 Structure-based classification of arenavirus GP1s 151
5.5.4 Circular dichroism analysis of WWAV and LORV GP1 154

5.6 Expression and purification of Lujo virus GP1 158
5.6.1 Rationale 158
5.6.2 Construct design, expression, purification, and crystallisation of LUJV GP1 159

5.7 Discussion 161

5.8 Conclusion 168

Chapter 6: Future perspectives and concluding remarks 170

6.1 Future perspectives 170
6.1.1 Structural studies of viral glycoprotein-receptor complexes 170
6.1.2 Towards a method for high-throughput identification of antibody and receptor footprints 172
6.1.3 Biochemical and cellular validation of the determinants of receptor specificity 175

6.2 Concluding remarks 177

Appendix A: List of publications 181

Appendix B: Bibliography.....183
Appendix C: List of Abbreviations.....200

Chapter 1:

Introduction

Chapter 1: Introduction

1.1 General perspectives: zoonotic viral diseases as a global health threat

1.1.1 Viruses as a global health threat

Viruses are obligate intracellular parasites that, as a consequence of their abundance and vast biological diversity, have profound influence on the global biosphere (Koonin et al., 2006, Rohwer et al., 2009). To this end, viruses are likely the most abundant biological entities on earth and are proposed to have impinged upon every form of cellular life (Suttle, 2005, Breitbart and Rohwer, 2005). Whilst not all viruses are pathogenic, they are a leading cause of disease worldwide and represent a persistent threat to global health, economy and biodefense.

Viruses that possess a single-stranded RNA genome are the most common etiological agents of emergent human illness, a feature that is in part due to the utilisation of RNA as an information storage molecule. By comparison with DNA viruses, low-fidelity RNA replication confers enhanced mutation rates upon RNA genomes, which in turn facilitate heightened genetic change and promote genomic economy (Drake and Holland, 1999, Duffy et al., 2008)(**Fig 1.1.1**). In the face of ever-changing environmental conditions, the inaccurate replication of RNA genomes permits rapid adaption to novel pressures (Nichol et al., 2000). Furthermore, most pathogenic RNA viruses are, or were, zoonotic and thus transmitted from animals to humans, in a process that is underpinned by a complex interplay of biological and anthropological factors (Nichol et al., 2000), such as population expansion, environmental encroachment, cultural nuances, and viral adaptation.

Numerous historical and contemporary viral epidemics have highlighted the threat posed to human health by zoonotic viral pathogens. In the centenary of the 1918 influenza pandemic that killed an estimated 50 million people globally (CDC, 2018), Ebola virus (EBOV) continues to emerge sporadically in human populations, following an epidemic in 2014 that resulted in over 11,000 fatalities (from 28,000 suspected infections), despite garnering a

significant global response (Frieden and Damon, 2015). In conjunction with ongoing outbreaks of established human pathogens such as EBOV, Middle East respiratory syndrome coronavirus (MERS-CoV), and Nipah virus (NiV) (WHO, 2018), it is likely that the diversity and magnitude of emerging viral disease events is underrepresented as a result of a paucity of diagnostic tools and misdiagnoses (Pernet et al., 2014).

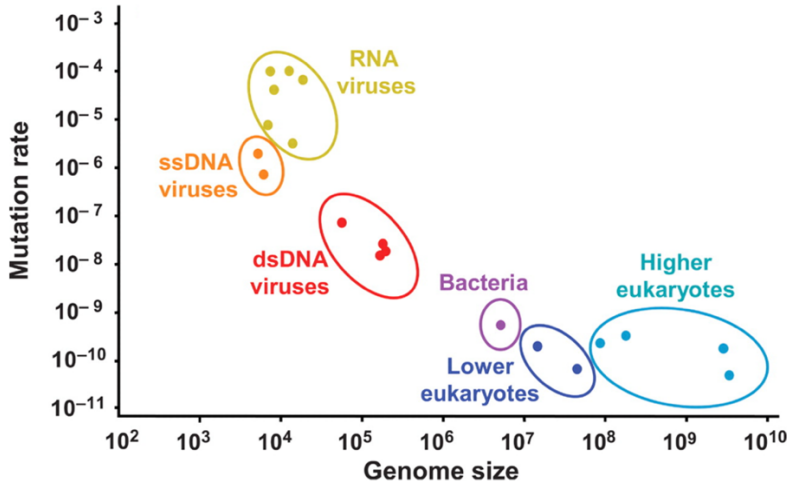


Figure 1.1.1: Genomic size and mutation rates amongst diverse biological entities. Per-site mutation rate versus genome size. Shown from left to right the enteties are: RNA viruses, tobacco mosaic virus, human rhinovirus, poliovirus, vesicular stomatitis virus, bacteriophage Φ6, and measles virus. Single-stranded DNA viruses are bacteriophage ΦX174 and bacteriophage m13. Double-stranded DNA viruses are bacteriophage λ, herpes simplex virus, bacteriophage T2, and bacteriophage T4. The bacterium is *Escherichia coli*. Lower eukaryotes are *Saccharomyces cerevisiae* and *Neurospora crassa*. Higher eukaryotes are *Caenorhabditis elegans, Drosophila melanogaster, Mus musculus*, and *Homo sapiens*. Figure adapted from Gago et al. (2009).

1.1.2 Viral zoonoses

Zoonosis is the process by which pathogenic agents may be reciprocally transmitted between animal and human hosts. Consistent with genetic and therefore biological similarity to human hosts, known reservoirs for zoonotic human pathogens are overwhelmingly mammalian and, to a lesser extent, avian (Morse et al., 2012). However, invertebrate species are also established transmission vectors for important viral pathogens, such as Dengue and Zika virus (Kilpatrick and Randolph, 2012), and are increasingly implicated as a reservoir for an expansive and diverse array of novel RNA viruses (Li et al., 2015). In some instances, following zoonotic transmission,

pathogens such as EBOV, MERS-CoV and NiV are capable of human-to-human spread (Gurley et al., 2007), highlighting the pandemic potential of zoonoses that are not 'self-limiting' in being restricted to cross-species transmission. Notably, some contemporary viruses that exclusively circulate within human populations, such as human immunodeficiency virus (HIV), originated from zoonotic events (Wolfe et al., 2007) (**Fig 1.1.2**).

Exposure to a virus is not the sole prerequisite for subsequent infection. Indeed, the factors that dictate zoonotic potential are numerous. Amongst other things, viruses must: obtain entry to an organism; circumvent host immunological defences; gain cellular entry; replicate both in sufficient quantity and, in some instances, with the aid of specific non-virally encoded maturation factors (Dutch, 2010); and ensure cellular and organismal escape to permit subsequent transmission (Rosenberg, 2015). Many of these factors restrict the effective transmission of viruses between dissimilar species and, to an extent, account for the apparent preponderance of mammalian-borne zoonotic pathogens.

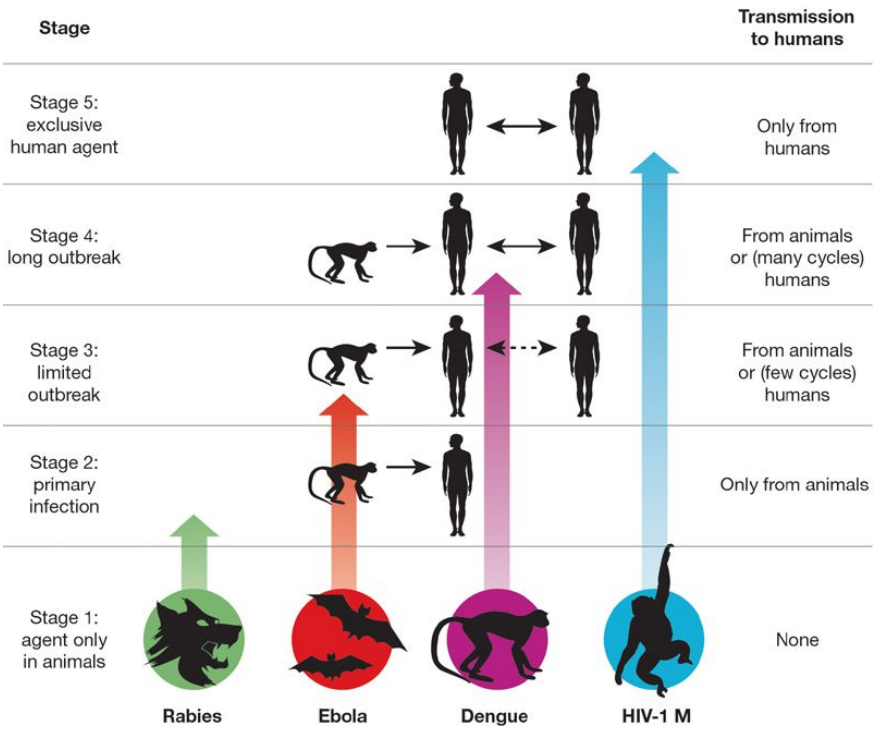


Figure 1.1.2: A five stage classification of zoonotic viruses. Schematic representation the five primary stages of zoonotic infections that range from those still only acquired directly from infected reservoir species (e.g Rabies) to those that exclusively circulate in human populations (e.g HIV). Figure taken from Wolfe et al. (2007).

Of the myriad factors that restrict the permissive host range of viral zoonoses, cellular entry is ubiquitous and of critical importance. As a consequence of genomic economy, RNA viruses typically possess a minimal complement of virally-encoded proteins, of which at least one mediates cellular-attachment. The specificity of an attachment protein for cognate host entry receptors underscores both the species and cell-type tropism of viral pathogens. For example, studies of Influenza A virus, which is responsible for frequent and numerous zoonoses, have demonstrated that mutations within the attachment-mediating hemagglutinin (HA) protein of avian origin viruses confer acquired specificity for human receptors with demonstrable ability to enhance human infectivity (Harvey et al., 2004). Should a virus be unable to mediate cellular attachment, even after successful organismal entry, subsequent cellular ingress is improbable and thus infection is unlikely to be established.

1.1.3 General principles of viral attachment and fusion

As obligate intracellular parasites, viruses must ensure delivery of their genome into permissive host cells to permit subsequent replication. For membrane-bound viruses—such as the arenaviruses and paramyxoviruses that are the focus of this study—cellular entry requires the merger of viral and host-cell lipid bilayers, either at the cell surface or within membrane-bound sub-cellular compartments. Spatial and temporal control of the entry of membrane-bound viruses is tightly regulated in response to environmental stimuli, by viral attachment and fusion proteins.

Preceding cellular entry, viruses must attach to the host cell surface. Although mechanisms of cellular attachment are diverse, viruses typically utilise virion-displayed proteins to bind one or more host-cell receptors. Such receptor-specific attachment is a principal determinant of the species, tissue and cell type tropism of a virus and therein may also underscore other features of the pathobiology. The range of host molecules that serve as viral receptors is diverse and reflects the varied modes of virus entry that they support (Yamauchi

and Helenius, 2013). Host proteins, lipids and glycans are all utilised by a variety of viruses, and amongst other factors, receptor localisation, abundance, and trafficking have profound influence on viral entry and the subsequent escape of progeny (Grove and Marsh, 2011). Indeed, the influenza virus—which mediates cellular attachment through recognition of sialic-acid—encodes a receptor-destroying enzyme (neuraminidase) that is crucial for liberating budding virions from the cell surface (Shtyrya et al., 2009), whereas HIV-1 induces the down-regulation of its cellular receptor (CD4) for the same purpose (Wilen et al., 2012).

Whilst some receptors solely mediate virus accumulation at the cell surface, others, such as transferrin receptor-1 (TfR1), which is utilised by New World arenaviruses, promote endocytosis and facilitate access to specific cellular compartments with unique molecular compositions (Choe et al., 2011). To this end, pathogens such as EBOV and LASV require access to intracellular compartments in order to interact with secondary receptors that are critical entry factors (Carette et al., 2011, Jae et al., 2014). Ultimately, whether alone or in combination, the binding of primary cellular receptors, additional interaction with secondary- or co-receptors, and environmental stimuli, such as low the pH encountered upon endocytic uptake, stimulate diverse signaling cascades that culminate in fusion protein activation and subsequent genome delivery into the cytosol (Carette et al., 2011, Wilen et al., 2012, Jae et al., 2014).

Despite diversity in structure, viral fusion proteins are mechanistically unified in coupling environmental sensation (pH or ligand-binding), to the triggering of large-scale structural rearrangements that bring about the merger of viral and host-cell membranes. Viral fusogens are conventionally grouped into three structurally-related classes (I, II, and III), of which the class I fusion protein of influenza virus (hemagglutinin) is the most extensively studied (Harrison, 2008). Structural studies have provided a wealth of information regarding the relatively long-lived pre- and post-fusion conformations of an array of viral fusion proteins, spanning all three classes (White et al., 2008, Harrison, 2015). However, the dynamic and

metastable nature of fusion proteins has, until recently, precluded structural elucidation of intermediate states (Xu and Wilson, 2011, Halldorsson et al., 2018).

Fusion proteins utilise energy derived from ligand-triggered structural transitions, from a metastable pre-fusion to stable post-fusion conformation, to overcome the kinetic barrier associated with membrane merging, thereby catalysing the process (Harrison, 2015). Following activation, fusion proteins insert a segment of hydrophobic amino acids, the 'fusion peptide' or 'loop', into the opposing cellular membrane. Refolding of the fusion protein brings together the previously opposing fusion peptide and C-terminal transmembrane domains, bringing both viral and host membranes into sufficiently close proximity to permit membrane fusion (**Fig 1.1.3**).

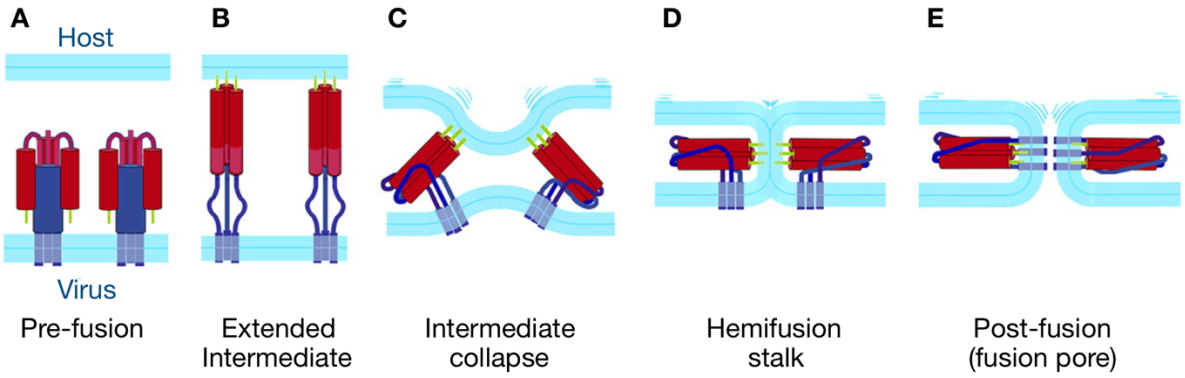


Figure 1.1.3: A general mechanism for viral fusion. Schematic representation of a proposed model for membrane fusion mediated by the type I fusion protein of influenza virus. (A) In the pre-fusion conformation the viral fusion protein (red and blue) adopts a compact and metastable trimeric architecture in which the fusion peptide (green) is shielded. (B) Triggering of the fusion protein occurs in response to environmental cues (e.g pH change or ligand-binding). Activation results in conformational changes in the protein that culminate in the formation of an extended intermediate in which the fusion peptide is inserted into the target membrane. (C) Collapse of the intermediate around the stable trimeric N-terminal regions (red) occurs, bringing the N- and C-terminal (blue) regions of the protein together. (D) Collapse of the extended intermediate brings the opposing membranes together and results in the formation of a 'hemifusion stalk' in which the proximal but not distal membrane leaflets have merged. (E) A fusion pore opens as the distal leaflets fuse and the N- and C-terminal regions further reorganise to form a stable post-fusion conformation. Figure adapted from (Floyd et al., 2008).

Class I fusion proteins, like those utilised by the *Paramyxoviridae* and *Arenaviridae*, are synthesised as inactive precursors that require proteolytic processing into at least two

polypeptides, of which the major C-terminal fragment is the fusogen and the N-terminal fragment is often the attachment protein (Harrison, 2015). A noteworthy exception is the paramyxoviral F protein, which has no dedicated host receptor-binding functionality (Chang and Dutch, 2012). In the primed pre-fusion conformation, the fusion peptide is typically located near the N-terminus and is proposed to insert into target membranes following the formation of an extended intermediate in which the N- and C-termini of the fusogen are located at opposing ends of the molecule (Floyd et al., 2008). The asymmetric collapse of the extended intermediate around a stable N-terminal trimeric core, towards the post-fusion conformation, brings together the termini and therefore the two membranes with which they are associated (Xu and Wilson, 2011) (**Fig 1.1.3**). Unlike fusogens from other classes, the class I proteins are trimeric in both their pre- and post-fusion conformations and form a canonical α-helical coiled-coil or 'six-helix bundle' in the post-fusion state, following productive fusogenic rearrangements.

1.1.4 Therapeutic targeting of the viral surface

Commensurate with the possession of minimal genomes, RNA viruses typically encode a limited number of proteins, a sub-set of which are displayed on the surface of mature virions. As either the sole surface antigen, or one of a limited number, attachment glycoproteins are a key target of the antibody-mediated immune response (Hangartner et al., 2006).

The majority of virus-specific antibodies elicited during viral infection do not neutralise the pathogen (Hangartner et al., 2006, Robinson et al., 2016). A paucity of naturally-elicited neutralising antibodies can be partially attributed to specificity for epitopes that are not accessible on mature virions, or ones that do not sufficiently interfere with protein functionality to mediate effective neutralisation. To this end, protein engineering of viral surface proteins, in order to generate a focused B-cell mediated immune response, has garnered widespread interest (Pantophlet et al., 2003, McLellan et al., 2013, Mann and Ndung'u, 2015, Hastie et al., 2017).

Furthermore, protein engineering has recently been utilised to isolate broadly-neutralising patient-derived antibodies with engineered selecting antigens, designed to exclude selection of highly species-specific epitopes (Clark et al., 2018). Structure-based engineering of viral surface glycoproteins has utility in generating optimised vaccine components for prophylactic control, or therapeutic antibodies for reactive administration. Such rationalised engineering efforts are critically dependent upon high-resolution structural data and thus support a requirement for characterisation of viral surface antigens.

1.1.5 Aims

Here, through crystallographic and biochemical study, we attempt to characterise the molecular determinants of viral zoonosis at the stage of host-cell attachment for two groups of emerging viral pathogens: the arenaviruses and henipaviruses. It is hoped that delineating the structural features that define the differential receptor tropism characteristics of two non-pathogenic viruses, which are close genetic relatives of lethal human pathogens, will provide a template for enhanced understanding of the interplay between receptor tropism and pathogenesis.

Furthermore, in classifying the diversity amongst viral surface antigens, we aim to provide a blueprint for improved therapeutic interventions, such as rationalised drug or therapeutic antibody design, and vaccination strategies.

1.2 General henipavirus biology

1.2.1 Henipaviral emergence and disease

Nipah (NiV) and Hendra (HeV) are zoonotic viral pathogens that constitute the prototypic members of the genus *Henipavirus*, within the family *Paramyxoviridae*. HeV emerged in Brisbane (Australia) in 1994, in an outbreak of severe and fatal respiratory disease in horses and humans (Selvey et al., 1995). NiV subsequently emerged in a much larger outbreak in Malaysia

and Singapore that began in 1999 and resulted in respiratory and mild neurological disease in pigs, and fatal encephalitis in humans (Chua et al., 2000).

The emergence of both HeV and NiV was a consequence of zoonotic transmission from bats of the *Pteropus* genus (flying fox), via intermediate hosts (Eaton et al., 2006, Halpin et al., 2011). The first documented human cases of both HeV and NiV were individuals with occupational exposure to intermediate equine and porcine hosts, respectively. Subsequent NiV outbreaks, however, have evidenced infection resulting from both direct bat-to-human and human-to-human transmission, highlighting the epidemic potential of NiV (Hsu et al., 2004).

Since their emergence, both NiV and HeV have caused frequent and fatal outbreaks of severe respiratory and neurological illness in pigs, cattle, goats, horses, and humans across Australasia, South-East Asia and the Indian sub-continent (Rahman and Chakraborty, 2012). Indeed, since its re-emergence in Bangladesh in 2001, NiV has been the etiological agent responsible for annual disease events associated with high mortality (40-77%) (Hsu et al., 2004, Rahman and Chakraborty, 2012). Since its emergence, HeV has infected over 90 horses, many of which succumbed to the disease or were culled for control purposes, and infected seven humans, four of whom died. Unlike NiV, HeV zoonosis is exclusively documented in cases of spillover from intermediate hosts, principally horses, and though numerous people have been exposed to infected animals, only those who received high levels of exposure appear susceptible to infection (Anon, 2017).

Whilst the geographic distribution of henipaviral spillover and disease events is typically considered to be restricted by the home range of the Pteropus bat reservoir (Halpin et al., 2011), increasing serological and RNA sequencing studies evidence the widespread existence and spillover of novel henipaviruses in Africa (Weiss et al., 2012, Pernet et al., 2014), from both pteropid and non-pteropid reservoirs (**Fig 1.2.1**). Furthermore, antigenic cross-reactivity of henipa-like viruses in Brazil, and the implication of New World bats as reservoir species, extends

the likely distribution of henipaviruses to encompass the Western hemisphere (Drexler et al., 2012, de Araujo et al., 2017).

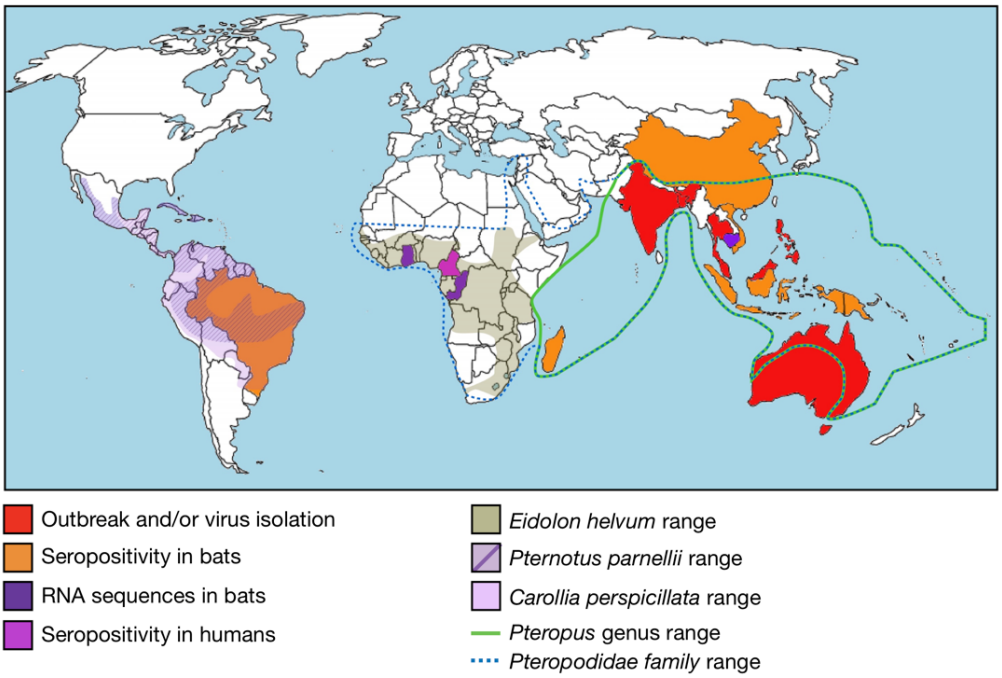


Figure 1.2.1: Evidence for the global distribution of henipaviruses. The global range of implicated henipaviral reservoir bat species and evidence of henipavirus presence. Figure kindly provided by Ms Victoria Avanzato.

Combined with the complete absence of any specific therapeutic or vaccination options, the severe and zoonotic nature of HeV and NiV infection has, uniquely amongst paramyxoviruses, resulted in their classification as Biosafety Level 4 pathogens. Not only do the prototypic henipaviruses represent an acute threat to global human health, but they also present significant economic and biosecurity risks.

1.2.2 Newly identified henipaviruses

Whilst the discovery of the prototypic henipaviruses was a result of their emergence, novel henipa- and henipa-like viruses have recently been detected prior to confirmed human zoonosis. Nucleotide screening of distinct bat populations across five African countries identified at least 23 distinct and novel henipaviral clades (Drexler et al., 2012), and provided

the full genome sequence for a representative virus, termed Ghanaian bat henipavirus (GhV, formerly GH-M74a or Kumasi). GhV has subsequently been subject to structural and functional characterisation, which revealed shared receptor utilisation with the prototypic henipaviruses, despite geographical and genetic distinction (Lee et al., 2015).

The Cedar virus (CedPV), obtained from pteropid bat urine in Queensland (Australia), represents only the third henipavirus to be isolated, after HeV and NiV (Marsh et al., 2012). Whilst the reservoir could not be unambiguously defined, the virus likely derived from *Pteropus alecto* or *Pteropus poliocephalus* colonies from which the urine was collected. Genetic and serological similarity identified CedPV as a novel henipaviral species and, despite close phylogenetic relatedness to HeV and NiV, CedPV appeared to be non-pathogenic in small animal models that are susceptible to henipaviral disease (Marsh et al., 2012).

The severe pneumonia-induced death of three miners in Mójiāng Hani Autonomous County (China) prompted sampling of resident animal species (bat, rat, shrew) in the cave where the eitological agent was likely contracted (Wu et al., 2014). Sampling revealed the presence of a rat-borne henipa-like virus, termed Mójiāng virus (MojV), that may have caused the disease. Identification of MojV in rats represented the first evidence of henipaviruses in a non-domestic, non-bat species and thus MojV challenges the paradigm that wild henipaviral reservoirs are exclusively chiropteran. Genetic, structural, and functional analyses of MojV highlight significant divergence from the prototypic henipaviruses (Rissanen et al., 2017), again evidencing the diversity amongst the genus and increasing the number of known reservoir species.

The continued discovery of novel henipaviruses underscores the global distribution and diversity of the genus, and thus highlights henipaviruses as an important group of zoonotic pathogens with an undetermined, and possibly underappreciated risk to human health.

1.2.3 Genome organisation and protein coding strategy

The henipavirus genus is a constituent of the *Paramyxovirinae* sub-family within the *Paramyxoviridae* family (Lefkowitz et al., 2018) (**Fig 1.2.3i**). The characteristic single-stranded, negative-sense RNA genome of henipaviruses encodes six major genes: nucleocapsid (N), phosphoprotein (P), matrix (M), fusion (F), attachment (G) and RNA-dependent RNA polymerase (L) (**Fig 1.2.3ii**).

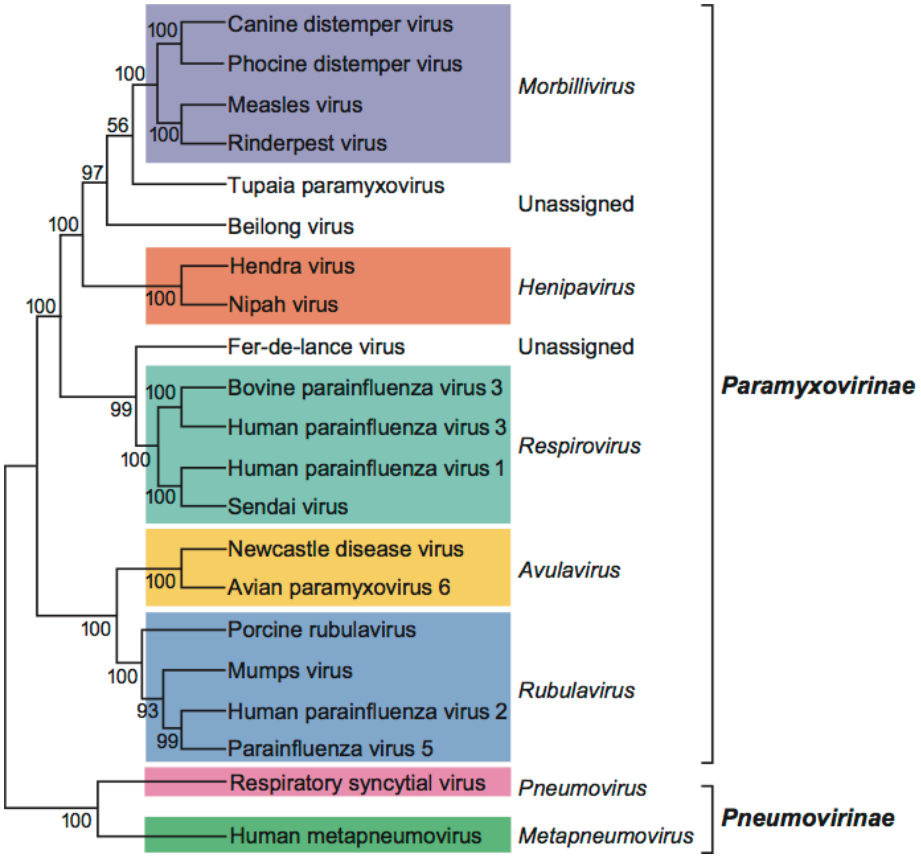


Figure 1.2.3i: Phylogenetic relationships within the *Paramyxoviridae* family. Phylogenetic analysis of the L proteins of members of the family *Paramyxoviridae*. Phylogenetic analysis was performed using MEGA4.1 on the amino acid sequence of the L proteins from representative members of the family *Paramyxoviridae*. Figure adapted from Lefkowitz et al. (2018).

Like other members of the *Paramyxovirinae*, the henipavirus P gene encodes multiple unique protein transcripts utilising a combination of an RNA editing mechanism and an alternate start codon (Thomas et al., 1988). The insertion of non-templated G nucleotides at a highly conserved editing site yields transcripts from alternate reading frames and results in the

production of two non-structural proteins, V and W, whilst an alternate start codon yields C, all of which are known to antagonise the innate anti-viral interferon response of the host (Park et al., 2003) (**Fig 1.2.3ii**). Intriguingly, whilst both HeV and NiV retain the ability to produce four unique protein transcripts from the P gene, CedPV appears to lack both RNA editing capability and the coding capacity for a V or W transcript (Marsh et al., 2012). Given the established immunomodulatory function of both the V and W proteins, their absence in CedPV may play a critical role in dictating the apparently non-pathogenic phenotype therein (Satterfield et al., 2015).

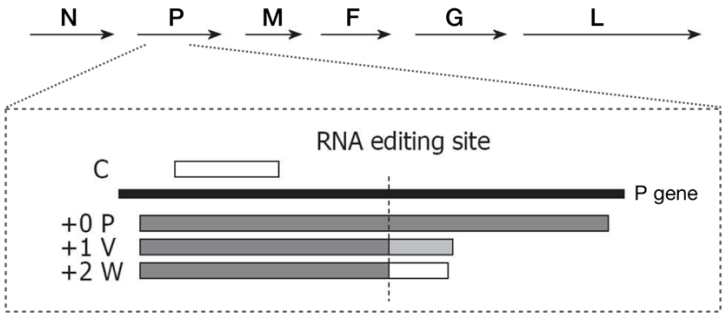


Figure 1.2.3ii: HeV and NiV genome and coding strategy. Schematic representation of the genome organisation and coding strategy of prototypic henipaviruses. The nucleoprotein (N), phosphoprotein (P), matrix (M), fusion (F), attachment (G) and polymerase (L) are displayed. The RNA editing and alternate start codon coding strategy for producing multiple transcripts from P is displayed in an expanded box. The number of non-templated G nucleotides necessary for P V and W transcript generation is indicated. Figure adapted from Audsley and Moseley (2013).

1.2.4 Viral architecture and structural proteins

Henipaviral virions are pleomorphic and enveloped by a phospholipid bilayer obtained during budding at the host-cell surface (Hyatt et al., 2001). The virion envelope is decorated with two virally-encoded transmembrane glycoprotein antigens: G and F. Whilst G and F have some demonstrable budding properties alone, the matrix protein appears essential to the assembly of new virions and ensuring genome incorporation (Patch et al., 2007) (**Fig 1.2.4**).

Encapsulated within the envelope is the linear ribonucleoprotein (RNP) core, which comprises the negative-sense single-stranded RNA genome that is intimately associated with the N protein, in a strict ratio of one N protein per six nucleotides (Halpin et al., 2004). The L

and P proteins also associate with the nucleoprotein complexes and are required to form the transcription competent transcriptase complex (N,P,L) which is responsible for both mRNA transcript generation and genome replication (Eaton et al., 2004).

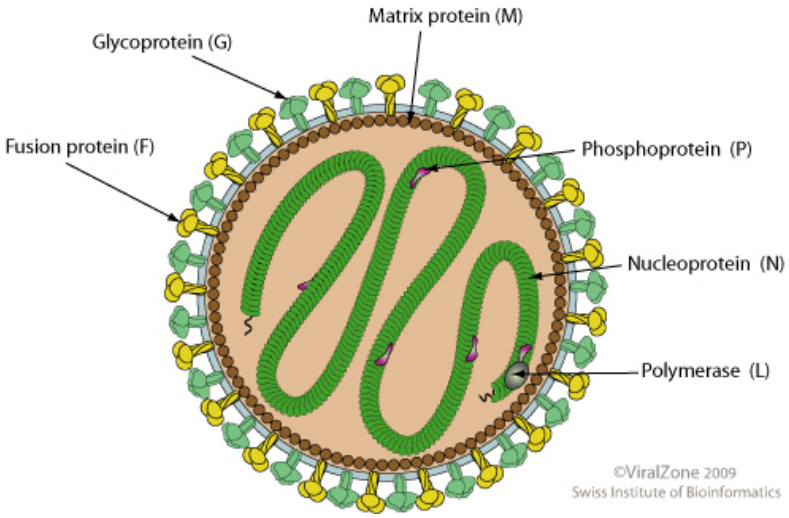


Figure 1.2.4: Ultrastructure and organisation of henipavirus virions. Schematic representation of henipavirus virions. Each structural protein is displayed. The transmembrane fusion (F) and attachment (G) glycoproteins decorate the virion surface. The matrix (M) protein underlies the host-derived phospholipid membrane. The single-stranded negative-sense RNA genome is associated with the nucleoprotein (N), which when further associated with the polymerase (L) and phosphoprotein (P) to form transcription- and replication-competent transcriptase complexes. Image obtained from Viral Zone (online, acceded 26/09/2018) available from: https://viralzone.expasy.org/85?outline=all_by_species

1.2.5 Cellular attachment and membrane fusion

Henipavirus cellular entry is mediated by the concerted interactions and structural transitions of the envelope-displayed G and F glycoproteins (Chang and Dutch, 2012). The henipaviral attachment glycoprotein is a homo-tetrameric type-II transmembrane glycoprotein composed of four distinct domains: a short N-terminal intraviral domain; a single-pass transmembrane domain; a principally α-helical stalk (Yuan et al., 2005); and a C-terminal globular receptor-binding domain (Steffen et al., 2012, Xu et al., 2015).

The F protein is a homo-trimeric class-I transmembrane glycoprotein that is synthesised as an immature precursor (F_0). F_0 requires post-translational cleavage by the non-virally encoded protease cathepsin-L, yielding F_1 and F_2 that remain covalently associated through a

disulphide bond (Pager and Dutch, 2005, Pager et al., 2006). Like other class I membrane fusion proteins, F undergoes thermodynamically favourable large-scale structural rearrangements from a metastable pre-fusion conformation to a post-fusion conformation (Dutch, 2010). Distinct from some class I fusion proteins, such as the GP2 of the *Arenaviridae*, fusion activation is pH-independent and is intimately coupled to receptor recognition by a cognate G protein. Furthermore, the crystal structure of the pre-fusion state of NiV F revealed higher-order oligomerisation of F proteins, suggestive of an additional level of concerted structural changes during F triggering (Xu et al., 2015). The prevailing model of henipaviral entry implicates the coupling of receptor engagement by G to allosteric modulation that directly triggers F. Additionally, the oligomeric self-association of F may further transduce an allosteric signal for triggering that together mediate the spatially and temporally regulated activation of F and, therefore, the initiation of cellular entry (Xu et al., 2015).

Though a wealth of biochemical data exist, implicating both the globular head and stalk domains as critical determinants of paramyxoviral F activation by G, none are conclusive, nor do they universally account for idiosyncrasies observed across different paramyxoviruses (Steffen et al., 2012, Bose et al., 2015). However, integrated functional and structural data support a stalk-exposure model in which receptor binding by the globular head of G triggers exposure of the G stalk domain (Yuan et al., 2005, Yuan et al., 2011), which has been shown to be essential to, and in some instances sufficient for, F triggering (Bose et al., 2014) (**Fig 1.2.5**).

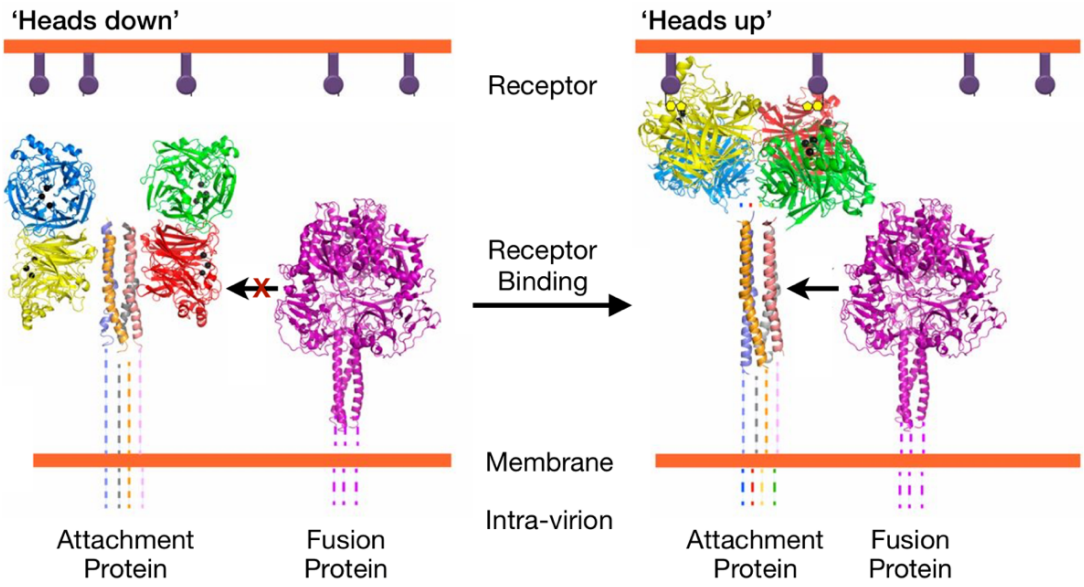


Figure 1.2.5: The stalk-exposure model of F activation. Schematic representation of the stalk-exposure model for F activation by the attachment-mediating glycoprotein of paramyxoviruses. In the unliganded 'heads down' state the globular domain of the attachment protein shields the stalk domain. Upon receptor binding the head domains move to an upward 'heads up' conformation which exposes the stalk domain for association with and triggering of F. Adapted from Bose et al. (2014)

1.2.6 B-class ephrins as viral receptors

The henipavirus attachment glycoprotein utilises a conserved protein architecture to mediate cellular entry through binding of the B-class ephrins (Negrete et al., 2005, Bowden et al., 2010b). Ephrins are membrane-bound proteinaceous ligands that participate in bidirectional cellular signaling through cognate interaction with the eph family of cell-surface displayed tyrosine-kinases (Lisabeth et al., 2013). The eight ephrins encoded by the human genome are subdivided into two classes, A and B, that associate with cellular membranes via glycophosphatidylinositol (GPI) anchors or single-pass transmembrane helices, respectively (1997). The A-class ephrin ligands comprise ephrin-A1 to ephrin-A5, and the B-class, ephrin-B1 to ephrin-B3.

Whilst the prototypic henipaviruses, HeV and NiV use both ephrin-B2 and ephrin-B3 for cellular entry, CedPV and GhV are only known to utilise ephrin-B2 (Negrete et al., 2005, Marsh et al., 2012, Lee et al., 2015). Despite the ephrin-specific binding properties of the

different henipaviral G proteins, only minor alterations to ephrin-B1 are required to permit its use as a functional entry receptor for NiV (Negrete et al., 2006). This suggests that similarly subtle changes to henipaviral G proteins, achievable through only modest diversification, would permit differential ephrin utilisation. To this end, the potential for diversity in henipaviral receptor utilisation was recently evidenced by MojV, which is incapable of mediating cellular entry through any of the established paramyxoviral receptors (Rissanen et al., 2017).

Ephrins function endogenously in numerous and diverse biological processes that include axonal guidance and angiogenesis, and thus are widely expressed on numerous cell types (Pasquale, 2008). Such widespread expression, particularly on epithelial and neuronal cell types, is consistent with the vascular and neurotropic disease pathologies associated with acute HeV and NiV infection (Ang et al., 2018). Additionally, the widespread and highly conserved nature of ephrin ligands across animal species accounts for the large host rage of henipaviruses. Furthermore, ephrins are highly conserved to the extent that orthologues from the zoonotic reservoir (bats), and amplifier hosts (horses and pigs), share over 95% identity to the respective human protein (Pernet et al., 2012). Notably, the GH loop that is critical for G-binding is identical amongst the ephrin-B2 and -B3 orthologues of these species (Bowden et al., 2008a).

Functional utilisation of B-class ephrins by henipaviral G proteins involves the formation of an extensive protein-protein interface centered around the 'GH loop' of the ephrin molecule (Bowden et al., 2008a). Though the recognition modes are distinct, the GH loop of ephrins is also the principal site of recognition of ephrin ligands by their endogenous receptors, eph tyrosine-kinases (Bowden et al., 2009a). Given the overlapping interfaces utilised by both henipaviral G proteins and eph family tyrosine-kinases, development of therapeutic antivirals capable of targeting ephrin ligands appears less promising than targeting the viral G proteins themselves, principally due to the potential disruption of numerous homeostatic processes in which the eph-ephrin signaling axis is influential. Indeed, disrupting endogenous eph-ephrin

signaling through the use of any notional ephrin-targeting antiviral would likely necessarily induce significantly more off-target effects than effective targeting of viral G proteins.

1.3 General arenavirus biology

1.3.1 Arenavirus classification, distribution, and disease

The *Arenaviridae* family comprises two genra: the reptarena- and mammarenaviruses, so named due to their association with reptilian and mammalian hosts, respectively. The mammarenavirus comprises 35 recognised viral species that are further subdivided into two distinct groups based upon antigenic relatedness: the Tacaribe serocomplex (or New World (NW)) and the Lassa-Lymphocytic choriomeningitis serocomplex (or Old World (OW) arenaviruses) (Lefkowitz et al., 2018) (**Fig 1.3.1i**). Whilst no cladistic distinction is currently defined for OW viruses, the NW arenaviruses are sub-divided into four clades: A, B, C, and D (formerly termed A/B or A/rec).

Mammarenaviruses are typically associated with rodent hosts, which are the natural viral reservoirs, with the exception of Tacaribe virus (TACV), which is associated with Artibeus bats (Downs et al., 1963). However, despite no evidence of a rodent reservoir for TACV, the induction of fatal disease by experiential infection of Artibeus bats with TACV suggests that they are unlikely to be the *bona fide* natural reservoir (Cogswell-Hawkinson et al., 2012). Though phylogenetic and geographical isolation of arenaviral species and their natural hosts, or close relatives thereof, is widely thought to be suggestive of co-speciation, the phylogenetic incongruence of the viral and host lineages of two recently discovered Asiatic arenaviruses challenge this hypothesis (Bowen et al., 1997, Blasdell et al., 2016). In addition to their genetic and serological divergence, the OW and NW arenaviruses are geographically distinct, limited to the OW (Europe, Africa, and Asia) and NW (the Americas), in a distribution that is concomitant with the home range of their reservoir species. The prototypic OW virus,

lymphocytic choriomeningitis virus (LCMV), however, is unique in exhibiting a worldwide distribution that is a consequence of infecting the common house-mouse.

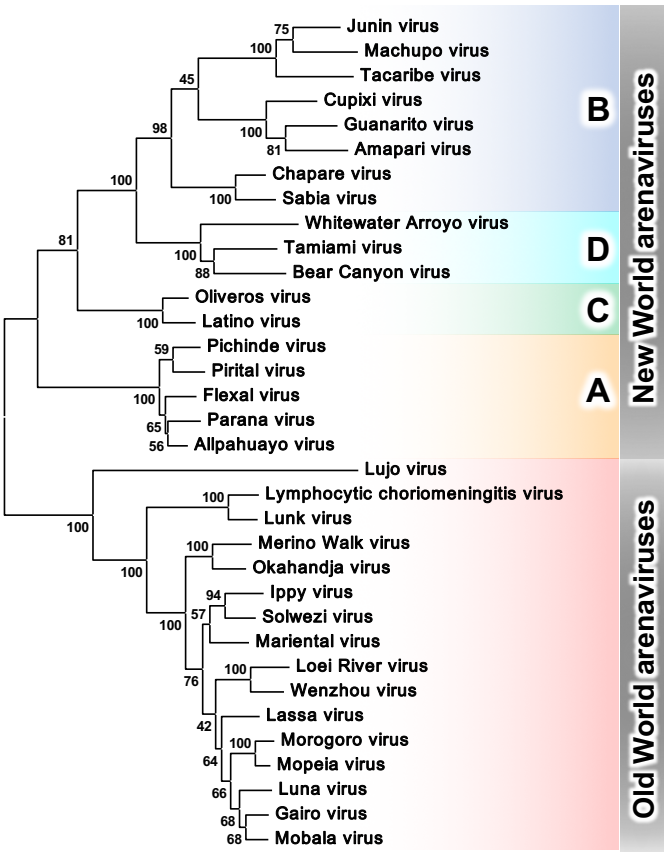


Figure 1.3.1i: Genetic phylogeny of the arenaviral GPC. A maximum-likelihood phylogeny of 34 arenaviral glycoprotein complex (GPC) sequences divides the family into Old World and New World (Clades A–D) groupings. Evolutionary analyses were performed using MEGA 7.

Both serogroups persist in rodents belonging to the family *Muridae,* with NW viruses infecting species of the *Sigmodontinae* sub-family, and OW viruses infecting the sub-family *Murinae.* Reservoir species exhibit chronic and largely asymptomatic infection that is propagated amongst rodents through both vertical and horizontal transmission (Mills et al., 1991). Sporadic spillover into human populations in close proximity to reservoir species is primarily mediated by cross-contamination of food products or aerosolisation of rodent excreta (McLay et al., 2014).

Though usually restricted to geographically constrained foci, arenaviral spillover events are broadly distributed, frequent, and often fatal (Gonzalez et al., 2007) (**Fig 1.3.1ii**). At least ten arenaviruses are associated with overt disease in humans, most of which cause fatal hemorrhagic fever (Moraz and Kunz, 2011). Whilst Chapare, Guanarito, Machupo, and Sabiá are all responsible for zoonotic hemorrhagic fevers in humans, their occurrence is relatively infrequent and small in scale (Lisieux et al., 1994, Delgado et al., 2008). Of the NW viruses, Junín (JUNV) —the causative agent of Argentine hemorrhagic fever— has posed the greatest threat to human health. Since its emergence in 1958, and prior to prophylactic interventions, JUNV infected between 100 and 800 people annually, who were principally agricultural workers in the endemic region of Argentina (Harrison et al., 1999). Case fatality of untreated patients with Argentine hemorrhagic fever ranges from 15%-20% and is associated with necrotic pathologies within vascular and neurological tissues (Shao et al., 2015).

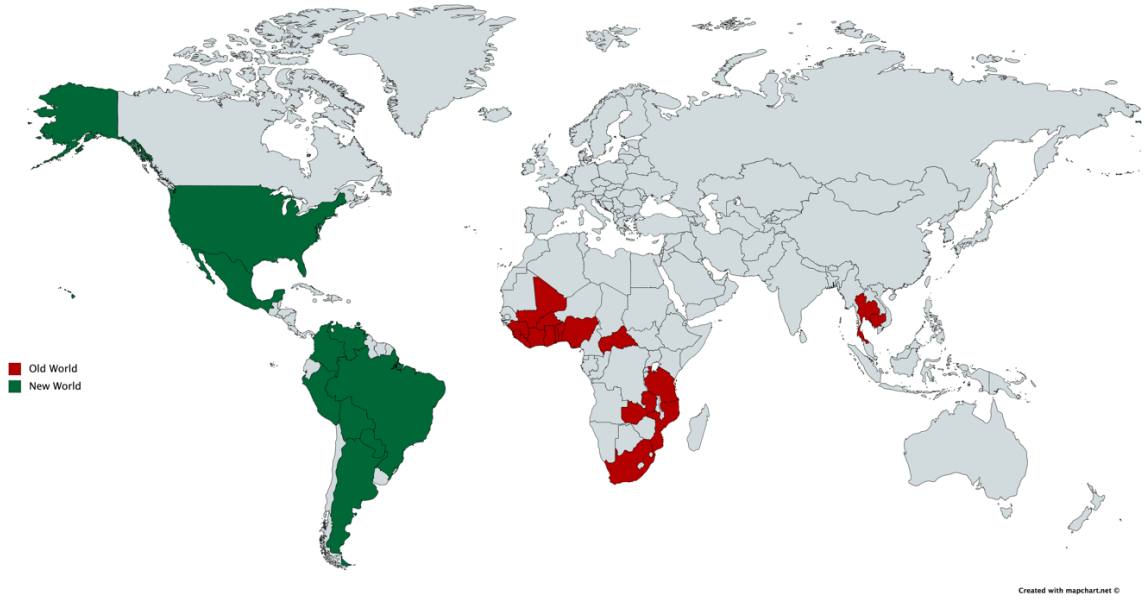


Figure 1.3.1ii: Global distribution of known arenaviral species. Global instances of arenaviruses. Contries with reported instances of any one of: RNA detection, serological evidence, virus isolation and human infection, are coloured. Regions are coloured according to NW (green) and OW (red) serogroups. Data integrated from numerous sources cited in the main text. LCMV is not considered due to its global distribution.

Unlike the restricted geographical distribution of most arenaviruses, the worldwide distribution of LCMV brings the virus into contact with a vast number of people. Whilst approximately 5% of people globally have signs of previous LCMV exposure, overt disease in immune-competent adults is extremely rare (Peters, 2006). Despite this, LCMV is associated with higher mortality in instances of congenital transfer and solid-organ transplantation, and may be underdiagnosed (Barton and Hyndman, 2000). Thus, many consider LCMV a neglected human disease agent.

Lassa virus (LASV) is responsible for the greatest human health burden of all arenaviruses (Gunther and Lenz, 2004). In endemic regions of West Africa, LASV causes up to 300,000 infections annually, of which 5,000-10,000 are fatal (Lukashevich et al., 1993, Shao et al., 2015). The symptoms of LASV infection are diverse and range in severity from mild fever or malaise to severe neurological and vascular pathologies that result in death (Houlihan and Behrens, 2017). Due to the diversity of symptoms and their occurrence in pathogen-rich, but relatively resource-poor, locations, LASV may often be misdiagnosed and thus its human impact underestimated.

1.3.2 Notable emergent and newly discovered arenaviruses

Since the discovery of LCMV in 1933, numerous novel arenaviruses have been reported, as a consequence of both an association with human disease and direct efforts to isolate novel viruses. In 2008 an outbreak of viral hemorrhagic fever, likely acquired in Zambia, killed four people of five infected, three of whom contracted the virus by direct contact with the index patient and one by tertiary infection (Paweska et al., 2009). Subsequently, the etiological agent was identified as a novel OW arenavirus, named Lujo (LUJV) (Briese et al., 2009). LUJV is only the second characterised hemorrhagic fever causing OW arenavirus, however, it is genetically distinct from all known arenaviruses and forms a unique phylogenetic branch off the ancestral node of the OW viruses (Briese et al., 2009) (**Fig 1.3.1**). Consistent with its genetic separation

from the known arenaviral groups, LUJV displays idiosyncratic usage of novel cellular entry receptors that are not shared with any known arenaviruses (Raaben et al., 2017).

In a recent survey for arenaviral RNA in rodent species in Cambodia, Thailand and Laos, two novel mammarenaviral species, Wēnzhōu (WENV) and Loei River (LORV), were identified (Blasdell et al., 2016). Phylogenetic relatedness to the LASV group of viruses, and association with OW rodent hosts, places both viruses in the OW serogroup. Though not conclusive, Wēnzhōu was linked to mild illness in humans (Blasdell et al., 2016). Thus, aside from LCMV, which is unusual in its world-wide distribution, WENV and LORV extend the known global range of arenaviruses to Asia and evidence the potential for zoonotic spillover therein.

Whitewater arroyo virus was isolated in New Mexico (USA) in 1996 in a targeted search for an arenavirus associated with the white-throated woodrat (*Neotoma albigula*). Phylogenetic and serological characterisation placed WWAV in the NW serogroup, in what is now known as clade D, most closely related to the founding member of this clade, Tamiami virus (TAMV) (Fulhorst et al., 1996). Though initially not associated with human disease, evidence of human infection by a WWAV or a WWAV-like virus has been observed and implicated as the likely cause of three fatalities in California (USA) (Enserink, 2000). Furthermore, recent serological studies also evidence WWAV (or WWAV-like) infection in humans across North America (Milazzo et al., 2011). While the prototypic WWAV strain (AV9310135) is incapable of utilising the human orthologue of the receptor that is responsible for the cellular entry of the closely related and pathogenic clade B viruses, a WWAV-like virus (AV96010151) can mediate cellular entry similar to the clade B viruses. It is unclear whether AV96010151 was the etiological agent of the Californian fatalities, however they nevertheless highlight the existence of WWAV strains with zoonotic potential, at least at the stage of cellular entry. Whilst serological evidence suggests the zoonosis of WWAV or WWAV-like viruses is infrequent and at low levels, it does highlight the existence of zoonotic hemorrhagic fever causing arenaviruses within clade D and

further extends the range of fatal arenaviral disease agents to encompass Central and North America (Milazzo et al., 2011).

1.3.3 Disease prevention and treatment strategies

Prophylactic and therapeutic interventions for the prevention and treatment of arenaviral disease are severely limited. Despite this, some have demonstrable efficacy, and though scarce, are more effective and numerous than the options available for henipaviral disease.

The off-label use of the broad-spectrum nucleoside analogue, ribavirin, is currently the only antiviral treatment option for arenaviral infection (Moreno et al., 2011). Amongst other mechanisms, ribavirin is incorporated into RNA as a consequence of its chemical analogy with adenine and guanine, thereby reducing the fidelity of viral replication to an extent that precludes effective life-cycle continuation. Due to the generalised mode of action of ribavirin, it has been used to treat numerous RNA and DNA viral diseases, including NiV, in an outbreak in which experimental ribavirin administration resulted in a 36% reduction in case fatalities amongst patients with acute encephalitis (Chong et al., 2001). Despite demonstrating efficacy in treatment during LASV infection (McCormick et al., 1986), ribavirin treatment is restricted to early infection and thus has limited utility (Carrillo-Bustamante et al., 2017). Combined, the limited efficacy, small window of opportunity and occasionally adverse side effects delineate ribavirin as a sub-optimal therapy (Rusnak et al., 2009).

The only effective prophylactic intervention for arenaviral disease is a live-attenuated JUNV vaccine, Candid #1 (Barrera Oro and McKee, 1991). Though not FDA approved, Candid #1 has a demonstrated protective efficacy of 95% and, as a consequence of targeted administration in endemic regions, has successfully reduced instances of Argentine hemorrhagic fever (Maiztegui et al., 1998). Despite evidence of cross-protection against MACV infection, the use of Candid #1 is limited to the endemic regions of Argentina (McLay et al., 2014). Furthermore, the success of Candid #1 has not been replicated for any other arenaviral disease and it remains

the only arenaviral vaccine. The success of Candid #1 and the ability of convalescent patient sera to effectively prevent and control JUNV disease, respectively, suggest an efficient B-cell response is able to control JUNV infection (Ruggiero et al., 1986). The efficient antibody-mediated control of JUNV therefore evidences the utility of structural studies and protein engineering for defining and modulating the antigenic properties of arenaviral surface antigens (Zeltina et al., 2017, Clark et al., 2018).

1.3.4 Genome organisation and viral architecture

Arenaviruses possess a bi-segmented RNA genome that encodes four principal polypeptides with an ambisense coding strategy. The large (L) segment encodes the RNA-dependent RNA polymerase (L) and the matrix protein (Z), separated by an intergenic region that possesses secondary structure that is implicated in transcription termination and mRNA stability (Pinschewer et al., 2005). The small genomic segment (S) encodes the nucleoprotein (NP) and the glycoprotein precursor (GPC), similarly separated by an intergenic region. L and NP are encoded in the negative-sense direction and therefore the corresponding mRNAs are transcribed directly from genomic RNA. The Z and GPC genes are encoded in the positive-sense orientation and are transcribed from anti-genomic RNAs.

Arenaviral virions are enveloped, pleomorphic and decorated by a single glycoprotein complex, GPC, which is randomly distributed across the surface (Li et al., 2016) (**Fig 1.3.4**). The Z protein is the principal mediator of arenavirus budding and is known to interact with GPC, potentially for recruitment to the site of viral budding (Perez et al., 2003, Capul et al., 2007). The genomic viral RNA is encapsidated by the N protein, which mediates genome incorporation into progeny virions through interaction with the Z protein (Loureiro et al., 2012). Interestingly, the C-terminal region of the N protein possesses novel endonuclease activity that mediates immune suppression by catalysing the digestion of dsRNA which, as a common pathogen associated molecular pattern (PAMP), is known to stimulate anti-viral immune effectors (Hastie

et al., 2011). The L protein, in conjunction with RNA-bound N protein, constitutes the minimal functional unit required for transcription and genome replication (Lee et al., 2000). Thus, as might be expected from entities with such minimal genomes as arenaviruses, the proteins encoded therein are highly functionalised and mediate diverse interactions both with virally- and host-encoded macromolecules.

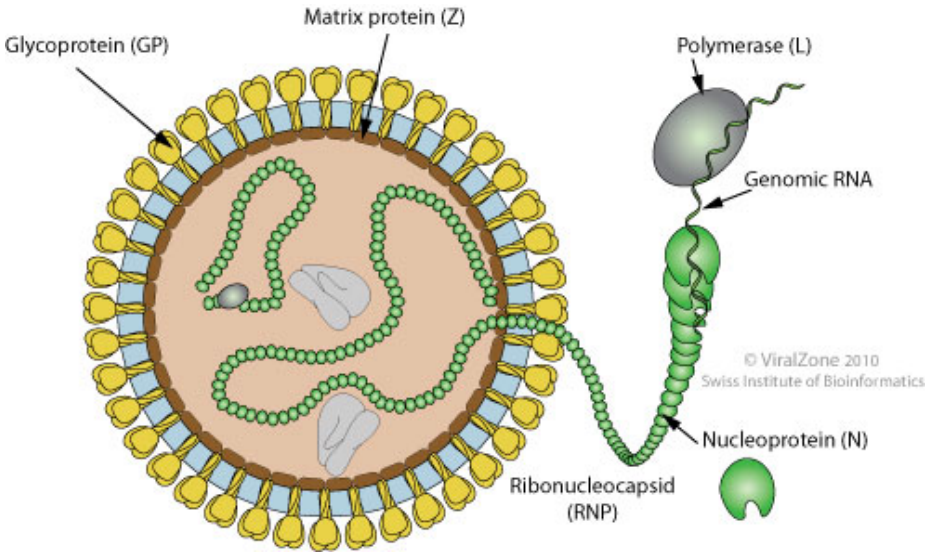


Figure 1.3.4: Arenaviral architecture. Schematic representation of the arenavirus virion. The enveloped virions are decorated with trimeric glycoprotein spikes (termed GP or GPC), that are in contact with an underlying layer of matrix protein (M). The nucleoprotein (N) encapsidates the bi-segmented genome and together with the RNA-dependent RNA polymerase (L) forms the minimal unit required for transcription and replication. Arenavirus virions often contain host cell ribosomes obtained during viral budding (shown in grey), which contributed to a grainy or sandy appearance that resulted in the *Arena* nomenclature (arena deriving from the Latin word for sand). Image available from (https://viralzone.expasy.org/501?outline=all_by_species; accessed 25/09/2018).

1.3.5 Cellular attachment and fusion

GPC is the sole proteinaceous antigen present on arenavirus virions. GPC is therefore responsible for orchestrating host recognition and subsequent fusion of the viral and host-cell envelopes, in a process that couples environmental pH sensing to both subtle and large-scale structural rearrangements of itself and the target cellular membrane (Nunberg and York, 2012).

GPC is expressed as a polyprotein precursor that requires numerous post-translational modifications for complete maturation (**Fig 1.3.5**). The highly conserved and unusually long (58

amino acids) stable signal peptide (SSP) is cleaved by the host signal peptidase, after which it is retained as an integral part of the GPC complex (York and Nunberg, 2016). SSP has been demonstrated to be important for interactions with the Z protein and is indispensable for the pH-induced membrane fusion activity of the mature GPC complex. Though no structural data exist for SSP, it is predicted to form a hairpin-like structure comprising two transmembrane-spanning domains separated by a short extra-viral loop, a myristoylated N-terminus, and an unmodified C-terminus, both on the viral interior (Shao et al., 2016).

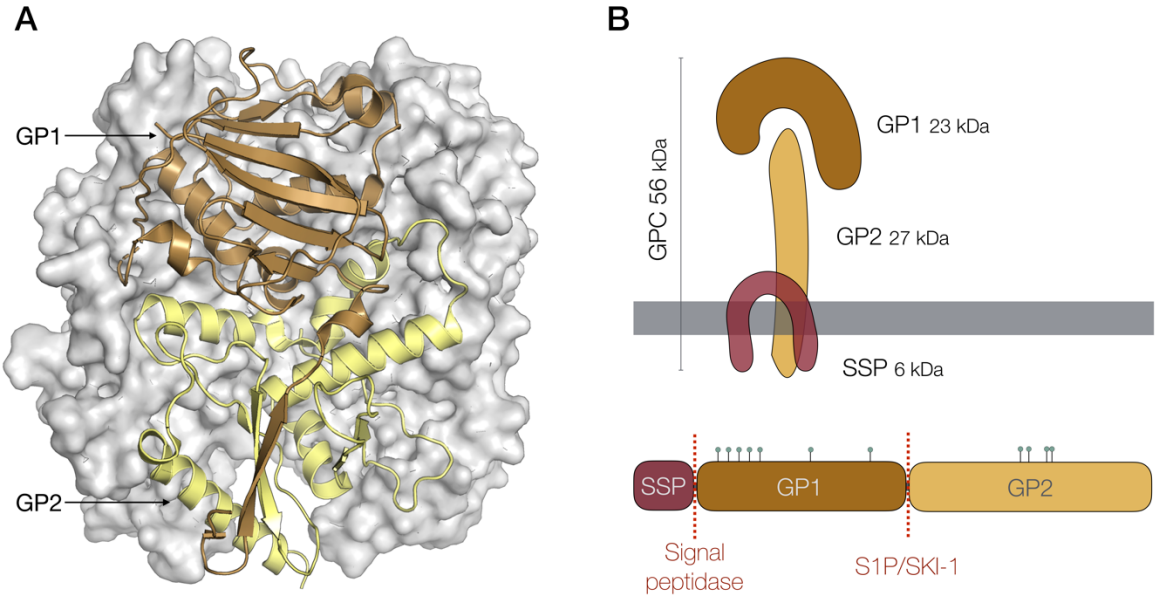


Figure 1.3.5: The LASV GPC. Crystal structure and schematic representation of the LASV GPC. (A) Crystal structure of the trimeric ectodomain of LASV GP1-GP2 complex (PDB 5VK2). The GP1 (brown) and GP2 (yellow) are shown for one protomer. The two additional protomers are shown as grey surfaces. SSP and the transmembrane region of GP2 were absent from the crystallised construct and therefore not displayed. (B) Schematic representation of a LASV GPC protomer (top), coloured as in A, with the addition of SSP (red) and the membrane (dark grey). Protein-only molecular masses for the LASV proteins is indicated as an example. Schematic representation of the LASV polyprotein precursor with proteolytic cleavage sites are shown as dotted lines, and N-linked glycosylation sequons as pins.

The nascent GPC polypeptide is inserted into membranes of the endoplasmic reticulum and proceeds to self-associate to form trimers. The immature trimeric complex is subsequently cleaved in the early golgi by the site-1-protease (also referred to as the subtilisin-like kexin protease-1; S1P/SKI-1) to generate the distinct attachment-mediating, GP1, and fusion, GP2,

proteins. The SSP-GP1-GP2 remain non-covalently associated, both within the hetero-trimeric protomer and between protomers in the trimeric complex (Nunberg and York, 2012). In addition to myristoylation and two distinct proteolytic cleavage events, GPC maturation involves extensive N-linked glycosylation. The highly glycosylated GPC surface is proposed to form a dense glycan shield that facilitates evasion of the antibody-mediated immune response (Watanabe et al., 2018). Furthermore, the high glycan density likely results in the steric restriction of glycan processing which, in LASV GPC, yields under-processed oligomannose-type glycans that may function in promoting cellular entry through the C-type lectin DC-specific ICAM-3-grabbing nonintegrin (DC-SIGN) (Goncalves et al., 2013).

The GP1 sub-component of the arenaviral GPC is responsible for cellular-attachment through the binding of specific host receptors. The OW and NW serogroups utilise different receptors for cellular entry. NW virus entry is primarily associated with binding of transferrin receptor-1 (TfR1) orthologues of rodent reservoir species. Interestingly, whilst the incidental ability to utilise human TfR1 (hTfR1) is not the sole predictor of the ability to enter human cells *in vitro*, it does appear to be a decisive factor in human pathogenesis (Flanagan et al., 2008, Choe et al., 2011). The crystal structure of the GP1 from the NW MACV in complex with hTfR1 confirmed earlier biochemical assays in demonstrating that the attachment glycoproteins of NW viruses bind TfR1 at a membrane-distal 'apical domain' (Abraham et al., 2010) (Discussed in **Chapter 4**). It is noteworthy that the domain of TfR1 utilised by NW arenaviruses is distinct from that known to mediate endogenous transferrin-binding (Lawrence et al., 1999). Thus, in stark contrast to the ephrin receptors, which are bound in an overlapping manner by both henipaviral G proteins and their endogenous binding partners —the ephrin receptors— TfR1 engagement by NW GP1s is distinct from regions of known function (Chrencik et al., 2006). It is therefore plausible that the apical domain of TfR1 would experience limited functional constraint and thus contribute to the postulated co-adaption of NW viruses and their hosts

(Bowen et al., 1997). Given the apparent lack of endogenous function and importance in viral entry, the apical domain of TfR1 is an attractive target for the development of antivirals.

TfR1 functions endogenously in iron homeostasis and as such is ubiquitously expressed. The expression of TfR1 is elevated on activated and rapidly dividing cells, such as cellular immune effectors promoted in response to viral infection (Gatter et al., 1983). To this end, macrophages and lymphocytes, as well as endothelial cells, appear to be susceptible to NW viruses *in vivo*, concordant with the observed disease pathologies (Choe et al., 2011). The up-regulation of immune cells expressing TfR1, in response to viral infection, is proposed to promote the rapid acceleration of viral replication. Furthermore, continual recycling of TfR1, which is crucial to transferrin release, effectively mediates viral internalisation into acidified endocytic compartments that promote pH-dependent membrane fusion (Choe et al., 2011).

OW and NW clade C viral species mediate cellular entry through interaction with the conserved glycoprotein α-dystroglycan (α-DG) (Cao et al., 1998). The GP1 interaction with α-DG occurs via a highly-modified O-linked 'matriglycan' that has, to date, precluded any structural analyses of the interaction owing to the complexity of the matriglycan biosynthetic process (Kunz et al., 2005). Similar to TfR1-utilising arenaviruses, α-DG binding stimulates the uptake of viruses into acidified compartments that stimulate the pH-dependent activity of GP2, however, unlike TfR1-tropic viruses, the process is clathrin-independent (Borrow and Oldstone, 1994). Interestingly, LASV undergoes an obligatory pH-induced receptor switch towards an intracellular receptor, lysosome-associated membrane protein 1 (LAMP1) (Jae et al., 2014). Furthermore, LUJV was also demonstrated to require the tetraspanin CD63 for stimulation of pH-dependent fusion activity (Raaben et al., 2017), highlighting an emerging role for intracellular receptors as viral entry factors.

The crystal structure of the LASV GPC ectodomain, comprising only GP1 and the extra-viral portion of GP2, provided high-resolution affirmation of an earlier cryo-electron microscopy derived reconstruction of GPC from *bona fide* virions (Li et al., 2016, Hastie et al.,

2017). LASV spike complexes comprise a globular ectodomain of approximately 10 nm by 10 nm, in which the GP1 sub-component is positioned atop the GP2. Unusually for viral class I fusion proteins, GP2 has a tripodal arrangement of transmembrane spanning regions, which may suggest a role for SSP—the presence of which is also a unique feature of the arenaviral GPC—in bridging the subunits. Interestingly, despite protein engineering efforts to stabilise the pre-fusion conformation of the LASV GP1-GP2 complex, a protomer spanning antibody fragment was required to 'lock' the GPC in a trimeric arrangement, which suggests that the transmembrane domain, intra-virion domain, or SSP, or combination thereof, are crucial to trimerisation (Hastie et al., 2017).

The pre-fusion arenaviral GP2 adopts a mixed α/β fold that bares little similarity to other class I fusion proteins (Hastie et al., 2017). Despite this, the post-fusion conformation of GP2 forms the canonical six-helix bundle indicative of this class of fusogen (Igonet et al., 2011, Parsy et al., 2013). During endosomal trafficking, the increasingly acidic pH environment likely destabilises the pre-fusion conformation of GP2, resulting in fusogenic rearrangements that ultimately culminate in the merger of viral and host-cell membranes (White et al., 2008). Interestingly, LASV GP1 has been shown to detach from the GPC complex at low pH (<4) (Li et al., 2016), potentially relieving interactions that stabilise the pre-fusion state of GP2 and thus potentiating triggering. To this end, we, and others, demonstrate striking structural transitions within GP1 that likely form as a consequence of acidification and detachment from GP2, suggesting an unusual mutual dependence of both GP1 and GP2 for stabilisation of metastable conformations (Cohen-Dvashi et al., 2015, Li et al., 2016, Pryce et al., 2018). Whilst the precise functional importance of GP1 detachment is undetermined, it is likely that the dissociation of GP1 potentiates fusogenic rearrangements within GP2, and thereafter GP1 may also serve as an immunomodulatory factor.

Chapter 2:

Materials and methods

Chapter 2: Materials & methods

2.1 Molecular biology

2.1.1 Cloning strategy and expression vectors

Construct boundaries were determined using a combination sequence alignment against related proteins with known structures and disorder prediction utilising RONN (Yang et al., 2005). Constructs for protein expression were cloned into the high copy number plasmid vector pHLsec, to facilitate the rapid production of high quantities of recombinant protein (Aricescu et al., 2006) (**Fig 2.1.1**). Unless otherwise stated, gene fragments were inserted between the AgeI and KpnI restriction sites, incorporating the vector-derived secretion signal sequence and C-terminal hexa-histidine tag to aid purification and detection. Secreted proteins retain an N-terminal vector-derived cloning sequence, ETG, and a C-terminal GTK preceding the hexa-histidine tag. Efforts were made to avoid terminating construct reading frames with asparagine residues that would lead to the introduction of a non-native N-linked glycosylation sequon (N-X-S/T, where X is any amino acid except proline), NGT in this instance.

For the production of Fc-tagged paramyxovirus G-proteins, generated for cellular assays performed by collaborators in the laboratory of Prof. Benhur Lee (Icahn School of Medicine at Mount Sinai, USA), relevant G-proteins were clones into the pHL-FcHis vector between the AgeI and KpnI sites (Aricescu et al., 2006). Fc-tagged proteins were affinity purified utilising the hexa-histidine motif, adhering to an identical protocol as non-Fc-tagged protein constructs.

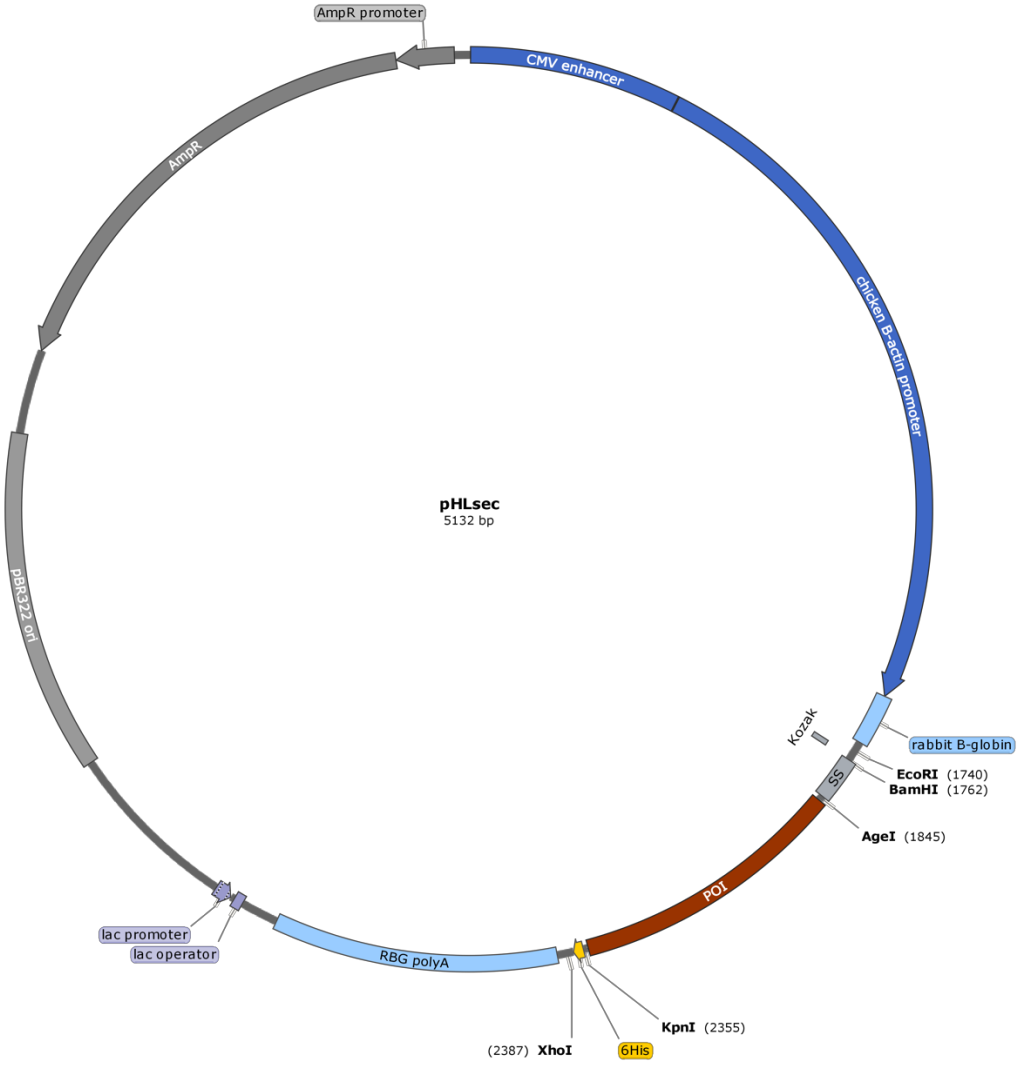


Figure 2.2.1: Salient features of the pHLsec vector. Annotated plasmid map for the pHLsec mammalian expression vector (Addgene plasmid catalogue number 99845). Salient features are displayed. A surrogate 500 bp protein of interest (POI) is displayed in red for illustrative purposes only.

2.1.2 cDNA templates

Synthetic genes encoding nucleotide sequences of interest were codon optimised for recombinant expression in human cell expression systems (codon optimised and synthesised by GeneArt). Lyophilised gene fragments (strings) were resuspended in sterile filtered water to a final concentration of 100 ng/μL for subsequent PCR amplification. Genes supplied within plasmids were resuspended as above and either subjected to PCR or transformed directly into *E. coli* for template amplification.

2.1.3 Polymerase chain reaction

Amplification of the entire gene open reading frame or specific truncation constructs was performed utilising a two-step polymerase chain reaction (PCR) protocol. Phusion® High-Fidelity PCR Mastermix with HF buffer (New England Biolabs) was used for template amplification adhering the manufacturer's recommended protocol for a 50 µl reaction utilising 100 ng of template. Components were mixed by pipetting prior to thermocycling. Reactions were activated for 30 seconds at 98°C; a 10 second 98°C denaturation step and a 45 second 72°C combined annealing and extension step was repeated 30 times; finally, reactions were allowed to elongate for 10 minutes at 72°C before cooling at 4°C. PCR fragments were purified using QIAquick PCR Purification Kits (Qiagen) adhering to the manufacturer's recommended protocol. Purified DNA fragments were analysed by agarose gel electrophoresis.

2.1.4 Primer specifications

Primers were synthesised by Sigma Aldrich or Eurofins and comprised three conserved features: a short tri-nucleotide (CGC) overhang; the restriction endonuclease recognition sequence; and the 5' or 3' region of the desired open reading frame. In instances where fragments were cloned into the XhoI site of pHLsec, tandem stop codons (TAATGA) were added at the 3' end of the open reading frame to ensure translation termination. The length of the open reading frame sequence was adjusted to optimise predicted primer melting temperature (T_m); the temperature at which half of primer-DNA duplexes dissociate. To facilitate a two-step PCR protocol, primer T_m was adjusted to within a range of 72°C to 75°C, using the NEB online T_m calculator (www.tmcalculator.neb.com) and 58-61°C using JaMBW T_m calculator (www.bioinformatics.org/jambw/3/1/9/index.html).

2.1.5 Agarose gel electrophoresis

TAE (Tris, Acetic acid, EDTA) agarose gel electrophoresis was used throughout cloning to monitor DNA purity and identity (size). Typically, samples were run on 1% w/v agarose gels containing 1:10,000 (v/v) SYBR™ safe (Thermo Fisher scientific) for 20–30 minutes at 100 V. Samples were premixed with appropriate amounts of 6x purple gel loading dye (NEB) and separated alongside a 1 kb DNA size standard ladder (NEB). DNA was subsequently visualised using a blue light transilluminator, for gel excision, or imaged with UV light using a ChemiDoc™ system (BioRad).

2.1.6 Enzymatic restriction digests

Purified PCR fragments and vector pairs were subject to restriction digestion by complimentary, commercially derived endonucleases (purchased from NEB). Digestion reaction mixtures were assembled in accordance with the manufacturer's standard protocol and the reaction left to proceed for 20–60 minutes at 37°C. Following digestion, DNA insert fragments were purified using QIAquick PCR Purification Kits (Qiagen), digested vector bands were excised and purified using QIAquick Gel Extraction Kits (Qiagen). The most frequently used endonuclease enzyme pair in this work was AgeI and Kpn1 for 5' and 3' cohesive end generation, respectively.

2.1.7 DNA ligation

Purified, endonuclease digested DNA fragments and vector were ligated using the New England Biolabs Quick Ligation™ kit. Ligation was carried out using the manufacturer's recommend protocol, however, in instances where ligation was unsuccessful, insert:vector molar ratios of 5:1 were utilised, with 50 ng of vector used consistently. Reactions were left to proceed for 15–20 minutes at room temperature prior to immediate use in transformation or freezing at -20°C.

2.1.8 Transformation

Pre-aliquoted DH5α™ subcloning efficiency™ *E. coli* (Invitrogen) were thawed gradually on ice. Newly ligated plasmids were added to the cells at a volumetric ratio of 1:10 (independent of concentration), respectively. In instances where pure, high concentration plasmid DNA was used, such as that arising from previous minipreps, volumetric ratios were reduced to 0.1–0.5:10. Typically, 30 μL of cells were used per transformation. Following DNA addition, cells were incubated on ice for 30 minutes. Cells were heat-shocked at 42°C for 45 seconds and immediately cooled on ice for a minimum of 5 minutes. To establish resistance and aid recovery, though not strictly necessary for bacteriostatic antibiotic selection (such as ampicillin), 200 μl of SOC (Super Optimal Broth, Sigma) medium was added to cells prior to incubation for 1 hour at 37°C in a rotating incubator (160–180 rpm). Cells were plated onto LB-Agar plates containing ampicillin at 100 μg/mL in order to select for cells possessing successfully ligated vector, since pHLsec confers ampicillin resistance. Plates were subsequently incubated at 37°C for 16–18 hours.

2.1.9 Plasmid amplification and purification

Singular colonies obtained from transformations were used to inoculate 5 mL of LB media supplemented with 100 μg/mL ampicillin. Cultures were incubated at 37°C in a rotating incubator (160–180 rpm) for 16–18 hours. Cells were pelleted by centrifugation at 3000–4000 x g for 15 minutes and plasmids were subsequently isolated and purified using QIAprep Spin Miniprep kits (Qiagen), in accordance with the recommended protocol. For large scale plasmid production, three litres of LB-ampicillin (six two-litre conical flasks each containing 500 mL LB) was inoculated and cultured, as above. Plasmids were extracted and purified from the resultant cell pellets using QIAGEN plasmid giga kits.

Purified plasmids were subject to restriction digestion and agarose gel electrophoresis in order to assess the success of ligation. Plasmids containing inserts of the anticipated length

were retained and subject to commercial nucleotide sequencing (Eurofins genomics, formerly GATC Biotech) for additional verification. Sequencing samples were prepared as per the stipulated requirements of the service: 5 μL of plasmid DNA at 80–100 ng/μL premixed with 5 μL of the desired primer at 5 μM. Primers targeting sequences at the 5' and 3' most regions of multiple cloning site in pHLsec were used for sequencing to facilitate obtaining viable coverage of the entire cloned fragment (**Table 2.1.9)**.

Table 2.1.9: Sequencing primers

Primer	Sequence
pHLsec forward	CACCAGCCACCACCTTCTGATAG
pHLsec reverse	GCTGGTTGTTGTGCTGTCTCATC

In all instances, DNA concentrations were estimated by absorbance at 260 nm (A^{260}) utilising a Nanodrop 1000 (Thermo) spectrophotometer.

2.2 Protein expression and purification

2.2.1 Rationale and outline workflow

Recombinant glycoproteins were over-expressed in a human embryonic kidney-derived immortalised cell line, HEK293T (ATTC® CRL-3216™). HEK293T cells are easy to handle, exhibit a fast growth rate that is amenable to rapid scale-up, and are readily transfectable. As a human cell line, HEK293T cells would be expected to yield proteins with near-native mammalian-type glycosylation and maximise the probability of successful folding and appropriate processing. Furthermore, glycosylation can be manipulated with the use of specific inhibitors of glycan processing enzymes, and further processed *in vitro* to reduce sample heterogeneity (Chang et al., 2007). Recombinant proteins secreted into the cell culture medium were buffer exchanged and concentrated before a tandem affinity and size-exclusion chromatography protocol that

generally yielded milligram quantities of high-purity protein suitable for biochemical and biophysical applications.

2.2.2 Mammalian cell culture

All mammalian cell culture procedures were performed within a laminar flow cabinet, using standard aseptic technique. Recombinant protein over-expression was carried out using adherent HEK293T cells grown in Dublecco's Modified Eagle's Medium (DMEM, Sigma) supplemented with 10% fetal calf serum (FCS, Sigma) and 1% each of non-essential amino acids (NEAA, Gibco) and L-glutamine (Gibco) (Aricescu et al., 2006). For maintenance and expansion, cells were cultured in T175 flasks (175cm^2 surface area, Corning) at 37°C, 5% CO_2 and passaged upon reaching 70–90% confluence, as assed by visual inspection using a light microscope. For passaging, cells culture medium was removed, and cells were washed with 10 mL of sterile phosphate buffered saline pH 7.4 (PBD, Gibco). The adherent cell monolayer was incubated with 5 mL of trypsin-EDTA (Life Technologies) for 5 minutes to degrade the extracellular matrix and therein facilitate resuspension. Dislodged cells were subsequently resuspended in 20 mL of the standard culture medium, 2 mL volumes of the resultant suspension were used to inoculate 23 mL of fresh culture medium in new T175 flasks.

2.2.3 Small-scale transient transfection

Small-scale expression tests were carried out in order to assess the suitability of protein constructs for large-scale expression. Six-well plates were inoculated with HEK293T cells such that 90-95% confluence was achieved at the time of transfection. Per construct, 4 µg of purified DNA was diluted in 250 µL of FCS-free DMEM (supplemented with 1% each of NEAA and L-glutamine); similarly, 10 µL of the transfection reagent Lipofectamine® 2000 (Invitrogen) was diluted into 250 µL of FCS-free DMEM in parallel. Following a 5 min incubation, the solutions were combined, mixed by gentle agitation, and allowed to complex for 20 min at room

temperature. The complexed DNA- Lipofectamine® solution was added to 2 mL of fresh DMEM (10% FCS, 1% NEAA, 1% L-Gln) in each well and mixed by gentle manual rocking. Cells were incubated for 48–72 hours prior harvesting.

Following protein expression, cell culture medium was aspirated from the wells and clarified by centrifugation at 4,000 x g for 5 minutes. The remaining adherent cells were dislodged from the wells with vigorous and repeated aspiration in 2 mL PBS. The cell culture supernatant and the cell suspension fractions were retained for western blot analysis in order to assess the extent of secreted protein expression or misfolding.

For all transfections, a mock well comprising all of the above components in the absence of plasmid DNA was utilised as a negative control. Similarly, a construct known to be effectively produced and secreted by the cells was used as a positive control.

2.2.4 Large-scale transient transfection

Cells were transfected upon reaching 90–95% confluence in expanded surface (2,125 cm^2) roller bottles (Cellmaster™, Greiner bio-one). Per roller bottle, 500 µg of high purity plasmid DNA was combined with 1 mg of Polyethylenimine (PEI, Polysciences Inc.), a cationic polymer utilised to mediate DNA transfection (Boussif et al., 1995). Prior to transfection, the requisite amount of plasmid DNA was sterilised by the addition of chloroform (10% by vol.); DNA-chloroform mixtures were mixed vigorously before centrifugation for 10 minutes at 13,000 x g and subsequent DNA recovery by aspiration. Chloroform-sterilised DNA and PEI were allowed to complex at room temperature for a minimum of 10 min in FCS-free DMEM (50 mL per flask) before addition to 200 mL (per flask) of fresh DMEM (2% FCS), pre-measured in roller bottles. In most instances, kifunensine, a class I mannosidase inhibitor, was added to the DNA-PEI mix at 5 µM following complexation (Elbein et al., 1990). The addition of kifunensine restricts glycan processing to yield high-mannose type glycoforms, predominantly $Man_9GlcNAc_2$ (Bowden et

al., 2008b), that are susceptible to enzymatic cleavage by endoglycosidase F_1. For protein expression, cells were incubated at 37°C, 5% CO_2 with constant rolling rotation for 72–96 hours.

2.2.5 Western Blot analysis of small-scale expression tests

Western blotting was utilised to comparatively quantify secreted protein concentration and therefore assess the suitability of a given construct for large-scale expression. Protein samples were separated according to molecular weight using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) using commercial 4–12% Bis-Tris PAGE gels (Life Technologies). Samples were subject to a consistent current at 160–180 V for 40–60 minutes, or until the dye front reached the end of the gel. Proteins were subsequently blotted onto a nitrocellulose membrane, a component of pre-assembled commercial iBlot™ transfer stacks, using an iBlot™ 2 gel transfer system (Thermo Fisher). Blotted membranes were blocked with 10% milk powder in PBS (w/v) for a minimum of one hour at room temperature. To mediate the detection of his-tagged proteins, membranes were incubated (1h at 25°C) with a mouse-derived monoclonal anti penta-his antibody (ID: 34660, Qiagen) diluted 1:1000 (v/v) in PBS containing 5% milk. Signal amplification and detection was achieved with the use of a peroxidase-conjugated goat-derived anti-mouse IgG antibody (Sigma), diluted 1:2000 (v/v) in PBS containing 5% milk, incubated for 1h at 25°C. Following each antibody incubation step, membranes were washed thoroughly with PBS supplemented with 0.05% (v/v) tween-20, typically involving three five-minute washes, each with an intervening rinse step. Signal development and detection was achieved using the Clarity™ Western ECL substrate (Bio-Rad) and a ChemiDoc™ system (BioRad), respectively.

In all instances, proteins were separated alongside two molecular weight ladders: (1) the BenchMark™ pre-stained protein ladder (Thermo Fisher) was utilised to assess the efficacy of protein transfer; and (2) the his-tagged BenchMark™ protein ladder, was used for size determination and confirmation of antibody staining success.

In instances where protein identification was required for purposes other than analysis of small-scale expression tests, an identical western blotting protocol was utilised. Alternative antibody combinations, if used, are detailed where relevant.

2.2.6 Cell harvesting and diafiltration

Following protein expression, cell supernatants were collected and clarified by centrifugation at 4,000 x g for 20 minutes prior to filtration through 0.22 µm vacuum filter units (Millipore). Filtered supernatants were buffer exchanged into 10 mM Tris pH 8.0, 150 mM NaCl, and concentrated using an ÄKTA flux cross flow filtration system (GE Healthcare) fitted with either 10 kDa or 50 kDa molecular weight cut-off (MWCO) membranes.

2.2.7 Nickel-affinity chromatography

Hexa-histidine tagged protein constructs were purified by metal-affinity chromatography. Five millilitre HisTrap™ HP (GE Healthcare) nickel affinity columns were utilised in conjunction with an Äkta Purifier (GE Healthcare) liquid protein chromatography system at room temperature, controlled by the UNICORN software package (GE). The system was equilibrated in 150 mM NaCl, 10 mM Tris pH 8.0 and an imidazole gradient established by machine-controlled mixing of 1.0 M imidazole in the same buffer. The hexa-histidine motif specifically binds immobilised nickel ions chelated by the stationary phase; histidine-tagged proteins were competitively displaced by the addition of imidazole, a histidine side chain mimetic (Hochuli et al., 1988). Protein samples were supplemented with 10 mM imidazole prior to loading, in order to increase the stringency of binding and thus reduce non-specific interactions with media components. Non-specifically bound proteins were eluted in a gradient of 0-50 mM imidazole until a stable baseline was achieved. Proteins of interest were typically liberated by addition of 250 mM imidazole.

2.2.8 Protein concentration determination

Protein concentrations were determined routinely throughout purification and subsequent handling steps. In all instances, protein concentration was estimated by virtue of the absorbance of light at 280 nm, primarily by aromatic side chains, utilising a Nanodrop 1000 (Thermo) spectrophotometer. The predicted extinction coefficients for individual protein constructs were calculated using the 'ProtParam' tool online (available at: https://web.expasy.org/protparam/), accounting for the entire secreted sequence inclusive of vector-derived residues. Protein samples were assumed to obey the Beer-Lambert Law (Absorbance (A)= molar extinction coefficient (e) x molar concentration (c) x path-length (l)) for the purposes of concentration determination.

2.2.9 Enzymatic deglycosylation

Partial enzymatic deglycosylation with endo F_1 (Grueninger-Leitch et al., 1996), was performed at 1:100 (endo F_1: Protein) mass ratio, and the reaction left to proceed at room temperature overnight. EndoF_1 acts upon the high-mannose type glycoforms (predominantly $Man_9GlcNAc_2$) present as a consequence of kifunensine treatment, cleaving the glycosidic bond of the two N-acetlyl glucosamine moieties that constitute the di-N-acetylchitobiose core, thus generating proteins with truncated, single GlcNAc glycans (Maley et al., 1989). The efficacy of deglycosylation of new protein constructs was assesses by SDS-PAGE and size-exclusion chromatography. Enzymatic deglycosylation with endo F_1 was performed for all constructs that were subsequently subject to crystallisation trials, in order to reduce the conformational heterogeneity of the samples.

In some instances, complete enzymatic deglycosylation was performed utilising the enzyme peptide-N4-N-acetyl-beta-D-glucosaminyl-asparagine amidase (PNGase F), acquired commercially (New England Biolabs) or kindly provided by Yasunori Watanabe. Unlike endo F_1, PNGase F removes entire N-linked glycans through hydrolysis of the innermost asparagine-

GlcNAc amide bond, and in doing so, converts the asparagine to an aspartic acid residue (Maley et al., 1989). For PNGase F-mediated deglycosylation, reactions were performed under non-denaturing conditions according to the manufacturer's protocol.

2.2.10 Size-exclusion chromatography

Size-exclusion chromatography was utilised as a final polishing step prior to crystallisation or further down-stream experiments that required high purity, monodisperse protein. Superdex™ 200 Increase 10/300 GL (GE Healthcare) columns were utilised in conjunction with an Äkta Purifier (GE Healthcare) and, unless otherwise stated, were equilibrated in 150 mM NaCl, 10 mM Tris pH 8.0. Samples were concentrated using Millipore™ Amicon™ centrifugal filter units with an appropriate MWCO and filtered using 0.22 µm centrifugal filter units (Millipore™) immediately prior to injection. Samples were typically loaded in less than 400 µL volumes using a 500 µL loop, to account for laminar flow effects. Wherever possible, SEC was performed using the same machine to facilitate comparison between protein samples by negating variables such as tubing lengths. Fractions were assessed by SDS-PAGE and subsequently pooled and concentrated for crystallisation or other applications.

2.3 Protein crystallisation

The formation of protein crystals for structure elucidation by X-ray diffraction requires milligram quantities of highly pure protein that is both compositionally and structurally homogeneous (Chayen and Saridakis, 2008). To this end, proteins were freshly purified, or re-purified, by SEC immediately prior to crystallisation experiments.

Proteins were concentrated and filtered using Millipore™ Amicon™ centrifugal filter units with appropriate MWCOs and subject to a commercial pre-crystallisation test (PCT™, Hampton) in order to estimate a protein concentration amenable to crystal growth. Crystallisation experiments were performed using the sitting drop vapour diffusion method.

Drops comprising 100 nL of protein and 100 nL of precipitant solution were dispensed into 96-well Greiner CrystalQuick® plates with the aid of a crystallisation robot (Cartesian Technologies). Plates were sealed manually and left to equilibrate against a reservoir of 95 µL of precipitant, housed within a Formulatrix® storage vault at either 21°C or 4°C, and subject to automated imaging (Walter et al., 2005).

Proteins were initially subject to a range of in-house screens (Division of Structural Biology, University of Oxford), derived from a number of commercially available precipitant solutions (Hampton Research and Molecular Dimensions). Initial crystal hits were optimised by precipitant dilution screening and supplementation with additives (Hampton Research). Conditions yielding crystals are detailed in the relevant chapters.

2.4 Structure determination

2.4.1 Crystal harvesting and cryo-protection

Crystals were harvested using LithoLoops™ and cryoprotected with glycerol or ethylene glycol. Cryoprotectants were used to supplement (25% v/v) the mother liquor from which the crystals were obtained; the resulting solution was either added directly to the crystallisation drop or crystals were transferred into the cryoprotectant. Cryo-protected crystals were immediately cryo-cooled in liquid nitrogen and maintained at liquid nitrogen temperatures thereafter. Many crystals were harvested with generous help and guidance of Dr. Karl Harlos.

2.4.2 X-ray data collection and reduction

X-ray diffraction data were primarily collected remotely at the I02 and I04 beam-lines at Diamond Light Source (UK). Data were collected at 100 K and diffracted X-rays recorded on a Pilatus 6M detector (Dectris). Diffraction data from a number of Whitewater Arroyo Virus (WWAV) GP1 crystals were collected *in vacuo* at beamline I23 (Diamond Light Source, UK) with the generous help of Dr. Kamel El Omari. Performing X-ray diffraction data in a vacuum negates

absorption and scattering effects incurred by the beam and thus permits the use of long wavelength X-rays that are suitable for experimental phasing techniques (detailed in **Chapter 4**) (Wagner et al., 2016).

Diffraction images were handled in accordance with Bragg's law, wherein a set of reflections is recorded, each with measurable position (diffraction angle) and intensity, a consequence of X-ray diffraction by planes of atoms within the crystal lattice that obey the Bragg condition (Bragg and Bragg, 1913), as defined by Bragg's Law:

$$n\lambda = 2d\sin\theta$$

Where n is an integer, λ is the wavelength of the incident X-rays, d is the separation of parallel lattice planes and θ is the scattering angle. X-ray diffraction data were indexed, integrated and scaled using xia2 (Winter, 2010). Utilising knowledge of the detector geometry and reflection position, diffraction data can be 'indexed' thus assigning miller indices to the reflections and therein defining space-group parameters. Observed reflections were 'integrated' in a process that determines their boundaries and intensity, and that of the background. Due to variation in a number of parameters during data collection, recorded intensities cannot be considered absolute, thus equivalent indices were 'scaled' to unify the intensity data on a comparative scale.

High resolution cut-offs were automatically determined by xia2 using its standard criteria ($CC_{1/2}$>0.5, merged I/σI >1, unmerged I/σI >0.25) and manually assessed via monitoring the resolution dependence of statistics such as the signal to noise ratio (I/σI), correlation coefficient ($CC_{1/2}$) (Karplus and Diederichs, 2012), completeness and R_{merge}/R_{pim}.

2.4.3 Phase estimation

In order to derive an electron density map that accurately describes the molecules present within the crystal, structure factor amplitudes and phase angle must be determined for each

observed reflection (Rhodes, 2006). Whilst amplitude information is obtained directly from the measured intensities of the reflections, phase information can-not be recorded experimentally. It is therefore necessary to estimate the phase information, a process that can be achieved both experimentally or, more commonly, computationally (McCoy et al., 2007).

In all cases, initial phase estimation was attempted by performing molecular replacement (MR) using the program Phaser (McCoy et al., 2007), which utilises phase information from similar proteins of known structure, in order to provide an initial estimate of the phases of the protein of interest. Models used for MR were obtained from the Protein Data Bank (PDB) and manually manipulated (in a text-editor) to remove atoms, such as those in variable loops, in order to decrease the number of atomic clashes that may preclude a successful solution. Extraneous water and ligand molecules were also removed. Resultant MR solutions were assessed on three criteria: overall Z-score, log-Likelihood Gain (LLG) and the quality of the resultant electron density maps.

As phase estimation by molecular replacement failed to yield a solution for WWAV GP1, the single-wavelength anomalous dispersion (SAD) method was utilised to experimentally determine initial phases (**Chapter 4**). SAD-based phasing depends upon the specific and characteristic absorption of X-rays of a given wavelength, typically by heavy metals. Absorption of X-rays by such atoms gives rise to a non-compliance with Friedel's Law, which states that the intensities of reflections that are separated by 180° in the reciprocal lattice are equivalent, from which phases can be approximated (Rhodes, 2006). Specific and expanded details of phase determination are supplied in the relevant thesis chapters.

2.4.4 Model building and refinement

As phases are not measured, but estimated, the improvement of electron density maps resulting from phase estimation was necessary. Multiple rounds of iterative model improvement were performed. Such iterative improvement is necessary since the working model provides phase

information for the subsequent map calculation, thus appropriate manual real-space improvement of the model, in response to observed map features, yields more accurate phase estimates and concomitant map improvement.

Reciprocal-space refinement was performed using PHENIX (Adams et al., 2010) and manual real-space model improvements were carried out using COOT (Emsley and Cowtan, 2004). The suitability of chosen refinement strategies and model quality were assessed by monitoring the R-factors: R_{free} and R_{work}. The R-factor describes the agreement between the experimentally observed structure factor amplitudes (F_{obs}) and those calculated from the current model (F_{calc}):

$$R = \frac{\Sigma \, ||F_{obs}|-|F_{calc}||}{\Sigma \, |F_{obs}|}$$

R_{work} describes the R factor for reflections undergoing refinement, with amplitudes calculated from 'worked' models. In order to prevent overfitting, an R_{free} was calculated against an independent test set of randomly selected reflections (5%), removed prior to the commencement of refinement (Brünger, 1992). The numerical difference between R_{work} and R_{free} was generally maintained within 5%; divergence of the values is indicative an inappropriate refinement strategy or model over-fitting and thus was avoided. Model quality was additionally validated by monitoring metrics of protein geometry, principally the root-mean-square-deviation (r.m.s.d) of bond angles and lengths relative to known standards, and analysis of peptide torsion angles (Ramachandran et al., 1963). The stereochemical properties of all structures were validated using MolProbity (Chen et al., 2010).

2.5 Structural analysis

2.5.1 Molecular visualisation and presentation

Molecular graphics were generated using The PyMOL Molecular Graphics System (Schrödinger, available online: www.schrodinger.com/pymol). The atomic coordinates used, if not data generated within this project, were obtained from the PBD and acknowledged accordingly. Electrostatic potential calculations were performed using the Adaptive Poisson-Boltzmann Solver, APBS (Baker et al., 2001).

2.5.2 Solvent accessibility determination

Determination of the solvent accessibility of protein residues, and the representation of sequence alignments was performed using ESPript (Gouet et al., 2003) utilising multiple sequence alignments generated using MultAlin (Corpet, 1988). ESPript generated accessibility scores were validated by manual inspection of the solvent accessible surface calculated by PyMOL.

2.6 Structural phylogenetic analysis

Pairwise evolutionary distance matrices were created using the Structural Homology Program, SHP (Stuart et al., 1979), and subsequent un-rooted phylogenetic trees were generated using the program PHYLIP (Felsenstein, 1989). In all instances, structures involved in the analysis had all extraneous ligands, proteins chains, and water molecules removed. Where non-crystallographic symmetry related copies of the proteins were present within the asymmetric unit, the most complete chain was utilised for analysis.

2.7 X-ray footprinting mass-spectrometry

2.7.1 Sample irradiation

Protein samples were buffer exchanged, using SEC, into 10 mM Phosphate 100 mM NaCl, which is established as an appropriate buffer for XF-MS analyses. In order to minimise the presence of hydroxyl radical scavenger species, particularly glycerol, all Amicon™ centrifugal filter units were washed thoroughly with the sample buffer prior to protein concentration or filtration. Unliganded henipaviral attachment glycoproteins were concentrated to 5 μM and their receptor-bound complexes to 3 μM (assuming 1:1 stoichiometry). Samples irradiation was performed at beamline 3.2.1 of the Advanced Light Source (ALS, Berkeley, California, USA), with the guidance of Dr. Sayan Gupta. Samples were exposed to X-rays for 0, 10, 25, 50, 75 and 100 ms using an automated microfluidics system (Gupta et al., 2016a). Following exposure, samples were immediately quenched with 10 mM (final concentration) L-methioninamide hydrochloride (Sigma) and stored at –80°C.

2.7.2 Mass spectrometry

Mass spectrometry analysis was performed by the 'mass spectrometry and proteomics' facility at the University of St Andrews. In order to increase the coverage of observed peptides, the standard operating procedure was adapted to include an orthogonal peptide digest with endoproteinase GluC, in addition to trypsin digestion. Furthermore, PNGase F was used to deglycosylate samples in response to an absence of detected glycopeptides in initial experiments. Data were analysed using the Mascot Server (Matrix Science).

2.8 Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy was utilised to infer information pertaining to the secondary structure content of arenaviral GP1 glycoproteins. CD experiments were performed at beamline B23 at Diamond Light Source (UK), accessed via the proposal-based mail-in service.

Experiments were performed using a nitrogen-flushed module A/B end-station/ChirascanPlus spectrophotometer (Jávorfi et al., 2010, Hussain et al., 2012). Data transfer and the obtained results were processed using CDApps and secondary structure estimation from CD spectra was carried out using the CDApps Continll algorithm (Provencher and Gloeckner, 1981).

Whitewater Arroyo Virus (WWAV) and Loei River Virus (LORV) GP1 glycoproteins were buffer exchanged, using SEC, into buffers comprising 150 mM NaCl and either 10 mM Tris pH 8.0- or 20-mM citrate pH 5.0. WWAV GP1 and LORV GP1 protein solutions were concentrated to 6.7 mg/mL and 7.6 mg/mL, respectively. Buffer that had passed through the SEC column was retained and used to generate baseline spectra.

Theoretical CD spectra were calculated using the PDB2CD online server (Mavridis and Janes, 2017) in order to provide a comparative model that accounts for the modulations of the CD spectra by factors other than absolute secondary structure content, such as the relative orientations of such secondary structural elements.

2.9 Enzyme-linked immunosorbent assays

Enzyme-linked immunosorbent assays (ELISA) were performed to facilitate the detection of protein-substrate binding events, utilising antibody-mediated detection and signal amplification, analogous to western blot analyses

ELISAs were performed in 96-well Nunc-immuno™ PolySorp® plates (sigma) to which 350–1400 ng (70 µL of 5–20 µg/mL) of Lassa virus (LASV) virus-like particles (VLPs, kindly provided by Dr. Sai Li) were added and bound to the plate overnight at 4°C. Plates were blocked with PBS supplemented with 5% (w/v) milk powder (100 µL per well) for one hour at room temperature. Following blocking, plates were washed thoroughly with PBS supplemented with 0.05% tween-20, hereafter termed PBST). For assays in which pH was an experimental variable, wells were equilibrated with a 15 min incubation in 100 µL of an appropriate pH buffer (10mM SPG pH 5.0, 150mM NaCl or PBS pH 7.4). Milk powder was not used beyond this point, owing

to the potential for aggregation induced by low pH. Serial dilutions of purified Lysosomal-associated membrane protein 1 (LAMP1) were added to the appropriate wells and left to equilibrate for 2 hours at room temperature with gentle rocking. Wells were subsequently washed five times with the appropriate pH buffer supplemented with 0.05% tween-20, to remove unbound LAMP1. Detection of specifically bound LAMP1 was mediated by antibody binding to a C-terminal hexa-histidine tag present on LAMP1 but not VLPs. An anti-hexa-histidine rabbit-derived antibody (ab9108, abcam) was diluted 1/1000 in the appropriate pH buffer supplemented with 0.05% tween-20, added to the wells in 50 µL volumes, and allowed to incubate for 1h at room temperature. The secondary antibody, an HRP-conjugated goat-derived anti-rabbit antibody (ab97080, abcam), was used under identical conditions to the primary. Following each antibody binding step, plates were washed five times with the appropriate assay buffer supplemented with 0.05% Tween-20 to remove non-specifically bound antibody. Detection was mediated using the Pierce™ TMB substrate kit (Thermo Fisher) and binding was quantified by virtue of optical absorbance at 450 nm at between 30 minutes and one hour after substrate addition.

2.10 Sample preparation for glycan composition analysis

The composition of N-linked glycans within LAMP1 was determined using mass spectrometry (Behrens et al., 2016). Global and site-specific glycan analyses were kindly performed by Mr. Yasunori Watanabe (Crispin Laboratory, University of Southampton, UK) as detailed previously (Behrens et al., 2016, Watanabe et al., 2018). For such analyses, recombinant secreted LAMP1 ectodomain was produced in HEK293T cells in the absence of kifunensine and purified as detailed above.

2.11 Numerical data representation and statistical analyses

All data plots were generated, and subsequent statistical analyses performed using GraphPad Prism version 7.0c for Mac (GraphPad Software, La Jolla California, USA). Specific details of statistical methods are detailed in the relevant thesis chapters.

Chapter 3:

Crystal structure of the Cedar virus attachment glycoprotein

Chapter 3: Crystal structure of the Cedar virus attachment glycoprotein

3.1 Preface

This chapter presents the crystal structure of the attachment glycoprotein, termed 'G', from the Cedar henipavirus. Aspects of this project were carried out in collaboration with the laboratory of Professor Benhur Lee at the Icahn School of Medicine at Mount Sinai (USA). Cellular and *in vitro* binding assays that identified Cedar virus entry receptors were perfomed by Mr. Kristopher Azarm, in some instances utilising reagents described within this chapter. The crystal structure of CedPV G and its characterised receptor-binding characteristics are novel findings that are as yet unpublished.

3.2 Summary of findings

The *Henipavirus* genus, of the *Paramyxoviridae* family, contains two lethal human pathogens that cause fatal encephalitis and respiratory failure in humans. The attachment glycoprotein, G, of henipaviruses is solely responsible for host-cell attachment and is therefore a principal determinant of host-species and cell-type tropism (Negrete et al., 2005). The ability of the prototypic henipaviruses, Hendra (HeV) and Nipah (NiV), to utilise numerous mammalian ephrin-B2 and -B3 orthologues underscores their broad host range, which is unusually large amongst the *Paramyxoviridae*, and can account for spillover into human populations (Eaton et al., 2006). Intriguingly, Cedar virus (CedPV), a close genetic relative of HeV and NiV, displays a distinct but overlapping receptor repertoire with HeV and NiV. Analysis of the crystal structure of the attachment glycoprotein of CedPV reveals that it likely possesses a conserved mode of ephrin-B2 recognition, and thus has zoonotic potential. Furthermore, we rationalise CedPVs ability and inability to utilise ephrin-B1 and -B3, respectively, and propose that ephrin-B3 utilisation, and the tissue tropism dictated thereby, may contribute to the extreme pathobiology associated with HeV and NiV infection.

3.3 Introduction

The *Henipavirus* genus constitutes a group of emerging zoonotic pathogens responsible for frequent outbreaks of severe disease in livestock and humans (Selvey et al., 1995, Chua et al., 2000). The prototypic henipaviruses Hendra (HeV) and Nipah (NiV) are unique amongst paramyxoviruses in their classification as biological safety level 4 (BSL-4) select agents (Eaton et al., 2006). Classification of HeV and NiV as BSL-4 pathogens is, in part, due to their unusually broad host range, ability to cause fatal neurological and respiratory disease in humans, and the absence of any specific therapeutics or vaccine (Hooper et al., 2001). To this end, the henipavirus attachment glycoprotein, G, is responsible for cellular receptor recognition and is therefore a principal determinant of species and cell-type tropism (Negrete et al., 2005). The henipaviral attachment protein mediates receptor recognition and subsequent initiation of the pH-independent activation of a cognate fusion (F) protein, culminating in the merger of viral and cellular membranes and viral genome release (Chang and Dutch, 2012).

Like other paramyxoviral attachment glycoproteins, the henipaviral G protein is a homo-tetrameric complex, in which each protomer comprises a short N-terminal intra-virion domain; a single pass transmembrane domain; a tetramerisation-mediating stalk; and receptor-binding domain. The receptor-binding domains of all known paramyxoviruses utilise a conserved β-propeller architecture to mediate the recognition of sialic acid and diverse proteinaceous receptors (Bowden et al., 2010b). The henipavirus G protein recognises the B-class ephrin ligands (Negrete et al., 2005), a group highly conserved cell-surface proteins that function in concert with the Eph family of receptor tyrosine kinases to regulate cellular adhesion and guidance (Poliakov et al., 2004). Orthologues of the two established henipaviral receptors, ephrin-B2 and ephrin-B3, are extremely well conserved across vertebrates, and exhibit greater than 95% amino acid sequence identity across numerous henipavirus host and vector species, such as bats, horses, pigs and humans (Himanen et al., 2007). Furthermore, the expression of both ephrin-B2 and -B3 in neuronal and endothelial tissues, amongst others, is closely aligned

with the severe respiratory and neurological disease pathologies observed in instances henipaviral infection (Pernet et al., 2012). Thus, the relative ubiquity in tissue distribution and strict conservation of the ephrin ligands underscores henipaviral cellular tropism characteristics, accounting for observed disease pathologies and the extremely large permissive host range.

Serological studies suggest that henipaviruses occupy a broad geographic range, coincident with, but not restricted to, the home range of bat reservoir species belonging to the genus *Pteropus* (Halpin et al., 2011). Such wide-spread geographic prevalence, combined with the large host range and zoonotic potential of henipaviruses, highlights the prospect of the emergence and spillover of novel pathogens of unknown risk into human populations (Pernet et al., 2014). Indeed, Cedar virus (CedPV) was recently isolated from the urine of flying fox bats in Queensland (Australia), and thereafter classified as a novel henipaviral species based upon genomic and serological similarities (Marsh et al., 2012). Subsequently, CedPV was shown to utilise ephrin-B2 for cellular entry and, despite commonality with the highly pathogenic NiV and HeV, was demonstrated to be non-pathogenic in experimental infection models (Marsh et al., 2012). Intriguingly, whilst all known henipaviruses (NiV, HeV, GhV, and CedPV) utilise ephrin-B2 for cellular entry, only the highly pathogenic HeV and NiV utilise ephrin-B3 (Marsh et al., 2012, Lee et al., 2015). Thus, functional ephrin-B3-mediated cellular entry may contribute to the severe pathogenicity of HeV and NiV, that is not observed in CedPV.

Here, we sought to delineate the molecular determinants of ephrin ligand recognition and selectivity by the CedPV attachment glycoprotein. In defining the features of attachment glycoproteins that dictate receptor discrimination and relating the pathogenic outcomes of viruses that possess different ephrin tropism characteristics, these data provide a blueprint for assessing both the zoonotic and pathogenic potential of novel henipaviruses.

3.4 Expression, purification, and crystallisation of CedPV G

3.4.1 Construct design and expression testing

Constructs of CedPV G were derived from a synthetic codon-optimised gene encoding the entire G open reading frame (NCBI reference sequence: YP_009094086.1). Construct boundaries were determined based upon alignment with HeV G and NiV G, for which crystal structures were available (Bowden et al., 2008a). Primers corresponding to four N-terminal 'start' sites (E100, E180, N189, and K209) and a single C-terminal 'end' (C622) were synthesised and PCR performed for the four unique 5' and 3' combinations (**Fig 3.4.1**). Resultant DNA fragments were cloned into the pHLsec mammalian expression vector (Aricescu et al., 2006) (**Chapter 2 section 2.1**).

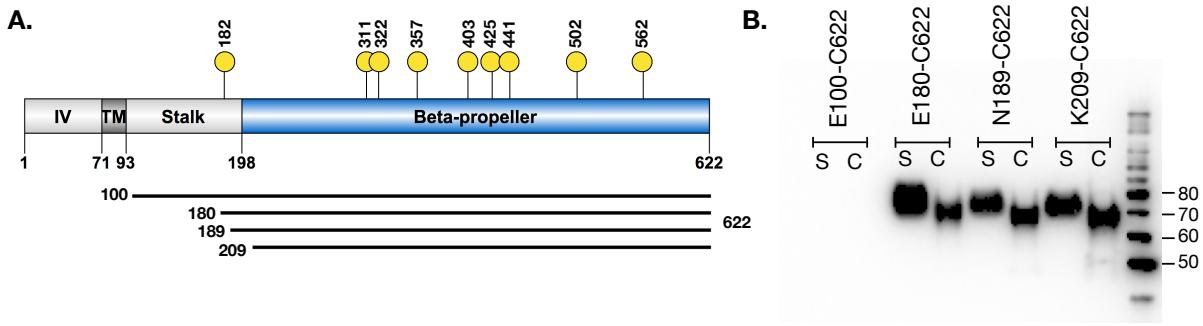


Figure 3.4.1: Construct design and small-scale expression testing of CedPV G. (A) Domain organisation and salient features of the CedPV G protein. The attachment-mediating β-propeller domain (blue), transmembrane (TM), and intra-virion (IV) regions are labelled. Putative N-linked glycosylation sites (NXT/S where X≠P) are displayed as pins. Cloned G constructs are illustrated as lines below the schematic. (B) CedPV G small-scale expression testing. Constructs illustrated in A were transiently transfected in HEK923T cells and subsequent protein expression analysed by his-tag mediated western blot detection, under reducing conditions. Cell lysates (C) and supernatants (S) for each construct were run in parallel.

Small-scale expression testing in HEK293T cells revealed that all constructs, other than E100–C622, were secreted from the cells, indicative of correct folding and post-translational modification (**Fig 3.4.1 A**). Secreted protein was present in approximately equal quantities for the N189–C622 and K209–C622 constructs. As such, the shortest, K209–C622, was selected for large-scale protein production, in an attempt to minimise flexible termini that may retard

crystallogenesis. Additionally, E180–C622, which was present in greater quantities than all other constructs, was also subject to large-scale expression.

3.4.2 Large-scale transient expression and purification

The constructs comprising E180–C622 and K209–C622 of CedPV G were used to transiently transfect HEK293T cells, in the presence of the α-mannosidase inhibitor kifunensine (Chang et al., 2007). Cell culture supernatants were clarified by centrifugation, and subsequently buffer exchanged into 10mM Tris pH 8.0, 150 mM NaCl (TBS). A combined nickel-affinity and size exclusion chromatography (SEC) schedule yielded pure CedPV G (**Fig 3.4.2 C**).

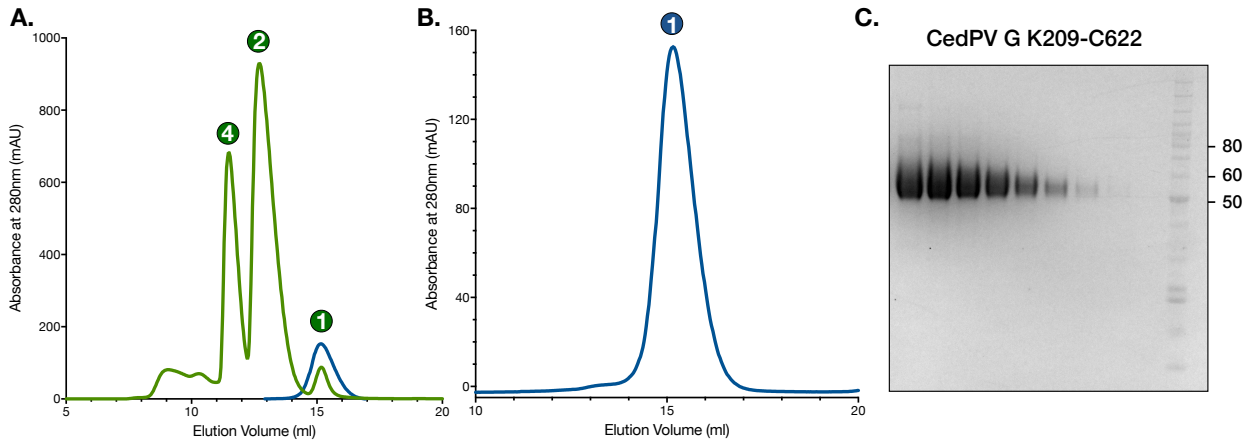


Figure 3.4.2: Purification of CedPV G. (A) Size exclusion chromatogram of kifunensine treated and endoglycosidase F_1 deglycosylated Cedar G E180–C622 (green). The protein forms putative tetramers (4), dimers (2), and monomers (1) with estimated molecular masses of 200, 100, and 30 kDa, respectively. The momomeric CedPV G K209–C622 construct is overlaid for comparison. (B) Deglycosylated CedPV G K209–C622 elutes in a symmetrical peak that would co-elute with the putative monomer peak observed in A. (C) Reducing SDS-PAGE analysis of CedPV G K209–C622, which has a proteinaceous mass of 48.8 kDa suggests that CedPV G is largely homogenously deglycosylated.

When assessed by SEC, CedPV G E180–C622 exhibited three species of interest (**Fig 3.4.2 A**), which eluted with estimated molecular masses of 200, 100 and 30 kDa, as determined by an empirically derived SEC standard curve, and were putatively attributed to tetrameric, dimeric and monomeric G protein, respectively. It is likely that residues within the

multimerisation-mediating stalk region of the protein, included within CedPV G E180–C622 construct, were responsible for the observed behavior. Due to the solution-state heterogeneity displayed by the sample, CedPV G E180–C622 was not deemed suitable for crystallisation experiments.

Expression of CedPV G K209–C622 and subsequent enzymatic deglycosylation with endoglycosidase F_1 (endoF_1), which acts upon the immature high-mannose glycoforms induced by kifunensine (Maley et al., 1989), yielded monodisperse, monomeric protein that appeared to comprise primarily completely deglycosylated glycoforms (**Fig 3.4.2 B & C**). CedPV G K209–C622 was thus deemed suitable for structural analyses, and is herein synonymous with CedPV G.

3.4.3 Crystallisation

Purified CedPV G was concentrated to 6.7 mg/ml and subjected to the full range of in-house crystallisation screens available at the Division of Structural Biology (Oxford, UK). Crystal trays were equilibrated and stored at 21°C. A single crystal of CedPV G was observed after 15 days in a condition comprising 5% (w/v) polyethylene glycol 6000 and 0.1 M MES pH 6.0.

3.5 Expression, purification, and crystallisation of CedPV G-receptor complexes

3.5.1 Construct design, large-scale transient expression, and purification

A construct comprising residues 25–168 of human ephrin-B2 (GenBank NM_004093), corresponding to the NiV G and HeV G binding-competent region of the protein, was kindly provided by Prof. Thomas Bowden (**Fig 3.5.1 i**). The construct, hereafter synonymous with ephrin-B2, is known to express in HEK293T cells and thus was not subject to expression testing or construct optimisation (Bowden et al., 2008a). An analogous construct to ephrin-B2, comprising residues 29–167 was designed for human ephrin-B1 (NCBI reference sequence NP_004420.1), based upon sequence alignment (**Fig 3.5.1 i**). The ephrin-B1 construct

comprising 29–167 is hereafter referred to as ephrin-B1. Small-scale expression by transient transfection in HEK293T cells revealed ephrin-B1 to be effectively secreted from cells, both when produced alone and when co-transfected in a one to one mass ratio with CedPV G.

Based upon previous observations (Bowden et al., 2008a) and the need for fewer purification steps, ephrin constructs were co-transfected alongside CedPV G. Ephrin-B1 and -B2 plasmid DNA was transfected in 1.5- and 3.0-fold excess (by mass) over CedPV G, respectively, in an attempt to saturate CedPV G binding sites. Both putative complexes expressed in milligram per litre quantities in HEK293T cells, grown in the presence of kifunensine.

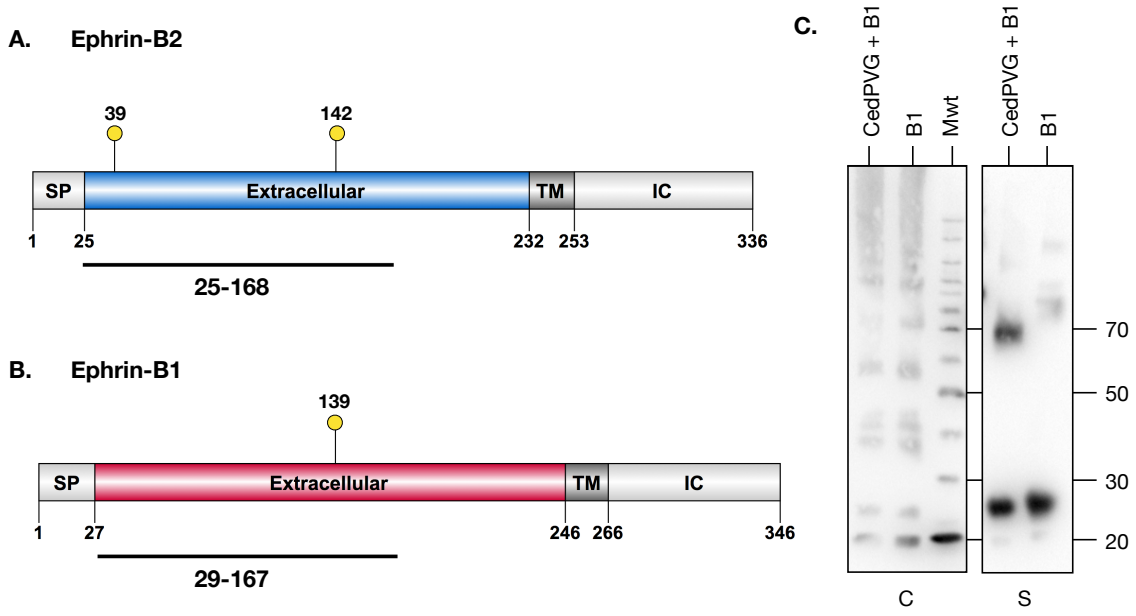


Figure 3.5.1i: Construct design and small-scale expression testing of ephrin-B1 and -B2. (A) Schematic representation of the human ephrin-B2 gene (EFNB2). The signal peptide (SP), extracellular (blue), transmembrane (TM), and intracellular (IC) domains are annotated. N-linked glycosylation sequons are shown as pins. The expression construct for the protein is illustrated with a bar beneath the schematic. (B) Schematic representation of the human ephrin-B1 gene (EFNB1), annotated as in panel A. The extracellular domain is coloured red. (C) Western blot analysis of small-scale expression test of ephrin-B1 alone (B1) or co-transfected with CedPV G (CedPV G + B1). Cellular and supernatant fractions were run in parallel, labelled C and S, respectively. Visualisation was mediated by his-tag detection. Molecular masses are displayed in kDa.

In both instances, uncomplexed ephrin-B1 or -B2 was easily separated from CedPV G-bound material (**Fig 3.5.1ii A**). Given an excess of ephrin-B1 and -B2, all CedPV G binding sites

were assumed to be occupied by stably bound receptor in a homogeneous complex population. Interestingly, putative complexes of both receptors displayed a range of elution volumes, despite SDS-PAGE confirmation of the presence of predominantly one CedPV G-ephrin-B1/-B2 species (**Fig 3.5.1ii B**). To assess the concentration dependence of the varying elution volumes of CedPV G-ephrin complexes, identical samples of CedPV G-ephrin-B2, at four concentrations, were subject to SEC (**Fig 3.5.1ii B**). At the highest concentration used, CedPV G-ephrin-B2 eluted in a continuum spanning 9.5 to 15.5 ml. The breadth elution volume reduced alongside concentration of CedPV G-ephrin-B2 loaded. The SEC and SDS-PAGE analyses (**Fig 3.5.1ii B & D**) suggest that CedPV G-ephrin-B2 undergoes concentration dependent oligomerisation in solution.

Intriguingly, SDS-PAGE of the CedPV G-ephrin-B2 complex reveals that the apparent molecular mass of CedPV G decreases with decreasing elution volume. As such, it appears that the less massive isoforms within the CedPV G population appear to preferentially elute as part of the larger oligomeric species. It is therefore possible that CedPV G isoforms bearing fewer glycans have an increased propensity to oligomerise, in a concentration dependent manner, and thus elute earlier than their apparently more massive counterparts bearing larger glycans. Given the observed solution-state heterogeneity that may be attributed to deglycosylation-induced oligomerisation, it would be of value to trial expression in an alternative system, such as HEK293S cells, that would be expected to yield homogeneously truncated glycans, small enough to be appropriate for crystallisation experiments but large enough to potentially inhibit the observed oligomerisation.

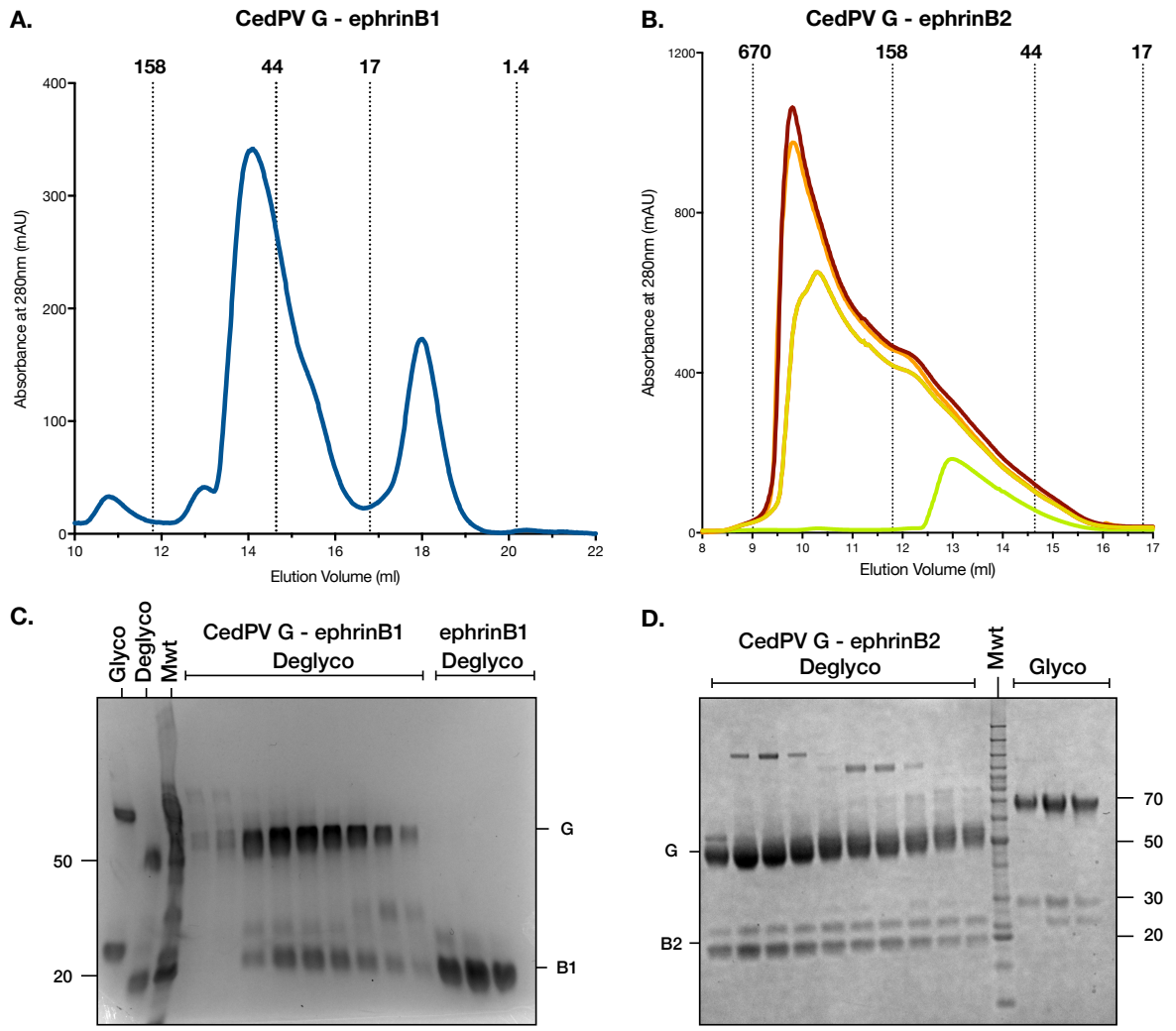


Figure 3.5.1ii: Purification of CedPV G–ephrin-B1 and –ephrin-B2 complexes. (A) Size exclusion chromatogram of kifunensine treated and endoglycosidase F1 deglycosylated CedPV G co-expressed in the presence of ephrin-B1. The molecular masses of size standards are indicated as dashed lines and correspond to γ-globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa), and Vitamin B12 (1.4 kDa). The asymmetric peak that straddles 44 kDa was putatively assigned as complex and the peak at 18 ml elution volume, assigned as excess ephrin-B1. (B) Size exclusion chromatogram of kifunensine treated and endoglycosidase F1 deglycosylated CedPV G co-expressed in the presence of ephrin-B2. Unlike the sample in A, excess ephrin-B2 was removed prior to the analysis, thus accounting for an absence of a peak at around 18 ml. Decreasing concentrations of a common pool of CedPV G-ephrin-B2 complex were loaded (coloured red to green with decreasing concentration) and run under identical conditions. The complex appears to exhibit concentration dependent oligomerisation, eluting with increasing apparent molecular mass in the higher concentration runs. Size standards are shown as in panel A, with the addition of thyroglobulin (670 kDa). (C) Reducing SDS-PAGE analysis of the peak displayed in panel A. Pooled protein collected prior to (glyco) and following deglycosylation (deglyco) with endoglycosidase F1 are shown to the left of the ladder (mwt). Tentative molecular weight positions were assigned owing to apparent gel artefacts. Fractions from the two peaks in panel A are displayed to the right of the ladder. (D) Reducing SDS-PAGE analysis of CedPV G-ephrin-B2 complexes. Fractions from the red chromatogram shown in B are displayed to the left of the ladder and representative fractions prior to endoglycosidase F1 treatment are displayed to the right.

3.5.2 Crystallisation

Purified CedPV G-ephrin-B2 and -B1 complexes were concentrated to between 4.2 and 12.2 mg/ml for crystallisation trials. Proteins were subjected to the full range of crystal screens developed in-house at the Division of Structural Biology (Oxford, UK), and were further screened based upon temperature (4°C and 25°C). Given the apparent glycan-dependent oligomeriation induced by CedPV G, both CedPV G-ephrin-B2 and -B1 complexes were screened for crystal formation using protein samples that were either treated with, or left untreated by endo F_1, though both expressed in the presence of kifunensine. Crystals were obtained in a variety of conditions for both complexes, however, none yielded useable X-ray diffraction beyond approximately 10 Å.

3.6 CedPV G structure determination

3.6.1 Crystal harvesting and X-ray diffraction data collection

The CedPV G crystal was harvested and cryo-protected by transfer into a drop of mother liquor supplemented with 25% glycerol. Following cryo-protection, the crystal was flash-cooled by immersion in liquid nitrogen and maintained at cryogenic temperature thereafter. X-ray diffraction data were collected to 2.78 Å at beamline I02 at Diamond Light Source (UK) and processed using xia2 (Winter, 2010) (**Table 3.6.1**).

Table 3.6.1: CedPV G data collection statistics.

Data collection statistics	CedPV G
Beamline	DLS I02
Wavelength (Å)	0.9795
Space group	$P6_5$
Cell dimensions *a*, *b*, *c* (Å)	201.5, 201.5, 112.9
α, β, γ (°)	90, 90, 120
Resolution range (Å)	58-2.78 (2.85-2.78)[a]
R_{merge} [b]	0.164 (>1)
Number of unique observations	65,603 (4,849)
I/σ(I)	15.4 (2.0)
$CC_{1/2}$	0.998 (0.777)
Completeness (%)	99.9 (99.6)
Multiplicity	20.5 (19.5)

[a] The value for the highest resolution shell is show in parentheses.
[b] $R_{merge} = \Sigma_{hkl} \Sigma_i|I(hkl;i) - \langle I(hkl)\rangle|/\Sigma_{hkl} \Sigma_i I(hkl;i)$, where I(hkl;i) is the intensity of an individual measurement and $\langle I(hkl)\rangle$ is the average intensity from multiple observations.

3.6.2 Crystal structure determination

Phase estimates were obtained by molecular replacement, using the crystal structure of the unliganded G protein from NiV (PDB 2VWD). Iterative rounds of model building and refinement were performed using COOT (Emsley and Cowtan, 2004) and PHENIX (Adams et al., 2010), respectively. Non-crystallographic symmetry restraints were implemented during refinement. Refinement statistics are presented in Table 3.6.2.

Table 3.6.2: CedPV G crystallographic refinement statistics.

Refinement statistic	CedPV G
Resolution (Å)	49-2.78
No. reflections	65,557
R_{work}[a]/R_{free}[b]	0.203/0.230
Molecules per a.s.u[c]	2
No. atoms	
Protein	6,698
Ligand/ion	0
Water	9
Glycan	182
B-factors	
Protein	86.5
Ligand/ion	n/a
Water	62.2
Glycan	102.8
Ramachandran(%)	
Favored	94.8
Allowed	5.2
Outlier	0
RMSD[d]	
Bond lengths (Å)	0.003
Bond angles (°)	0.730

[a] $R_{work} = \Sigma_{hkl}||F_{obs}| - k|F_{calc}||/\Sigma_{hkl}|F_{obs}|$
[b] R_{free} was calculated as for R_{work}, using a consistent set of reflections representing 5% of the total, removed prior to refinement.
[c] a.s.u: crystallographic asymmetric unit.
[d] RMSD: root mean square deviation from ideal geometry.

The two molecules present within the crystallographic asymmetric unit were essentially identical (RMSD 0.3 Å over 416 equivalent Cα positions) and electron density permitted modelling of the entire crystallised construct (residues K209–C622) in both instances. Owing to a lower *B*-factor, chain A (77) was selected over chain B (96) for further analyses. As such, all reference to CedPV G hereafter refers to chain A unless otherwise specifically stated.

3.7 CedPV G structure analysis

3.7.1 Structure overview

Analysis of the crystal structure of CedPV G reveals that the glycoprotein adopts the canonical six-bladed β-propeller fold that is observed both within the henipaviruses, NiV, HeV, and GhV (Bowden et al., 2008b, Bowden et al., 2010a, Lee et al., 2015), and more broadly across paramyxoviral attachment (G)/hemagglutinin (H)/hemagglutinin-neuraminidase (HN) proteins (Crennell et al., 2000, Bowden et al., 2010b). Each of the six blades, designated β1-β6, comprise four antiparallel β-strands, some of which are discontinuous, that assemble around a central axis (**Fig 3.7.1 A**). Like the prototypic henipaviruses, the β2 and β3 blades are decorated with extended loops at their 'top' surface that form three short 3_{10} helical segments (η1–η3) and three longer α-helices (α1– α3). Eight disulphide bonds stabilise the fold, seven of which are conserved amongst other henipavirus G proteins. The additional disulphide (C310–C376) in CedPV G connects blades β2 and β3 at the base of the molecule, via the first strand in each.

Comparison of CedPV G with unliganded HeV G (PDB 2X9M) and NiV G (PDB 2VWD) reveals structural similarity within the β-propeller scaffold, with RMSD values of 1.7 Å calculated over 385 Cα positions and 1.6 Å over 371 Cα positions, respectively (**Fig 3.7.1 B)**. While the core of β-propeller scaffold is well conserved, structural deviations map principally to loop regions that connect the strands within each blade, at the apex of the molecule and proximal to the putative receptor binding site. Interestingly, structural conservation within the β-propeller scaffold can be broadly divided into two continuous sections; the β1-3 and β6 blades and their associated loops exhibit substantially elevated RMSD values than strands β4-5 (**Fig 3.7.1 C**).

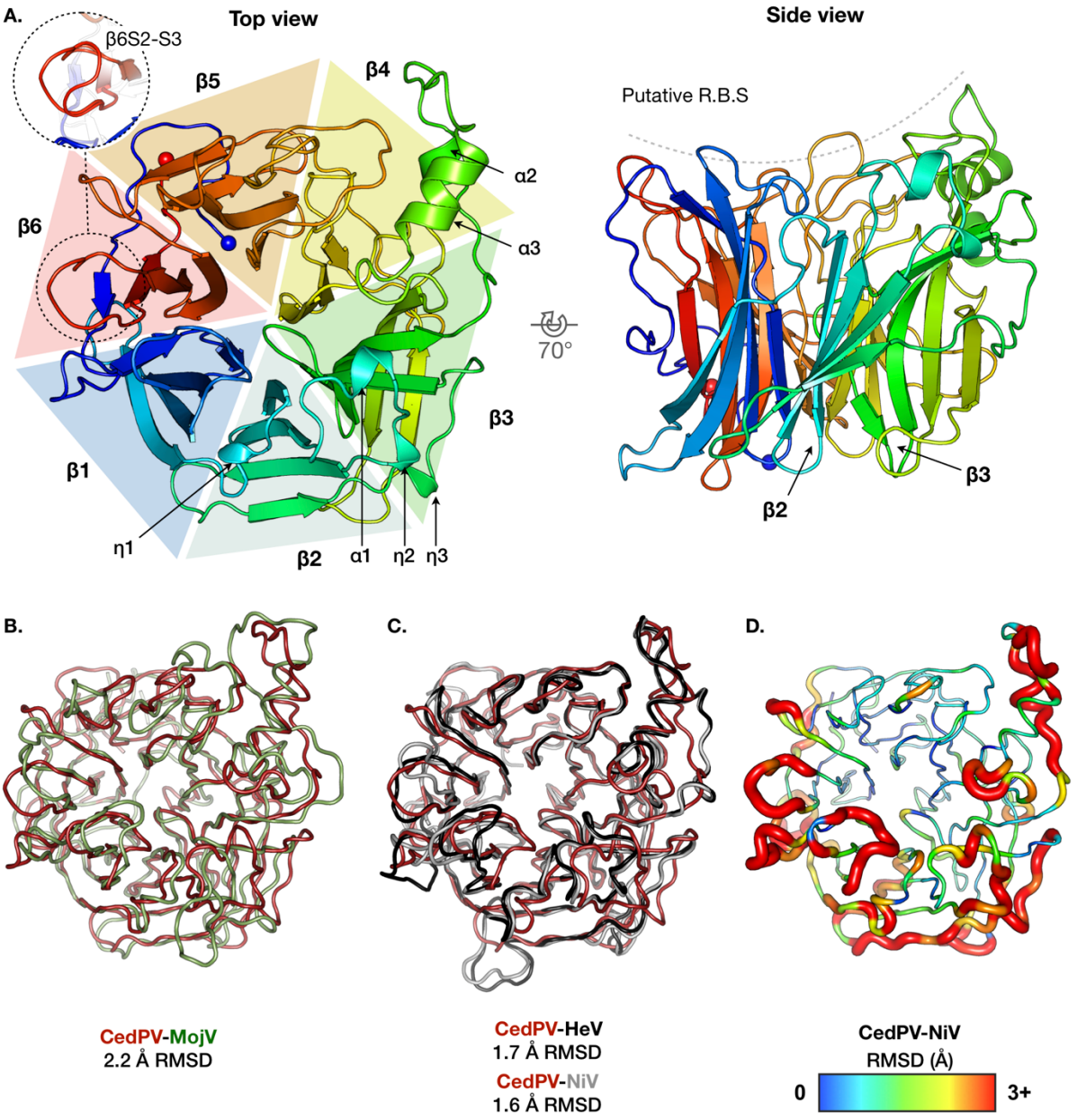


Figure 3.7.1: CedPV G adopts a conserved β-propeller fold. (A) CedPV G is displayed as a cartoon coloured from N- to C-terminus (blue to red), with the termini shown as spheres. The extent of each of the six 'blades' of the β-propeller is denoted with coloured triangles and are labelled (β1- β6). The region of highest variability (β6S2-S3 loop) between the asymmetric unit copies of the protein is shown with an inset panel (dashed line). Each blade comprises four antiparallel strands that run outward from the central axis. Alpha- and 3_{10}-helical segments are labelled as α1-3 and η1-3, respectively. The protein is displayed in a 'top' view (left) and rotated 70° to reveal a side view. The putative receptor binding surface (R.B.S) is broadly depicted with a dashed line. (B & C) Structural overlay of CedPV G (red) and the G proteins from the putative henipavirus Mojiang (MojV, green), and the prototypic henipaviruses Hendra (HeV, black) and Nipah (NiV, grey). Root-mean square deviation (RMSD) values, calculated using SSM superposition in coot, are displayed beneath. (C) Structure comparison between CedPV G and NiV G. RMSD between aligned Cα residues is depicted by both colour (ramped from blue to red with increased RMSD) and tube width (thin to thick with increased RMSD). Residues that failed to align or exhibited RMSDs greater than 3 Å were assigned RMSD values of 3 Å.

While both HeV G and NiV G are structurally very similar (0.9 Å RMSD), the conformation of the β6S2-3 loop differs substantially. Furthermore, variation is observed within the β6S2-3 loop in the ephrin-B2 bound and unbound states of HeV G (Bowden et al., 2008a, Bowden et al., 2010a). These observations suggest an inherent conformational plasticity within the β6S2-3 loop. Indeed, the analagous β6S2-3 loop in CedPV G exhibits two distinct conformations within each a.s.u copy of CedPV G and *B*-factors that are elevated above the average observed for the domain as a whole. Thus, the β6S2-3 loop of CedPV G, HeV G, and NiV G, is conformationally labile, and subject to reorganisation upon receptor binding, which may indicate a conserved functional role for this region, potentially in mediating receptor-induced activation of a cognate fusion protein (F).

3.7.2 Comparison with known paramyxoviral attachment proteins

Intriguingly, the henipavirus-like Mòjiāng virus (MojPV), classified as such based on genetic similarity with the known henipaviruses HeV, NiV, CedPV, and GhV (Wu et al., 2014), has been shown not to utilise ephrin-B2/-B3, or any of the common paramyxoviral entry receptors (Rissanen et al., 2017). Both MojV and CedPV exhibit pronounced sequence divergence within their respective G proteins (22% and 31% identical to NiV G, respectively). In light of the high level of sequence variation observed within CedPV G, a structure-based phylogeny was constructed to contextualise CedPV Gs structural relatedness to other paramyxoviral β-propellers (**Fig 3.7.2**).

Concomitant with sequence-based phylogenetic approaches, CedPV G co-localises within a cluster of ephrin-utilising henipaviruses (NiV, HeV, and GhV). CedPV G, consistent with both sequence and geographical similarity, is more closely related to the prototypic Asiatic henipaviruses than the African virus, GhV G. Localisation of CedPV G within this cluster supports initial observations that CedPV likely utilises the ephrin ligands as cellular receptors for viral entry (Marsh et al., 2012). As such, and in spite of substantial sequence divergence, the

structure of CedPV G is likely constrained by a requirement to maintain high-affinity ephrin-binding. Indeed, the structural dissimilarity of MojV G to the ephrin-utilising henipavirus and other paramyxoviral G/HN proteins, is indicative of adaptation to a novel receptor (Rissanen et al., 2017).

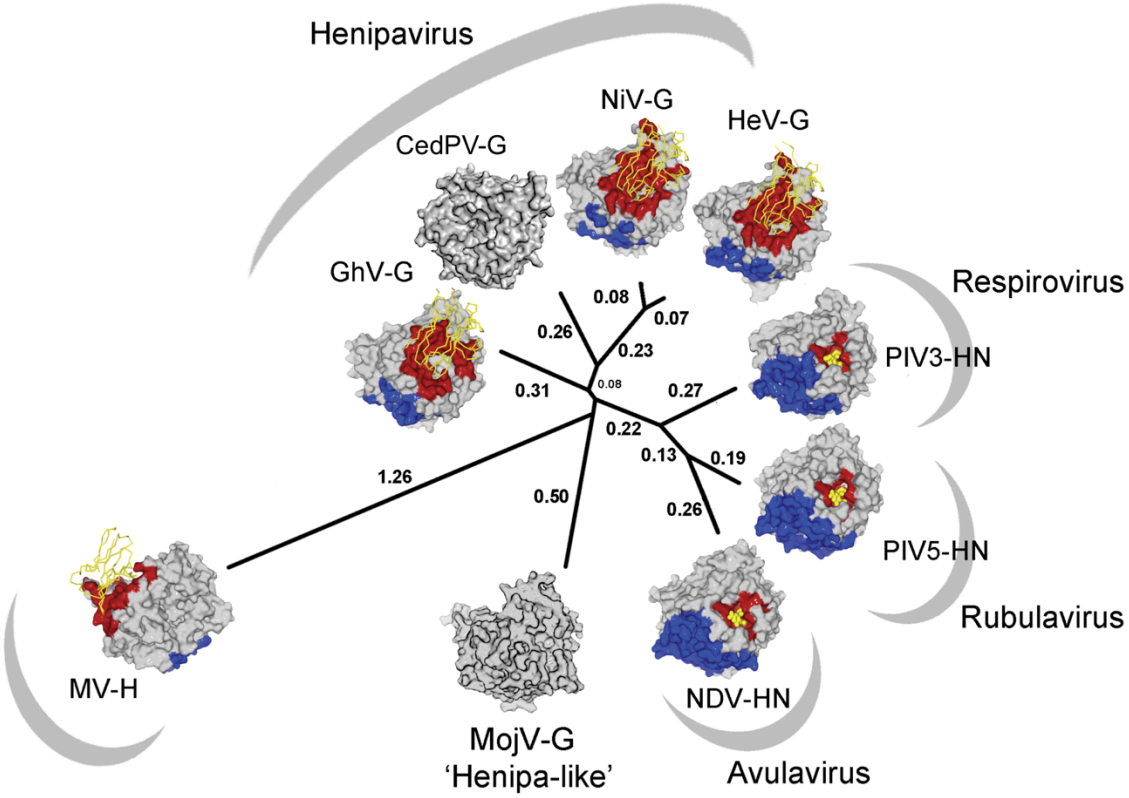


Figure 3.7.2: CedPV G is closely related to the ephrin-binding henipavirus G proteins. Structural phylogenetic analysis of paramyxovirus attachment (G), hemagglutinin (H), and hemagglutinin-neuraminidase glycoprotein structures, produced using SHP (Stuart et al., 1979). Respective G, H or HN proteins are shown as a gray surface. Proteins for which the receptor binding site (red surface with docked ligand in yellow) and homo-dimerisation interface (blue) are known, have them displayed. CedPV-G clusters with the genetically and geographically related heniparivus G proteins. Accession codes and full species names for the structures used in the analysis are: Ghanaian bat henipavirus (GhV-G), 4UF7; Cedar virus (CedPV-G, not-deposited); Nipah virus (NiV-G), 2VSM; Hendra virus (HeV-G), 2VSK; Human parainfluenza virus 3 (HPIV3-HN), 1V3C; Human Parainfluenza virus 5 (PIV5-HN), 1Z4V; Newcastle Disease virus (NDV-HN), 1E8V; Mojiang virus (MojV-G), 5NOP; Mumps virus (MV-H), 5B2D.

The β-propeller architecture is highly conserved amongst paramyxoviral attachment proteins (G/HN/H), despite pronounced sequence variability and requirements to recognise diverse ligands (Bowden et al., 2010b). Despite substantial primary sequence diversification, even amongst viruses that utilise common receptors (e.g CedPV and NiV are 31% identical yet

maintain ephrin binding), the necessity to conserve specific receptor tropisms likely constrains structural variability. As such, structural phylogenetic classification represents a powerful tool that can delineate distinct structural families that are likely to possess common receptor binding characteristics. Indeed, such analyses support a conserved mode of ephrin recognition by the G protein of CedPV that are consistent with *in vitro* viral entry experiments (Marsh et al., 2012, Laing et al., 2018).

3.7.3 Glycosylation of CedPV G and a model for self-association

The CedPV G protein exhibits the greatest number of N-linked glycosylation sequons of any know henipavirus attachment glycoprotein. Within the receptor-binding β-propeller domain, CedPV G possesses eight sequons whilst NiV, HeV, GhV, and MojV contain five, five, seven, and zero, respectively. Interestingly, only one site ($N502^{CedPV}$) is conserved, at the primary sequence level, with either of the prototypic henipaviruses, HeV and NiV. The equivalent glycosylation sites to $N502^{CedPV}$ in HeV and NiV has been demonstrated to modulate fusion activation. In both instances, mutational removal of the N-linked glycan at this position yielded hyperfusogenic viruses (Bradel-Tretheway et al., 2015). Thus, it is plausible that $N502^{CedPV}$ may possess conserved fusion-regulating functionality within CedPV.

Glycosylation of HeV and NiV has been demonstrated to have a profound impact upon protein stability, G-F interactions, cell-fusion, viral entry, and shielding from antibody-mediated neutralisation (Biering et al., 2012, Bradel-Tretheway et al., 2015). Given the absence of conserved glycans between the prototypic henipaviruses and CedPV G, it is plausible that no single glycan site, other than N502, is indispensable for general henipavirus G function. However, it is possible that the glycosylation of henipaviral G proteins is co-adapted to regulating homotypic fusion through species-restricted G-F interactions. Indeed, CedPV G has been demonstrated to be incapable of heterotypically activating NiV or HeV F proteins (unpublished data).

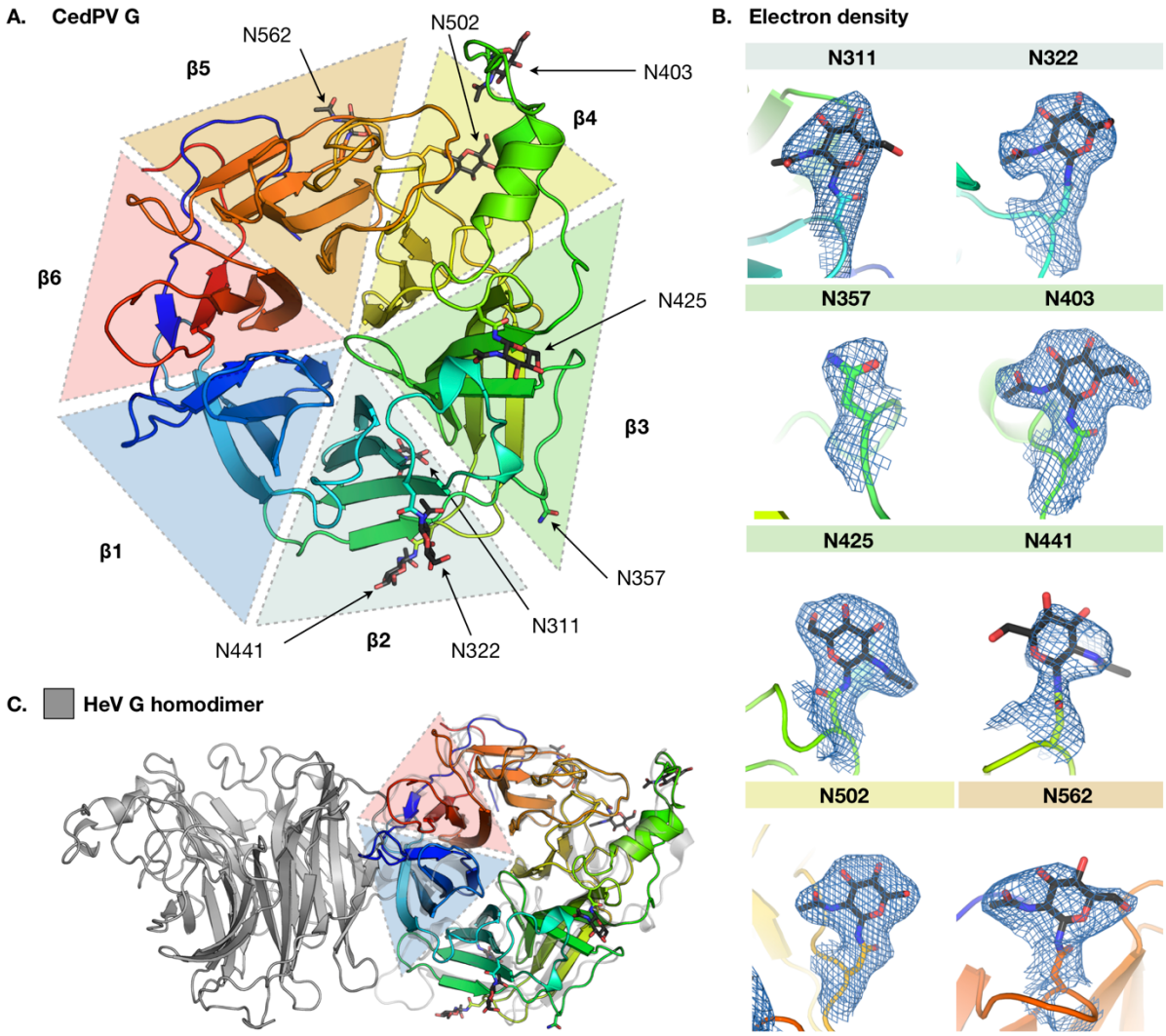


Figure 3.7.3: N-linked glycosylation of CedPV G is consistent with a conserved mode of self-assembly. (A) Distribution of N-linked glycans on CedPV G. CedPV G is shown as a cartoon coloured from N- to C-terminus (blue to red). Regions roughly defining the blades of the β-propeller are coloured from blue to red. Modelled N-linked glycans are shown as sticks coloured according to constituent elements: carbon is black, nitrogen is blue, and oxygen is red. N-linked glycosylation sites are found on the β2-β5 blades, inclusive. (B) Expanded view of modelled glycans and asparagine residues from canonical N-linked glycosylation sequons. Refined *2Fo-Fc* electron density maps for sites of interest are contoured to 1σ and displayed as blue mesh. (C) CedPV G glycosylation is consistent with the homo-dimer observed in HeV G. Dimerisation in HeV is mediated by the β1 and β6 blades (blue and red, respectively) on a surface devoid of glycosylation sites. CedPV G is shown superposed onto one molecule within the HeV G dimer (PDB 2X9M) which has one protomer shown as transparent, for clarity.

Despite the positional variability of N-linked glycans within ephrin-utilising henipaviral G proteins, the overall conservation of a highly glycosylated β-propeller is suggestive of a functional constrain placed upon maintaining glycosylation. As such, analogous studies

functional of CedPV G glycosylation would be of value to elucidate the functional importance of both the conserved N502 and additional glycosylation sites.

Of eight N-linked glycosylation sequons present within CedPV G, evidence for seven N-acetylglucosamine moieties was observed that permitted model building (N311, N322, N403, N425, N441, N502, and N562) (**Fig 3.7.3 B**). In each instance, the $2F_o$-F_c electron density was suggestive of single N-linked N-acetylglucosamine moieties, consistent with the complete action of endoF_1 upon the kifunensine-induced oligomannose-type glycans (Chang et al., 2007). An absence of unambiguous electron density at position N357 precluded modelling of an N-acetlyglucosamine moiety in either of the a.s.u copies of the protein. If present within the crystallised protein isoform, glycans at position N357 would have access to a large solvent channel and would thus likely be flexible. To this end, features of the $2F_o$-F_c and the F_o-F_c maps suggest this site may be glycosylated, however it is likely that the glycans are either present at low occupancy or are flexible. Owing to the size of the protein and the number of glycosylation sites, mobility shift on SDS-PAGE is not sufficiently accurate to determine the likelihood of glycosylation at this site, and as such, further experimentation would be required to confirm the presence of N-linked glycans at N357.

Localisation of the glycosylation sites within the β2-5 blades (**Fig 3.7.3 A**) is consistent with observations that henipavirus G proteins are likely to homodimerise at an interface comprising the β1 and β6 blades β-propeller (Bowden et al., 2010a). Overlay analysis with the crystal structure of dimeric HeV G reveals that CedPV G exhibits no obvious structural features that would preclude self-association in a mode similar to HeV G (**Fig 3.7.3 C**). Crystal packing within the a.s.u does not appear to be biologically relevant concomitant with this hypothesis, since the β1 and β6 blades of 'molecule A' contact sheets β3 and β4 of 'molecule B', forming an interface that contains a number of glycan moieties that, in a native un-trimmed context, would likely impede the protein-protein interactions observed in the crystal.

3.7.4 Mapping sequence conservation within CedPV G

Despite common utilisation of ephrin-B2 as a functional cellular receptor, the attachment-mediating G protein from CedPV possesses limited sequence identity with GhV G, HeV G, and NiV G (26-33%). To assess the distribution of conserved amino acids within CedPV G, structure- and sequence-based mapping approaches were employed (**Fig 3.7.4i & ii**).

Analysis of a sequence alignment of amino acids comprising the β-propeller domain from CedPV G, NiV G, HeV G, and GhV G reveals that conservation of amino acids within the henipaviral G proteins is not evenly distributed across the entire domain. Qualitatively, it appears that the β4-6 blades are more conserved than β1-3, with β2-3, specifically, displaying few strictly conserved residues. The distribution of conserved amino acids is in good agreement with structural conservation of the β4-5, and to a lesser extent β6, blades (**Fig 3.7.1 C & D**). Furthermore, the crystal structures of the ephrin-B2 bound G proteins from HeV, NiV, and GhV have shown that β4-6 contribute the majority of the ephrin-B2 interacting residues (**Fig 3.7.4i**). Thus, it is likely that the primary sequence conservation, and concomitant tertiary structure similarity, reflects a constraint upon maintaining a conserved mode of ephrin-B2 recognition.

When mapped onto the surface of CedPV G, it is evident that amino acid conservation is elevated within the putative receptor binding site (**Fig 3.7.4ii**). When superimposed onto the structure of CedPV G, ephrin-B2, obtained from alignment with the NiV G-ephrin-B2 complex (PDB 2VSM), is in proximity to a contiguous patch of conserved (both identical and similar) residues at the 'top' surface of the β-propeller. Consistent with sequence-based analysis, these conserved residues are principally found in loop regions within β1 and β4-6 (**Fig 3.7.4ii**). Quantitative analysis of the surface distribution of conserved amino acids further supports the notion that sequence diversification within CedPV G is constrained by the requirement to bind ephrin-B2. When compared with to NiV G, the sequence identity within the putative receptor binding site

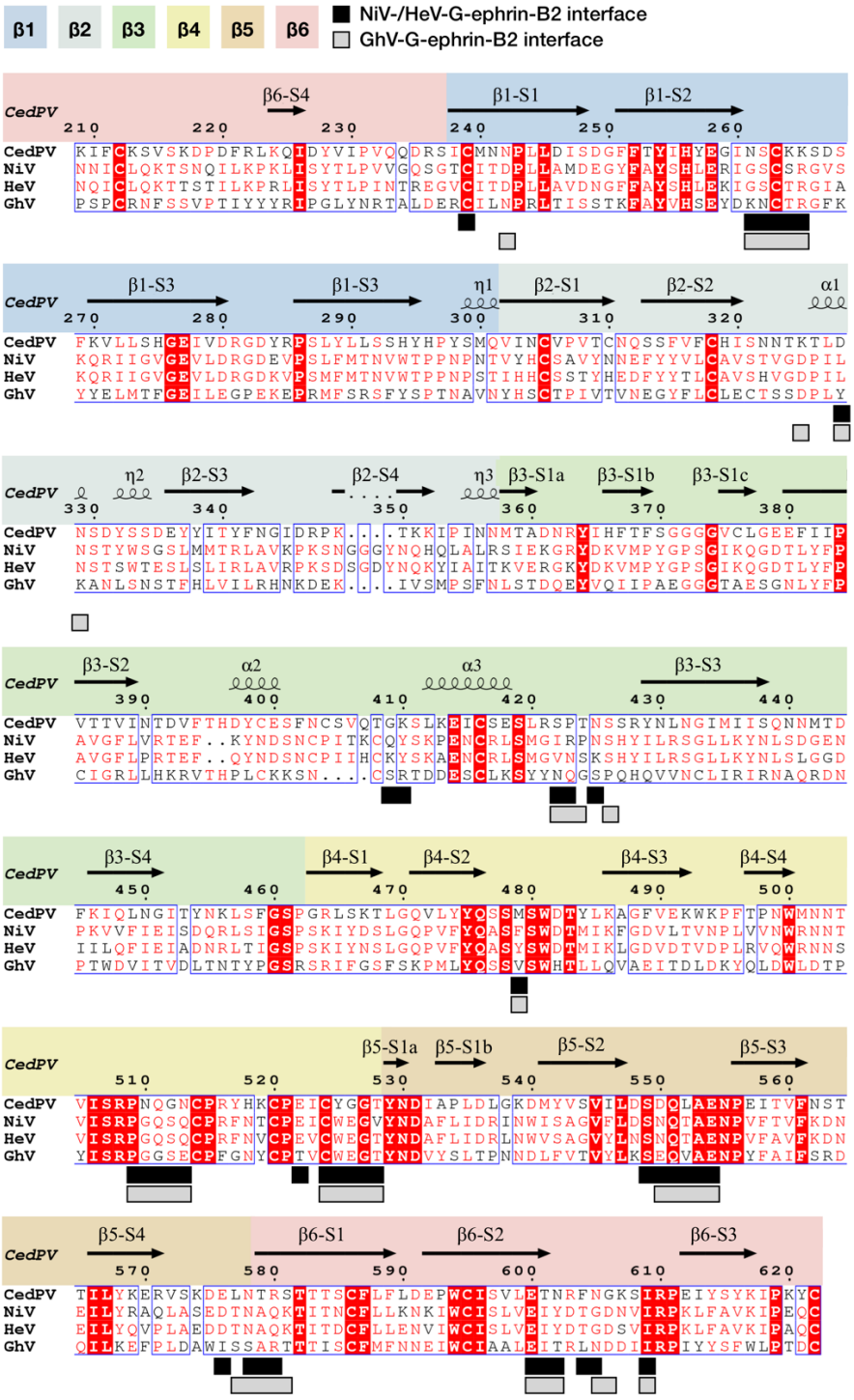
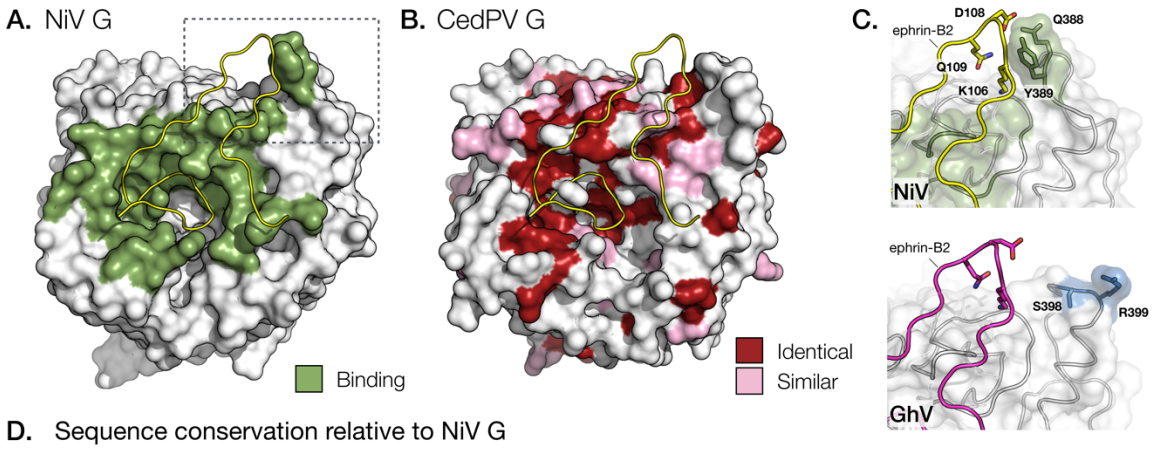


Figure 3.7.4i: Sequence alignment of the receptor-binding β-propeller domain of the ephrin-utilising henipaviruses. Sequence alignment of the attachment glycoproteins (G) from CedPV, NiV, HeV, and GhV. Absolutely conserved residues are highlighted in red, partially conserved residues are coloured red, and non-conserved residues are coloured black. Residues known to interact with ephrin-B2 in NiV/HeV G and GhV G complexes are denoted with black and grey boxes below the sequence, respectively. Secondary structure elements of the CedPV G crystal structure are annotated above the sequence and labelled according to type: β strands are shown with arrows and labelled according to β-blade (β1-6) and strand number within a blade (S1-4) e.g β1-S4; α helices are illustrated as coils and labelled α1-3; 3_{10} helices are also labelled with a coil and designated η1-3. The blades of the β-propeller are coloured from blue to red. Sequence alignments were determined by MultAlin (Corpet, 1988) and plotted using ESPript (Gouet et al., 2003).

of CedPV is elevated relative to the domain as a whole (40% versus 31%), as is the case in GhV G (42% versus 28%).

Despite overall sequence (28% identity) and, to an extent, structural divergence (2.2 Å RMSD) of GhV G, the observed ephrin-B2 binding mode is remarkably similar to that of NiV and HeV Gs (Bowden et al., 2008a, Lee et al., 2015). Thus, combined with CedPVs increased similarity to both NiV G and HeV G, the marked conservation of the ephrin-B2 binding mode in GhV G strongly suggests ephrin-B2 is likely to engage CedPV G in a similar fashion.



D. Sequence conservation relative to NiV G

	HeV G	CedPV G	GhV G	MojV G
Seq Identity	81%	31%	28%	22%
Binding site identity	78%	40%	42%	15%
Ephrin-B2/-B3 binding	yes/yes	yes/no	yes/no	no/no

Figure 3.7.4ii: Amino acid conservation within CedPV G is elevated at the putative ephrin-B2 interface. Structural mapping of the ephrin-B2 binding interface (as defined by Bowden et al. (2008a)) and amino acid conservation within henipaviral attachment (G) proteins. (A) NiV G is displayed as a grey surface with interacting regions of bound ephrin-B2 shown as a yellow tube. Non-interacting regions of ephrin-B2 are removed for clarity. Residues in contact with ephrin-B2 are coloured green. (B) CedPV G is shown oriented as in panel A, with the same region of ephrin B2 superposed, based on alignment of the NiV G-ephrin-B2 complex. Amino acids within CedPV G that are identical (red) or similar in physiochemical properties (pink) are mapped onto the surface. (C) Expanded view of the secondary binding within NiV G (dashed line in A) and the equivalent regions within the GhV G-ephrin-B2 complex and superposed complex from panel B. Critical residues within the secondary binding site of NiV G are shown as sticks. Aligned residues (blue) are shown for GhV G and CedPV G. Like GhV G, CedPV appears to lack the secondary binding site. (D) Quantitative sequence conservation distribution analysis. Amino acid sequence identity relative to NiV G is shown for the entire G protein and within the ephrin-B2 binding site. All ephrin-utilising henipaviruses exhibit elevated sequence conservation within the receptor binding site.

A comparison of the residues within the NiV G- and GhV G-ephrin-B2 complexes reveals a largely overlapping distribution of amino acids utilised to mediate receptor recognition (**Fig 3.7.4i)**. Interestingly, a single contact site within the α2-α3 loop of NiV G and HeV G is not conserved, nor substituted by other similar interactions, within the GhV G-ephrin-B2 complex (**Fig 3.7.4ii C)**. This 'secondary binding site', so called because it is discontinuous with the majority of the ephrin-binding surface, accounts for a 15% reduction in the GhV G-ephrin-B2 interface relative to that observed in the NiV G complex (Lee et al., 2015). Furthermore, reciprocal exchange of NiV and GhV residues within this loop resulted in mutant G proteins with decreased (NiV $G^{QY388-389RT}$) and increased (GhV $G^{RT399-400QY}$) ephrin-B2 binding and cellular entry properties, respectively. To this end, it is interesting to note that CedPV G does not retain the QY motif in its α2-α3 loop, but instead possesses a $GK^{409-410}$ motif that is unlikely to form the secondary binding site. The α2-α3 loop of CedPV G is longer than equivalent regions of both NiV and GhV, however, inspection of a model complex derived from superposition of NiV G- and GhV G-ephrin-B2 complexes, does not reveal any residues likely to mediate contacts with ephrin-B2 (**Fig 3.7.4ii**). Furthermore, as both the α2 and α3 helices and their connecting loops are essentially unaltered by ephrin-B2 binding in both HeV G and NiV G, it is unlikely that this region would be significantly different in ephrin-bound CedPV G (Bowden et al., 2008a, Bowden et al., 2008b, Bowden et al., 2010a). Thus, CedPV is unlikely to form contacts with ephrins outside of the continuous primary interface that, whilst apparently dispensable to ephrin-B2 binding, may be crucial to engender sufficiently high affinity binding to permit functional ephrin-B3 utilisation.

3.7.5 Structural features of the putative receptor binding site of CedPV G

The crystal structures of henipavirus G proteins in complex with ephrin-B2 have revealed that the interaction interface is extensive (up to 2,800 Å), and dominated by the insertion of the GH loop (F114–N129) of ephrin-B2 into a central cavity at the apex of the β-propeller, in a binding

mode that is strikingly similar in all cases (Bowden et al., 2008a, Lee et al., 2015). To assess the likelihood that CedPV G employs a similar mode of ephrin-B2 recognition, the structure of the NiV G-ephrin-B2 complex (PDB 2VSM) was superimposed onto CedPV G and the resultant interface analysed (**Fig 3.7.5**).

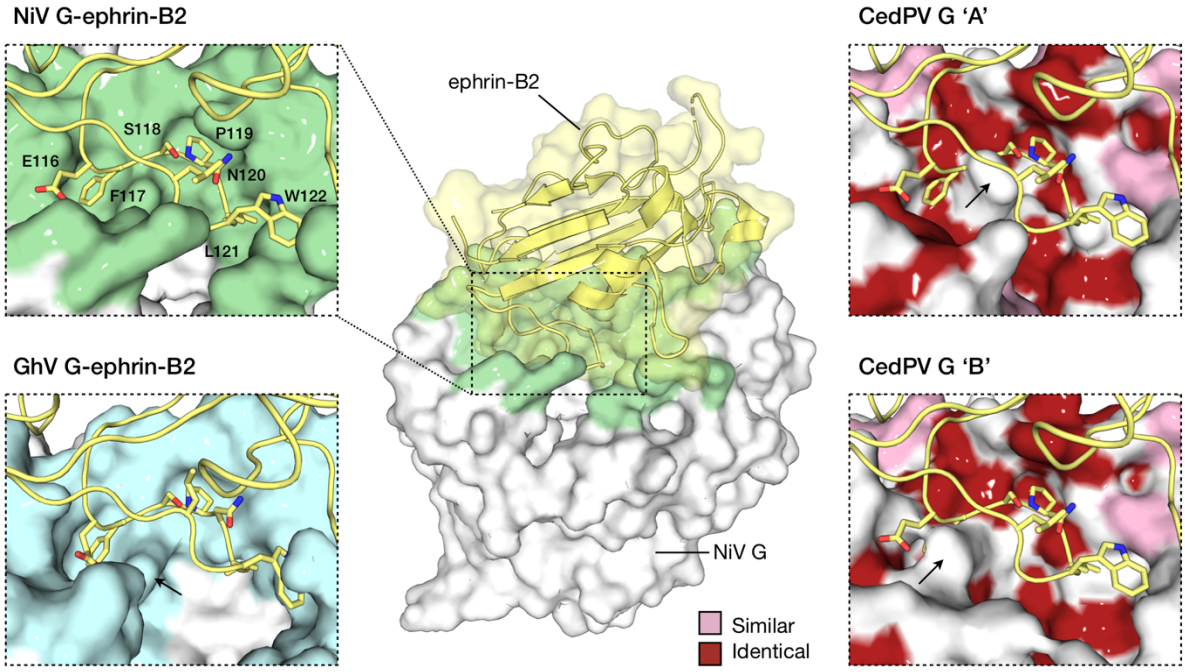


Figure 3.7.5: The ephrin binding surface of CedPV is topologically conserved. The overall architecture of the NiV G protein (grey surface) in complex with ephrin-B2 (yellow transparent surface & cartoon) is shown (PDB 2VSM). Residues of NiV G that contact ephrin-B2 are coloured green. Inset boxes show an expanded view of the GH loop of ephrin-B2 and the central cavity into which it is inserted (grey surface), indicated with a dashed line in the central panel. Critical interacting residues within the GH loop are labelled in the NiV G-ephrin-B2 panel only. (bottom left) The GH loop interaction within the GhV G-ephrin-B2 crystal structure (PDB 4UF7). Residues that interact with ephrin-B2 are coloured blue. (right top & bottom) Superimposition of ephrin-B2 onto the two a.s.u copies of CedPV G, labelled by chain identity (A and B). The surface of CedPV is coloured according to sequence conservation relative to NiV G, with similar residues coloured pink and identical residues coloured red. The position of a conformationally labile arginine side chain ($R580^{CedPV}$) that is also present within GhVG ($R569^{GhV}$) is indicated with an arrow in all instances.

Analysis of the superimposed ephrin-B2 contact surface of CedPV G reveals that CedPV G, like the other henipaviral G proteins, forms a depression at its apical surface that is sufficiently large to accommodate the GH loop of ephrin-B2 in a mode consistent with that observed in other henipavirus G-ehprin-B2 binding events (**Fig 3.7.5**). Conservation of the receptor binding surface of CedPV G is consistent with RMSD variance analysis (**Fig 3.7.1 C & D**) that revealed

the β4-6 blades of CedPV to be strikingly conformationally conserved in NiV G and HeV G, despite some local differences in amino acid composition. Consistent with this, the β4-6 blades form a surface that contribute a number of residues that accommodate the bulky side chains of Fl17 and Pl19 of ephrin-B2.

Interestingly, despite the overall topological complementarity of the proteinaceous surface, the side chain confomer of an arginine residue ($R580^{CedPV}$) in CedPV G appears to present a steric barrier to ephrin-B2 by extending upward into a position that would likely be occupied by the main chain atoms within the GH loop of ephrin-B2 (**Fig 3.7.5**). Analysis of the alternate copy of CedPV G within the crystallographic a.s.u reveals similar steric incompatibility between R580 and ephrin-B2. Both arginine residues, however, are not identically positioned in each a.s.u copy, suggestive that R580 is capable of undergoing positional rearrangement which would likely accommodate ephrin-B2 binding. Indeed, GhV G possesses an arginine at the equivalent position (global SSM alignment places Cα atoms within 1 Å) to $R580^{CedPV}$ ($R569^{GhV}$), that does not preclude ephrin-B2 binding. Thus, an arginine at position 580^{CedPV} is not necessarily incompatible with ephrin-B2 recognition, furthermore R580 likely only requires subtle rotameric rearrangement to permit ephrin-B2 binding.

3.7.6 Molecular determinants of differential ephrin tropism

Like all known classical henipaviruses (excluding MojV), ephrin-B2 has been established as a *bona fide* cellular entry receptor for CedPV. Interestingly, CedPV, like GhV, appears incapable of either binding or utilising ephrin-B3 (Marsh et al., 2012, Laing et al., 2018). Solution state binding and cellular entry assays performed by collaborators in the laboratory of Prof. Benhur Lee (Icahn School of Medicine at Mount Sinai, USA), have both confirmed previous receptor tropism studies, and further demonstrated that, uniquely, CedPV G is able to utilise ephrin-B1 for cellular entry (unpublished data). In support of this, SEC analysis of co-expressed CedPV G and ephrin-B1 demonstrate that the both proteins form a stable heteromeric complex (**Fig 3.5.1ii**

A). In the absence of a deposited co-ordinate file for the structure of ephrin-B1, and given its structural similarity with ephrin-B2 (Nikolov et al., 2005), structural analysis of the molecular determinants of differential ephrin tropism was performed with reference to the structure of NiV G bound to ephrin-B2 and ephrin-B3 (Bowden et al., 2008a, Xu et al., 2008).

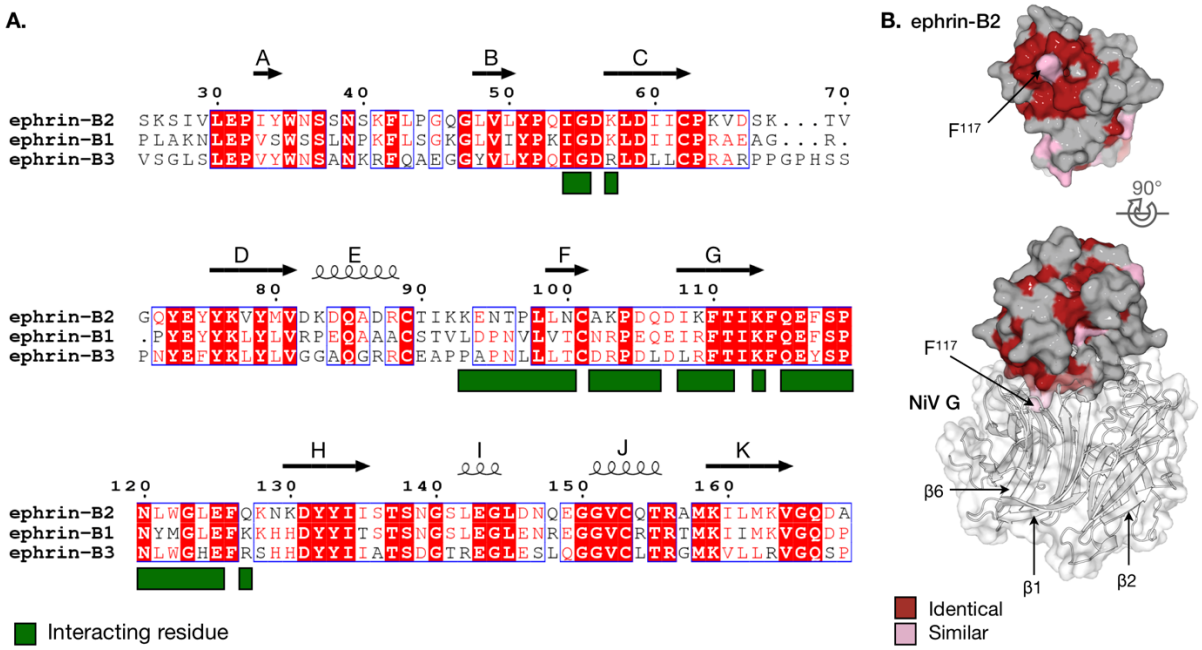


Figure 3.7.6i: Henipaviruses recognise a surface that is highly conserved amongst ephrin-B homologues. (A) Sequence alignment of the globular extracellular domains of human ephrin-B1, -B2, and -B3. Absolutely conserved residues are highlighted in red, partially conserved residues are coloured red, and non-conserved residues are coloured black. Residues known to interact with NiV and HeV G proteins are indicated with green boxes below the sequence. Secondary structure elements of ephrin-B2 are labelled above the sequence: beta strands are shown as arrows and helical segments as coils. Secondary structures are labelled A-K according to established ephrin nomenclature convention (Himanen et al., 2001). Sequence numbering corresponds to ephrin-B2 (UniProt P52799). Sequence alignments were determined by MultAlin (Corpet, 1988) and plotted using ESPript (Gouet et al., 2003). (B) Mapping ephrin B family conservation onto the surface of ephrin-B2 from its NiV G-bound complex (PDB 2VSM). Ephrin-B2 is shown as a grey surface coloured according to residues that are identical (red) or similar (pink) across ephrin-B1, -B2 and -B3. A prominent phenylalanine (F117) within the crucial GH loop of ephrin-B2 is labelled for orientation purposes. NiV G is shown as a transparent white surface with the visible β-blades labelled.

Sequence-based analysis of the B-class ephrins reveals that the G-interacting regions of the molecule are highly conserved, particularly the crucial GH loop, which constitutes the primary interacting motif of ephrin-B2 and -B3 (**Fig 3.7.6i A**). When mapped onto the surface of ephrin-B2, conservation within the GH loop constitutes a continuous surface with elevated

amino acid conservation relative to the domain as a whole (**Fig 3.7.6i B**). Interestingly, Fl17, which is centrally located within the conserved GH loop patch (**Fig 3.7.6i B**), is not absolutely conserved in ephrin-B3, which possesses a tyrosine at the equivalent position. Whilst the conserved hydrophobic cavity in CedPV G is large enough to accommodate the additional hydroxyl group present on tyrosine, it is plausible that binding of a phenylalanine is favoured due to the hydrophobic nature of the pocket. Indeed, substitution of Fl17 for tyrosine within ephrin-B2 (ephrin-B2^{Fl17Y}) diminished binding of ephrin-B2 to NiV G, though only slightly (K_d 35 nM versus 86 nM)(Bowden et al., 2008a).

In addition to the evidence implicating Fl17$^{ephrin\text{-}B2}$ as a determinant of differential ephrin utilisation, a leucine-tryptophan motif (L121-W122$^{ephrin\text{-}B2}$), conserved within ephrin-B2 and -B3, has been demonstrated to be critical for mediating NiV G binding (Negrete et al., 2006). Interestingly, L121$^{ephrin\text{-}B2}$ and W122$^{ephrin\text{-}B2}$, are not present in the equivalent positions within the otherwise highly conserved GH loop of ephrin-B1, which instead possesses a Y-M motif (Y121-M122$^{ephrin\text{-}B1}$). Substitution of the Y-M motif in ephrin-B1 for L-W (ephrin-B1^{LW}) is sufficient to permit NiV G binding, and thus henipavirus G protein residues that interact with the Y-M/F-W motif within ephrin-B1/-B2 may represent a key site for determining receptor tropism. Furthermore, the observation that ephrin-B1^{LW} bound NiV G with a similar affinity to wild-type ephrin-B2 is suggestive that interactions with the GH loop, specifically the Y-M/F-W motif, are the principal determinants of differential receptor affinity in this instance, and therefore specificity (Negrete et al., 2006).

Analysis of the NiV G-ephrin-B2/-B3 complex crystal structures reveals that the F-W motif of both ephrins is engaged by a hydrophobic surface that is primarily comprised of F458^{NiV} and W504^{NiV} (**Fig 3.7.6ii A).** Furthermore, R402^{NiV} is positioned above the tryptophan (W122$^{ephrin\text{-}B2}$/W125$^{ephrin\text{-}B3}$) of the TW motif, which likely forms Van der Waals (V.D.W) interaction with W122$^{ephrin\text{-}B2}$/W125$^{ephrin\text{-}B3}$ and is within hydrogen bonding distance of the main-chain of E94 within the EF loop of ephrin-B2. Interestingly, the EF loop of ephrin-B3 does not

come within hydrogen bonding distance of $R402^{NiV}$ and thus, along with $Y117^{ephrin-B3}$ may account for the reduced binding of affinity of NiV G with ephrin-B3, relative to -B2 (Negrete et al., 2006). Within CedPV G the equivalent residue to $R402^{NiV}$, $P422^{CedPV}$, would be unable to form a hydrogen bond with $E94^{ephrin-B2}$ and would likely form less extensive V.D.W interactions with W122/W125 of either ephrin-B2 or -B3, respectively. Therefore, if such interactions are not compensated in other regions of the molecule, that are less apparent, CedPV G likely mediates fewer contacts with ephrin-B2 and thus may bind it with lower affinity than NiV G and HeV G.

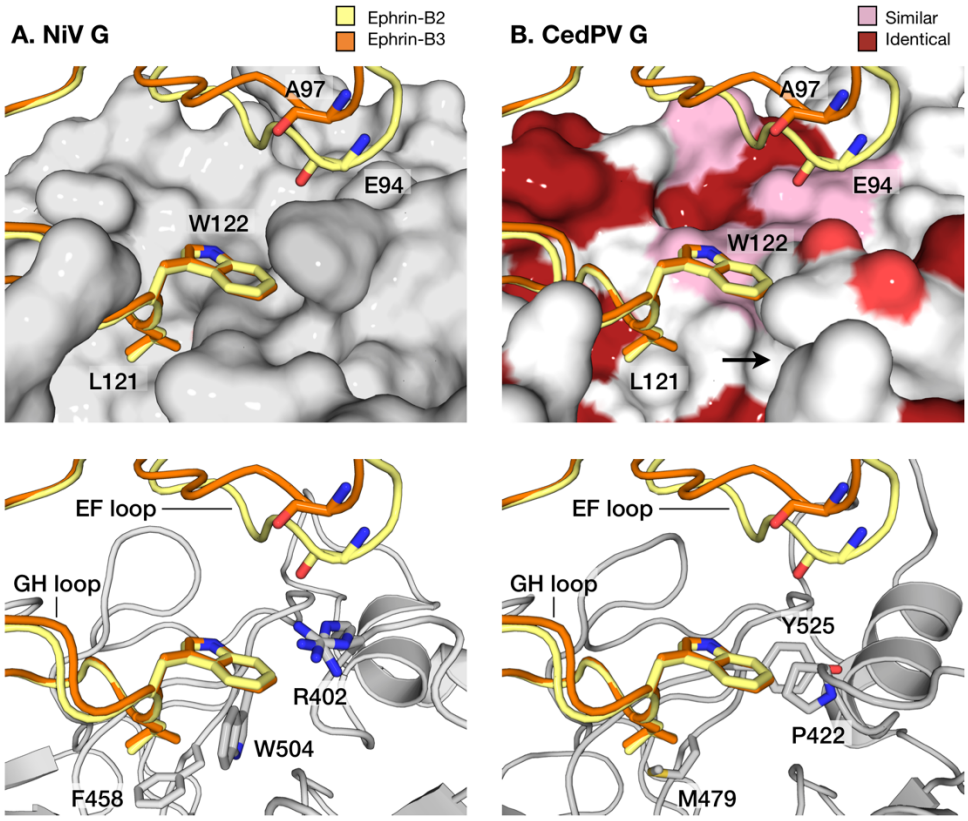


Figure 3.7.6ii: Differential recognition of the ephrin GH loop likely accounts for restricted B-class ephrin utilisation by henipaviruses. (A) The crystal structures of NiV G (grey) in complex with ephrin-B2 (yellow cartoon) and ephrin-B3 (orange cartoon) (PDB 2VSM and 3D12). NiV G is shown in surface (top) or cartoon (bottom) representation, displaying an identical view. L121 and W122 of the GH loop of ephrin-B2 and -B3 are shown as sticks and labelled. The main chain atoms of E94 and A97 within the EF loop of ephrin-B2 and -B3 (respectively) are shown as sticks. Ephrin residues are not labelled in the lower panel for clarity. Important NiV G residues discussed in the accompanying text are labelled in the lower panel. (B) The crystal structure of CedPV G with ephrin-B2 (yellow) and ephrin-B3 (orange) superimposed. The CedPV surface is coloured according to conservation with NiV G: red regions are identical, and pink are similar. The expanded cavity present in CedPV G is indicated with an arrow. (bottom) Salient features of the GH interacting region of CedPV G are displayed as in A.

The most striking feature of the putative GH recognition interface within CedPV G is the side chain orientation of $Y525^{CedPV}$. Whilst the equivalent residue in NiV G, $W504^{NiV}$, is positioned below $W122^{ephrin\text{-}B2}$, and amenable to the formation of extensive V.D.W interactions with both L and W within ephrin-B2 and -B3, $Y525^{CedPV}$ is orientated away from the LW motif (**Fig 3.7.6ii B**). The aromatic side of chain of $Y525^{CedPV}$ is positioned behind $P422^{CedPV}$ and is likely conformationally constrained by the surrounding environment. Thus, without conformational rearrangement, it is unlikely that $Y525^{CedPV}$ would be able to interdigitate the LW motif as does $W504^{NiV}$. As a consequence of the positioning of $Y525^{CedPV}$ behind $P422^{CedPV}$, the cavity below the LW motif within CedPV G is expanded relative to other henipaviral G proteins. It is therefore possible that expansion of this cavity within CedPV G relieves steric hinderance that would otherwise preclude the binding of the YM motif presented by ephrin-B1, thus permitting its utilisation. To this end, it seems likely that $Y121^{ephrin\text{-}B1}$ presents the greatest barrier to the recognition of ephrin-B1 in NiV G, HeV G, and GhV G, which all possess a conserved tryptophan at position $W504^{NiV}$ that adopts a similar conformation.

3.8 Discussion

Until recently, the genus *Henipavirus* comprised only the eponymous HeV and NiV, which constitute two of the most lethal known human pathogens (Marsh and Wang, 2012). Whilst serological and genetic surveys have identified numerous putative helipa-like viruses across a broad geographical range (Drexler et al., 2012), only GhV and CedPV have been completely sequenced, and in the latter case, successfully isolated (Marsh et al., 2012). Intriguingly, despite phylogenetic, geographical, and serological association with the prototypic henipaviruses, CedPV appears to be non-pathogenic in animal models (Marsh et al., 2012).

Consistent with structural and genomic phylogenetic relatedness to the prototypic henipaviruses, CedPV, like HeV and NiV, is capable of undergoing ephrin-B2-mediated cellular entry (Marsh et al., 2012, Laing et al., 2018) (**Fig 3.7.2**). Interestingly, whilst CedPV is incapable

of utilising ephrin-B3 as an alternate entry receptor, as NiV and HeV are, it is uniquely able to undergo ephrin-B1 mediated cellular entry. Structural characterisation of the receptor-binding module of CedPV G was performed to delineate the molecular determinants of the overlapping but distinct receptor tropism characteristics of henipaviruses.

Despite pronounced sequence divergence relative to the prototypic henipaviruses (31% identity to NiV G), the canonical six-bladed β-propeller architecture utilised by CedPV G is highly conserved (1.6 Å RMSD). In concert with overall structural similarity, elevated amino acid conservation within the putative ephrin-binding interface suggests a functional constraint is imposed upon the molecule by the requirement to maintain high-affinity ephrin binding. Indeed, superimposition-based docking of ephrin-B2 onto the putative receptor binding interface of CedPV revealed conserved features, indicative of a shared mode of receptor recognition, such a deep hydrophobic pocket into which $Fl17^{ephrin\text{-}B2}$ is likely inserted. The observation that GhV G, which is similarly divergent from the prototypic henipaviruses (28% identity to NiV G) as CedPV G, engages ephrin-B2 in a binding-mode that is strikingly similar to NiV and HeV (Lee et al., 2015), further supports a universal mode of ephrin-B2 recognition by henipaviruses.

Inspection of the crystal structure of CedPV G offers no obvious explanation as to its inability to bind ephrin-B3. Notably, the primary interacting region, the GH loop of ephrin-B3, is not substantially different from that of ephrin-B2. The crystal structure of GhV G, which is also incapable of binding ephrin-B3, in complex with ephrin-B2 revealed a substantially reduced molecular contact surface as compared with the NiV G- and HeV-ephrin-B2 complexes, principally due to the absence of a secondary interaction site (**Fig 3.7.4ii C**). Structural analysis revealed that CedPV G is also likely to lack this secondary interaction interface with a bound ephrin and thus, like GhV G, may exhibit a number of subtle structural incompatibilities, or reduced number of inter-molecular interactions, that preclude ephrin-B3 binding, likely as a consequence of an insufficiently high interaction affinity.

Through collaborative study with the laboratory of Prof. Benhur Lee (Icahn School of Medicine at Mount Sinai, USA), ephrin-B1 was identified as an alternate functional entry receptor for CedPV. Ephrin-B1 utilisation is a unique feature of CedPV that is determined by the G protein. Previous mutational studies of the ephrin GH loop have demonstrated that substitution of the Y-M motif in ephrin-B1 with the homologous L-W motif of ephrin-B2 and -B3, renders it capable of interacting with NiV G (Negrete et al., 2006), implicating this motif as a critical determinant of henipavirus G binding. It is therefore probable that the L-W/Y-M interacting region of CedPV dictates its ability to bind ephrin-B1. Indeed, analysis of the CedPV G crystal structure revealed an expanded cavity proximal to the likely position of the Y-M motif that, whilst sufficiently large to accommodate a tyrosine side chain in CedPV G, is significantly reduced in NiV, HeV and GhV. Furthermore, it is plausible that only Y within the Y-M motif restricts ephrin-B1 utilisation, though functional studies would be required to test this hypothesis. Complementary mutational analyses of the CedPV G protein would likely be more complex, given the multitude of factors that contribute to the local protein conformation. However, generation of a mutant CedPV G containing a substitution of $Y525^{CedPV}$ for the bulkier tryptophan that is present within NiV, HeV, and GhV would be of interest for future experimentation.

These, and other data, suggest that CedPV is a virus with zoonotic potential, owing to its ability to utilise at least two human cell-surface proteins for functional cellular entry (Marsh et al., 2012, Laing et al., 2018). However, despite its shared receptor usage with, and genetic relatedness to, the extremely pathogenic HeV and NiV, CedPV is non-pathogenic in small animal models (Marsh et al., 2012). Thus, despite ephrin-B2 utilisation underpinning many hallmarks of severe henipaviral disease, such as broad cross-species zoonotic transmission and cell-type tropism, it alone is not sufficient to account for the extreme pathogenicity of HeV and NiV. It remains undetermined whether the additional ability of HeV and NiV to utilise ephrin-B3, and the neurotropic dissemination it enables (Bergemann et al., 1998, Negrete et al., 2006),

particularly within the brainstem, contributes to the severe pathology of infection therewith. Furthermore, while the pathobiological implications of ephrin-B1 tropism are similarly unresolved, the novel discovery of ephrin-B1 utilisation by henipaviruses ought to warrant investigation.

3.9 Conclusion

Here, we report the crystal structure of the attachment glycoprotein from the only confirmed non-pathogenic henipavirus. These structural data highlight the potential for highly a conserved mode of ephrin recognition despite substantial sequence divergence amongst henipavirus G proteins. Furthermore, the structure of CedPV G enables rationalisation of the unique receptor tropism characteristics of the virus that will inform future experimental study. Given CedPVs non-pathogenic phenotype, the ability to utilise ephrin-B2, and the associated cellular tropism its usage engenders, can not solely account for the extreme pathogenicity of NiV and HeV. It does, however, dictate zoonotic potential. These data evidence the existence of henipaviruses with an expanded repertoire of functional entry receptors and provide a blueprint for understanding the molecular determinants of differential ephrin tropism. To this end, it is anticipated that structural data, such as those presented herein, may contribute to the foundation of predictive tools for the assessment of the receptor tropism characteristics of novel henipaviruses, and therein provide an assessment of zoonotic and pathogenic potential.

Chapter 4:

Crystal structure of the Whitewater Arroyo virus GP1

Chapter 4: Crystal structure of the Whitewater Arroyo virus GP1

4.1 Introduction

This chapter presents the crystal structure of the attachment glycoprotein, termed GP1, from Whitewater Arroyo virus (WWAV), the first structure from a clade D arenavirus. WWAV was isolated from its natural reservoir species, the white-throated woodrat (*Neotoma albigula*), and subsequently demonstrated to be antigenically and genetically most closely related to Tamiami virus (TAMV) (Calisher et al., 1970, Fulhorst et al., 1996). Despite initial characterisation as a clade A arenaviruses (Bowen et al., 1997) based upon nucleoprotein (NP) sequence, WWAV and other clade D viruses more closely resemble viruses from clade B within the glycoprotein (GP) gene (Archer and Rico-Hesse, 2002). WWAV and other North American arenaviruses were therefore reclassified as a distinct phylogenetic lineage termed clade D, (formerly clade A/B or A/rec).

WWAV, like other mammarenaviruses, causes persistent infection in rodent reservoir species. The geographic distribution of rodent hosts dictates that of the resident viral species, thus confining known clade D arenaviruses to North America (Reignier et al., 2008). Zoonotic transmission of New World (NW) arenaviruses to humans is sporadic, and underpinned by the ability of the GP1 attachment subcomponent of the arenavirus glycoprotein (GPC) to recognise the human orthologue of transferrin receptor (TfR1) (Radoshitzky et al., 2007). Similarly, the rodent TfR1 orthologues of arenaviral host species serve as functional entry receptors within viral reservoirs (Abraham et al., 2009). It is the additional and coincidental ability of a virus to utilise hTfR1 as a functional cellular receptor that not only contributes to zoonosis but also appears to be an absolute predictor of viral pathogenicity (Flanagan et al., 2008, Choe et al., 2011).

A WWAV-like virus has been implicated in the deaths of three people in California, USA (Enserink, 2000), and evidence of WWAV or WWAV-like virus infection has been detected in

humans across the southern United States (Milazzo et al., 2011). Whilst WWAV (type strain AV9310135) can not use hTfR1 as a functional cellular receptor *in vitro*, the WWAV isolate termed AV96010151, is able to utilise hTfR1 (Zong et al., 2014), highlighting the existence of WWAV variants with the ability to bind hTfR1 and cause human disease. These observations suggest that WWAV, along with other clade D arenaviruses, has the potential to emerge as a human pathogen.

4.2 Summary of findings

The GP1 glycoprotein of the NW arenavirus WWAV is responsible for host-cell attachment, and therefore a principle determinant of species and cell-type tropism. The ability of a NW arenavirus to bind human transferrin receptor 1 (hTfR1), in addition to the TfR1 orthologue of their respective host species, is a key determinant of human pathogenicity (Radoshitzky et al., 2007, Abraham et al., 2009). A soluble, receptor-binding competent construct of WWAV GP1 was transiently expressed in mammalian cells, purified, deglycosylated, and crystallised. Crystals of WWAV GP1 were generated in two conditions, buffered to pH 7.5 and 5.6, and X-ray diffraction data were collected to 2.4 and 2.0 Å resolution, respectively. As phase determination by molecular replacement was unsuccessful, long-wavelength SAD was used for experimental phase determination. The crystal structures reveal that WWAV GP1 adopts the canonical mixed α/β fold observed in all arenavirus GP1s solved to date and exhibits finer structural features that bridge both the Old world (OW) and NW serogroups. Structural analyses reveal that outside of the core secondary structure elements, WWAV GP1 differs substantially from other New World GP1s, highlighting the structural variation of the arenavirus attachment protein. To this end, amino acid conservation analysis further demonstrates the potential for diverse receptor-recognition modes across TfR1-tropic arenaviruses, mediated by poorly conserved GP1 surfaces. These data provide a molecular blueprint for understanding the determinants of the differential

hTfR1 usage in WWAV strains as well as a foundation upon which to assess the potential risks posed by other arenaviruses.

4.3 Expression, purification, and crystallisation of WWAV GP1

4.3.1 Construct design and expression testing

Constructs were derived from a codon-optimised synthetic GPC gene of the prototypic WWAV strain AV93101135 (NCBI reference sequence YP001911113.1). Construct boundaries were determined with the aid of a sequence alignment against the Machupo virus (MACV) GP1 (Bowden et al., 2009b), the only available arenavirus GP1 crystal structure at the time. Given the relatively limited amino acid similarity (47%) between the two proteins, constructs were finely screened around the boundaries defined in the crystallised MACV GP1 construct. Primers corresponding to three N-terminal 'start' residues (L67, A74, V81) and two C-terminal 'end' residues (S226, Q233) were synthesised, and PCR performed with the six unique 5' and 3' primer combinations. The resultant DNA fragments were cloned into pHLsec mammalian expression vector (Aricescu et al., 2006) (**Chapter 2: 2.1.1**).

Small-scale expression testing in HEK293T cells revealed that all constructs terminating at Q223 failed to secrete and were retained in the cellular fraction, indicative of misfolding (**Fig.4.3.1**). Constructs terminating at S226 were all detected in the culture supernatant and thus successfully folded and secreted; A74–S226 was present in the greatest quantity and therefore selected for large-scale protein production (**Fig.4.3.1**).

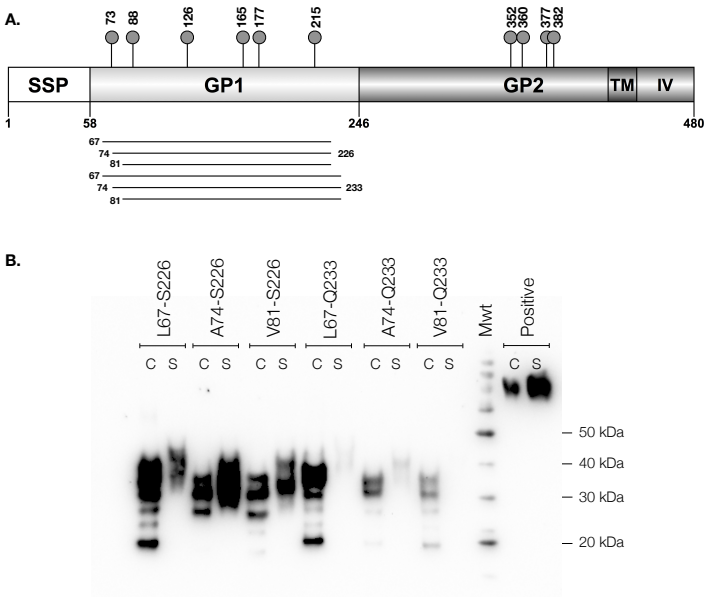


Figure 4.3.1: Construct design and small-scale expression testing of WWAV GP1. (A) Domain organisation and salient features of the WWAV glycoprotein complex, GPC. The three distinct polypeptides SSP (stable signal peptide), GP1 and GP2 are coloured from white to grey. The transmembrane (TM) and intra-virion (IV) regions within GP2 are labelled. Putative N-linked glycosylation sites (NXT/S where X≠P) are displayed as pins. Cloned GP1 constructs are illustrated as lines below GP1. (B) WWAV GP1 small-scale expression testing. Constructs illustrated in A were transiently transfected in HEK923T cells and subsequent protein expression analysed by western blot detection, under reducing conditions. Cell lysates (C) and supernatants (S) for each construct were run in pairs and compared to a protein known to express at high-levels as a positive control.

4.3.2 Large-scale expression and purification

The construct comprising residues A74–S226, hereafter synonymous with WWAV GP1, was used to transiently transfect HEK293T cells, in the presence of the α-mannosidase inhibitor kifunensine (Chang et al., 2007). Cell culture supernatants were clarified by centrifugation and filtration, and subsequently buffer exchanged into 10 mM Tris pH 8.0, 150 mM NaCl (TBS). A combined nickel affinity and size-exclusion chromatography (SEC) schedule yielded pure WWAV GP1 with an approximate yield of 2.4 mg/L of culture.

Enzymatic deglycosylation with endoglycosidase F_1 (endoF_1), which acts on the chitobiose core induced by kifunensine (Maley et al., 1989), yielded protein with an appreciable shift in its elution volume (~1 mL) from the SEC column (**Fig 4.3.2**). Reducing SDS-PAGE analysis of kifunensine-treated, glycosylated GP1 with endoF_1 treated material revealed five

distinct species ranging in molecular weight from approximately 20 kDa to 35 kDa (**Fig 4.3.2**). Within the accuracy permitted by this analysis, the observed mass range was consistent with the presence of at least four glycan sites (of five sequons), differentially occupied by high-mannose glycoforms ($Man_{6-9}GlcNAc_2$), assuming a predominant population of $Man_9GlcNAc_2$ with an approximate mass of 2 kDa. Deglycosylation with $endoF_1$ yielded two distinct molecular weight species, presumably differentiated by the presence of a single occupied glycan site. The lowest molecular weight species was observed in large excess, suggestive that $endoF_1$ is capable of removing the glycans contributing to the increased mass of the other species, and that no inherent properties of GP1 inhibit its removal. Therefore, the predominant species (>90% approx.) most likely exhibited homogeneous single GlcNAc glycoforms at four sites and was deemed suitable for crystallisation trials.

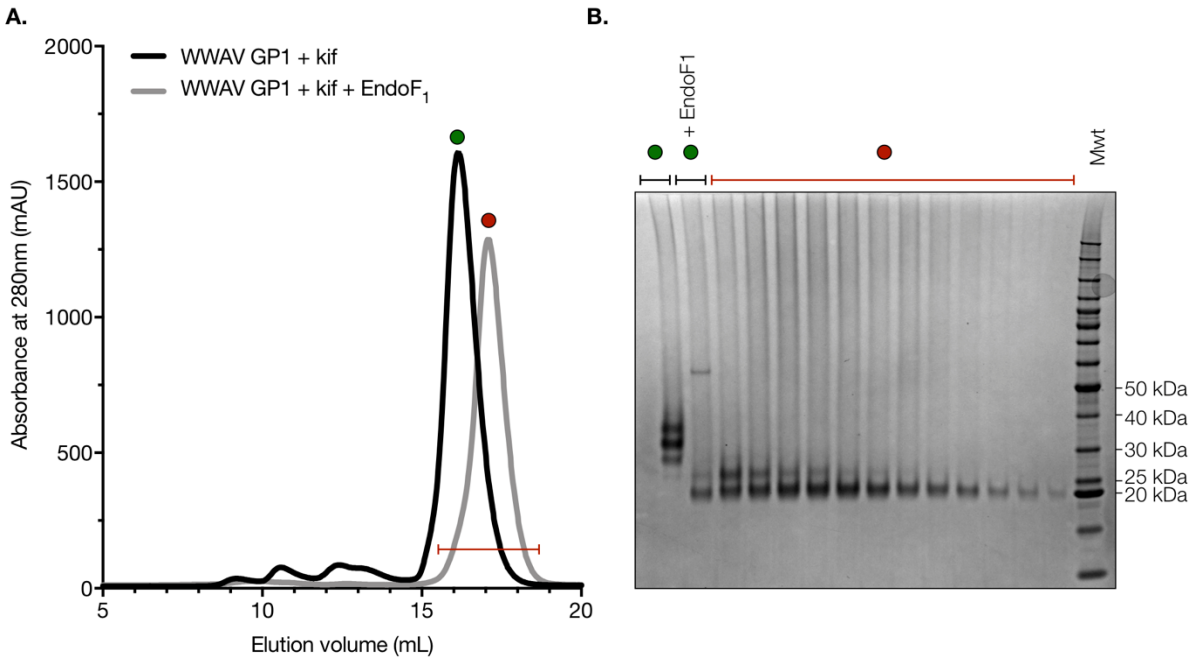


Figure 4.3.2: Purification and deglycosylation of WWAV GP1. (A) Superimposed size exclusion chromatograms of kifunensine treated WWAV GP1 pre- (black) and post- (grey) treatment with $endoF_1$, separated using a Superedex 200 Increase 10/300 GL column. Coloured circles and red bar denote species analysed by reducing SDS-PAGE, displayed in panel B. (B) SDS-PAGE analysis of WWAV GP1. Pooled fractions of purified kifunensine treated GP1 (green circle) were subject to $endoF_1$ deglycosylation for 16 hours at room temperature (green circle + $endoF_1$) prior to re-purification by SEC. Fractions spanning the deglycosylated GP1 peak (red bar) were analysed to assess the extent of deglycosylation. A molecular weight marker (Mwt) was included for size estimation.

4.3.3 Crystallisation

WWAV GP1 was concentrated to 10.1 mg/mL and subjected to the full range of crystallisation screens available at the Division of Structural Biology (Oxford, UK). Crystal trays were equilibrated and stored at 21°C, in the first instance. Crystals were initially observed after 48 hours, in a condition comprising 0.2 M potassium sodium tartrate, 2 M ammonium sulphate and 0.1 M tri-sodium citrate pH 5.6. Crystal morphology approximated elongated hexagonal bipyramids. Attempts to optimise these crystals with both dilution and additive screening, using the same batch of protein, were unsuccessful. This crystal form is hereafter referred to as pH 5.6.

A second crystal form was obtained in a condition comprising 11.0 mg/mL protein equilibrated against 1 M sodium acetate, 0.1 M HEPES-Na pH 7.5, and 0.05 M cadmium sulphate, which is hereafter termed the pH 7.5 form. Crystals appeared after 24 hours and continued to grow for 15 days prior to harvesting. Crystal arising from this condition formed elongated hexagonal prisms and were readily reproducible.

4.4 Expression and purification of transferrin receptor 1

4.4.1 Construct design and expression testing

Constructs were derived from a codon-optimised synthetic gene encoding the entire reading frame of transferrin receptor 1 (TfR1) from the rodent host of WWAV, *Neotoma albigula* (GenBank: AIA77429.1). Full-length *Neotoma albigula* ectodomain constructs (naTfR1^{ecto}) were designed based on sequence alignment with the crystallised construct of human TfR1 (hTfR1) (Lawrence et al., 1999). Truncated TfR1 constructs were designed in order to produce a variety of reagents with different properties that were anticipated to be advantageous to crystallisation. Truncated TfR1 constructs were designed to incorporate different combinations of structural domains; naTfR1^{apical} (S196–A381) lacked both the dimerisation-mediating helical and protease-like domains and comprised only the GP1-interacting apical domain of naTfR1 (Abraham et al.,

2010). The naTfR1$^{apical+protease}$ and TfR1$^{protease1+apical}$ constructs were designed to remove the helical domain, either alone or in conjunction with the second linear segment of the protease-like domain, respectively. Domain boundaries were structurally rationalised and devised to produce minimal structural domains with negligible vestiges of inter-domain linkers (**Fig 4.4.1**).

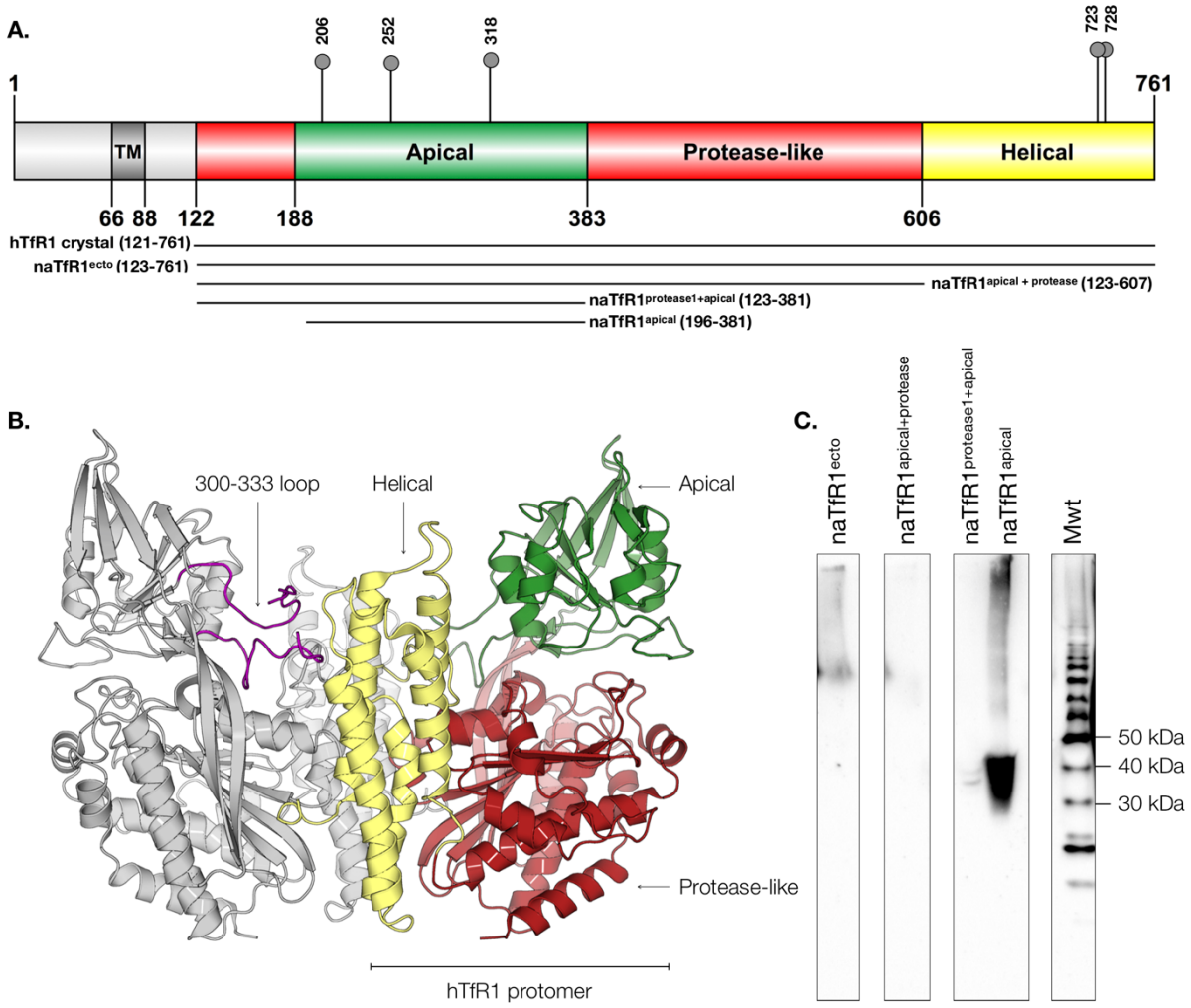


Figure 4.4.1: Construct design and small-scale expression testing of TfR1 from the WWAV host *Neotoma albigula*. (A) Domain organisation and salient features of the TfR1 protein. The protease-like domain is coloured red; the WWAV-interacting apical domain green; and the dimerisation-mediating helical domain yellow. Putative N-linked glycosylation sites are displayed as pins. Cloned constructs and the construct used to generate the hTfR1 crystal structure shown in B are illustrated below the diagram. (B) Crystal structure of hTfR1, coloured as in A. One protomer of the heterodimeric assembly is coloured in grey, and the large apical domain 300-333 loop is coloured purple therein, for clarity. (C) Small-scale expression testing of naTfR1 constructs. Constructs detailed in A. were transiently transfected in HEK293T cells, and the cell supernatants analysed by western blotting to detect secreted protein production.

The level of secreted protein of all constructs, except naTfR1^{apical}, was insufficient to merit up-scaling (**Fig 4.4.1**). Given the protein topology, failure of naTfR1$^{apical+protease}$ and naTfR1$^{proteasel+apical}$ secretion was unsurprising, however the extremely low levels of naTfR1^{ecto} was unanticipated, given that an equivalent hTfR1 construct expresses well under similar conditions (A. Zeltina, personal communication, 2017). The naTfR1^{apical} construct was secreted in large quantities and thus selected for large scale protein expression.

4.4.2 Large-scale expression, purification, and complexation

The naTfR1^{apical} construct was transiently expressed in HEK293T cells, in the presence of kifunensine. Cell culture supernatants were clarified by centrifugation and filtration, and subsequently buffer exchanged into TBS. A combined nickel affinity and SEC schedule produced pure deglycosylated naTfR1^{apical}, with an approximate yield of 0.8 mg/L of culture.

Kifunensine-treated naTfR1^{apical} appeared monodisperse in solution, eluting from SEC in a symmetrical peak centered around 15.8 mL (**Fig 4.4.2**). Enzymatic deglycosylation with endoF$_1$ was incomplete; naTfR1^{apical} eluted in an asymmetric peak that spanned the width of the non-deglycosylated protein peak and achieved maximal UV^{280nm} absorbance at 16.2 mL (**Fig 4.4.2**).

A small-scale complexation experiment was performed to assess whether purified TfR1^{apical} and WWAV GP1 would form a stable complex that could be isolated from its sub-components using SEC. In an attempt to saturate TfR1^{apical} binding sites, WWAV GP1 and naTfR1^{apical} were mixed in a 2:1 molar ratio and allowed to complex for two hours on ice. Two distinct species were observed when following SEC analysis of the complexation mixture, the larger eluted at approximately 15.5 mL and the smaller at 17 mL (**Fig 4.4.2**). Both species used for complexation derived from SEC fractions that each had elution volumes of 16.5 mL, thus the newly formed 15.5 mL peak, observed in the complexation run, likely represented authentic complex. Indeed, SDS-PAGE analysis revealed this peak to comprise both WWAV GP1 and

naTfR1^{apical} in an approximately one to one ratio, and the 17 mL peak primarily comprised WWAV GP1, in excess of naTfR1^{apical} (**Fig 4.4.2**).

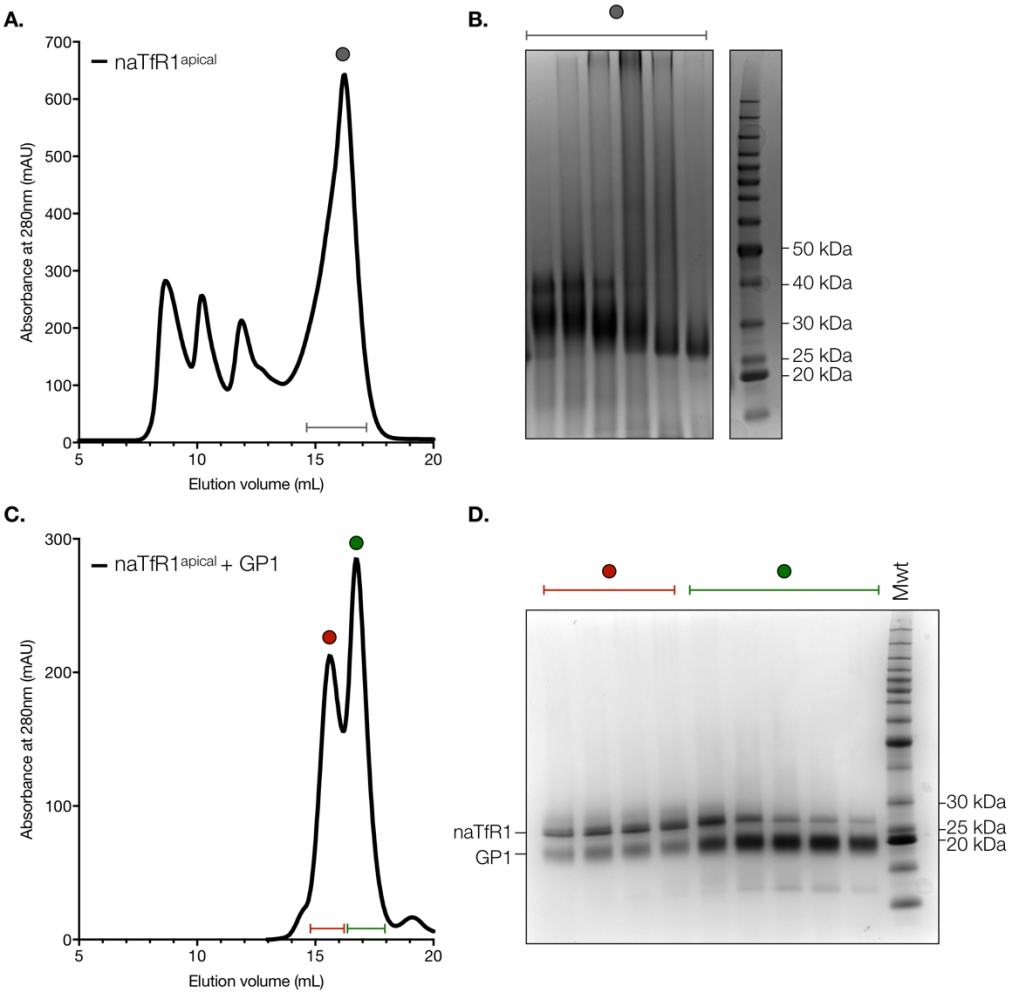


Figure 4.4.2: large-scale expression of naTfR1^{apical} and WWAV GP1 complexation. (A) SEC purification of endoF$_1$ treated naTfR1^{apical}, the peak of interest is illustrated with a grey circle. (B) SDS-PAGE analysis of fractions indicated with a grey bar in panel A. The asymmetry of the protein of interest peak in panel A resulted from incomplete deglycosylation of naTfR1^{apical}. Homogeneous deglycosylated naTfR1^{apical} from A and B was incubated with with WWAV GP1 before being analysed by SEC (C). Putative complex (red circle) and un-complexed component (green circle) peaks are labelled and analysed by SDS-PAGE in panel D. (D) Fractions analysed by SDS-PAGE are illustrated with red or green bars.

4.4.3 Crystallisation trials

Given the apparent absence of a stable complex—evidenced by a failure of a molar excess of WWAV GP1 to saturate naTfR1^{apical} (**Fig 4.4.2**)—purified naTfR1^{apical} and WWAV GP1 were mixed in a 1:1 molar ratio, concentrated to 7.3 mg/mL, and subjected to crystallisation screening. Of

the l2 screens initially tested, all exhibited widespread precipitation (>60% of conditions); GP1-TfR1 was therefore diluted to 4.3 mg/mL and re-screened. Despite the reduced concentration, screens consistently displayed above desirable levels precipitation, possibly due to inherent insolubility, rendering crystal nucleation and growth unlikely.

The naTfR1^{apical} construct was considered the most probable cause of the observed precipitation in the crystallisation trials. Indeed, the apical domain of TfR1 possesses a large internal loop (residues 300–330), containing a number of hydrophobic residues, which appears to be stabilised by the protease-like domain in the context of the full TfR1 ectodomain (**Fig 4.4.1**)(Lawrence et al., 1999). In the absence of such inter-subunit interactions, the 300–330 loop is likely to be conformationally labile. To address this, *in situ* proteolysis was performed with the aim of proteolytically cleaving flexible regions of the polypeptide chain, thus reducing the conformational heterogeneity of the sample and increasing the probability of successful crystal formation (Dong et al., 2007). To increase the likelihood of cleavage of the 300–330 loop, α-chymotrypsin was utilised, which preferentially cleaves the peptide backbone after bulky hydrophobic residues such as tryptophan, tyrosine and phenylalanine (Appel, 1986), of which there are four in this region. The equimolar naTfR1^{apical}-WWAV GP1 mixture (4.3 mg/mL final complex concentration) was supplemented with 0.002% (by mass) α-chymotrypsin and 200 mM non-detergent sulfobetain (NDSB-201, Sigma), to improve solubility (Vuillard et al., 1994), immediately prior to crystallisation screening. The resultant screen exhibited reduced precipitation, however, no crystal growth was observed (trays established 27/10/2017). Thus, future construct optimisation is warranted, specifically efforts to remove or modify the 300–330 loop, or optimise full-length ectodomain constructs.

4.5 WWAV GP1 structure determination

4.5.1 Crystal harvesting and X-ray diffraction data collection

WWAV GP1 crystals arising from both the pH 5.6 and pH 7.5 conditions were harvested and cryoprotected by transfer into their respective mother liquor supplemented with 25% glycerol. Crystals were subsequently flash-cooled by immersion in liquid nitrogen.

X-ray diffraction data for both crystal forms were collected at beamline I04 at Diamond Light Source and processed using xia2 (Winter, 2010) (**Table 4.4.1**). Crystals from the pH 7.5 crystal form were also exposed to long-wavelength X-rays (2.75520 Å) *in vacuo* at beamline I23 (Diamond, UK), to facilitate experimental phase determination. Data collection at I23 was kindly perfomed by Dr. Kamel El Omari.

Table 4.5.1: WWAV GP1 data collection statistics.

Data collection statistic	WWAV GP1 pH 7.5 Cd[a]	WWAV GP1 pH 7.5	WWAV GP1 pH 5.6
Beamline	DLS I23	DLS I04	DLS I04
Wavelength (Å)	2.75520	0.97950	0.97950
Space group	$P6_322$	$P6_322$	$P6_322$
Cell dimensions *a*, *b*, *c* (Å)	106.9, 106.9, 74.9	106.7, 106.7, 74.8	109.0, 109.0, 70.8
α, β, γ (°)	90,90,120	90,90,120	90,90,120
Resolution range (Å)	75-2.99 (3.04-2.99)[b]	29-2.43 (2.49-2.43)	55-2.08 (2.13-2.08)
R_{merge}	0.132 (>1)	0.068 (>1)	0.132 (>1)
I/σ(I)	25.2 (2.0)	22.8 (1.2)	18.0 (1.6)
$CC_{1/2}$	0.999 (0.723)	1.000 (0.638)	0.999 (0.588)
Completeness (%)	94.3 (85.0)	99.9 (100)	100 (100)
Multiplicity	28.3 (19.5)	18.8 (20.3)	19.2 (19.9)
Anomalous multiplicity	15.9 (4.8)		

[a] Cd denotes the cadmium-derived anomalous scattering data set used for phase determination.
[b] The value for the highest resolution shell is show in parentheses.
[c] $R_{merge} = \Sigma_{hkl} \Sigma_i|I(hkl;i) - \langle I(hkl)\rangle|/\Sigma_{hkl} \Sigma_i I(hkl;i)$, where I(hkl;i) is the intensity of an individual measurement and $\langle I(hkl)\rangle$ is the average intensity from multiple observations.

4.5.2 Crystal structure determination

For both WWAV GP1 data sets, phase determination by molecular replacement was not successful. Owing to the reproducibility of the pH 7.5 crystal form and the presence of cadmium in the crystallisation condition, diffraction data were collected at beamline I23 (Wagner et al., 2016), in order to maximise the detected anomalous signal. Phases were determined using the single-wavelength anomalous dispersion (SAD) method utilising the anomalous signal derived from uniformly bound cadmium atoms originating from the precipitant. Heavy atom sites and an initial trace model were generated with SHELXC/D/E using the HKL2map interface (Schneider and Sheldrick, 2002). The partial model produced by SHELXE was successfully used for the molecular replacement of the higher resolution pH 5.6 data (TFZ 23.6, LLG 343) (McCoy et al., 2007). Iterative rounds of model building and refinement were performed using COOT and PHENIX, respectively (Emsley and Cowtan, 2004, Adams et al., 2010).

Multiple attempts at independently generating a model for the 'native' pH 7.5 data, with appropriate R factors, were unsuccessful. To this end, the refined pH 5.6 model was used to phase the pH 7.5 data by molecular replacement. As such, both sets of 'native' data were assigned a consistent set of R_{free} reflections in order to mitigate the potential for model bias incurred by the use of models refined against other data. Model building and refinement was performed as above. Anomalous group refinement and secondary structure restraints were implemented in PHENIX (Adams et al., 2010). Occupancy refinement was also utilised to account for a cadmium atom bound at a crystallographic three-fold axis. Refinement statistics are presented in Table 4.5.2.

Table 4.5.1: WWAV GP1 structure refinement statistics.

Refinement statistic	WWAV GP1 pH 7.5	WWAV GP1 pH 5.6
Resolution (Å)	29-2.43	55-2.08
No. reflections	17,895	15,403
R_{work}[a]/R_{free}[b]	0.220/0.252	0.177/0.207
No. atoms		
Protein	1,160	1,170
Ligand/ion	5	0
Water	0	83
Glycan	42	42
B-factors		
Protein	105.2	48.1
Ligand/ion	108.1	n/a
Water	n/a	53.3
Glycan	140.1	75.0
Ramachandran(%)		
Favored	97.9	96.6
Allowed	2.08	3.42
Outlier	0	0
RMSD[c]		
Bond lengths (Å)	0.002	0.007
Bond angles (°)	0.55	0.86

[a] $R_{work} = \Sigma_{hkl}||F_{obs}| - k|F_{calc}||/\Sigma_{hkl}|F_{obs}|$
[b] R_{free} was calculated as for R_{work}, using a consistent set of reflections representing 5% of the total, removed prior to refinement.
[c] r.m.s.d.: root mean square deviation from ideal geometry.

4.6 WWAV GP1 structure analysis

4.6.1 Structure overview

Despite forming in distinct conditions and being of different, albeit hexagonal, macroscopic morphologies, both crystal forms belonged to the space group $P6_322$ each with one molecule in the crystallographic asymmetric unit. The structures obtained from both conditions were essentially identical (0.4 Å RMSD across 153 aligned Cα atoms). As such, reference of WWAV GP1 hereafter refers to the higher resolution pH 5.6 structure, unless otherwise specified.

Of the 153 GP1 residues in the crystallised construct, modelling of 148 (S75–G222) was permitted by the observed electron density, encompassing the majority of the globular receptor-binding domain. WWAV GP1 adopts the mixed α/β fold observed in other NW arenavirus GP1s (Bowden et al., 2009b, Zeltina et al., 2017) and OW GP1s, solved in complex with cognate GP2 molecules (Hastie et al., 2016, Hastie et al., 2017). The molecule broadly comprises two distinct halves: the concave β-sheet face, composing a 7-stranded anti-parallel β-sheet; and the convex, helical face, containing two α-helices, a 3_{10} helix and a number of connecting loops (**Fig 4.6.1**). Two additional short 3_{10} helical segments, η2 and η3, as defined by DSSP (Kabsch and Sander, 1983), are observed in the loops surrounding the periphery of the sheet face. The secondary structural elements are linked by numerous loops of varying lengths that primarily extend from the helical face and perimeter of the molecule. Loops are hereafter referenced with respect to the secondary structure elements that they connect.

Structural superimposition of WWAV GP1 onto Junín virus (JUNV) GP1 and MACV GP1 resulted in Cα RMSDs of 2.5 Å and 2.3 Å, respectively. Structural deviations map principally to the loops connecting secondary structure elements, which themselves are spatially conserved (**Fig 4.6.1**). The N-terminal region of the α-2 helix is extended in both JUNV and MACV. Whilst JUNV GP1 and MACV GP1 superimpose with an RMSD of 1.5 Å, the β1 and β2 strands of JUNV deviate away from β3 strand, unlike WWAV and MACV. The most notable structural difference is an 11 amino acid insertion following the β7 strand in MACV GP1, which is unique amongst known NW arenaviruses. This region, termed 'loop 10', engenders MACV a topologically distinct sheet face in which two unique 3_{10} helices and connecting loops sit atop the β1, 2, 3, and 7 strands (**Fig 4.6.1**).

Though classified as a NW arenavirus, superposition analysis reveals that WWAV GP1 exhibits similar overall levels (RMSD) of structural deviation with both NW and OW GP1s (**Table 4.6.1**). The RMSD values calculated from these superimpositions, however, arise from fewer aligned residues when comparing WWAV to OW, versus NW GP1s. Therefore, the overall

architecture of WWAV GP1 is more highly conserved with the NW viruses, as might be anticipated given shared receptor-tropism characteristics. Nevertheless, WWAV appears more structurally related to the OW GPIs than any other GP1 solved to date, possessing features that are conserved in the OW GPIs, such as the nature of the C185–C205 disulphide bond (**Fig 4.6.2**). A structural phylogenetic analysis of GP1 structures, detailed in chapter 5 (**Section 5.4.3**), provides a more accurate assessment of the structural relationship between GPIs across the entire family.

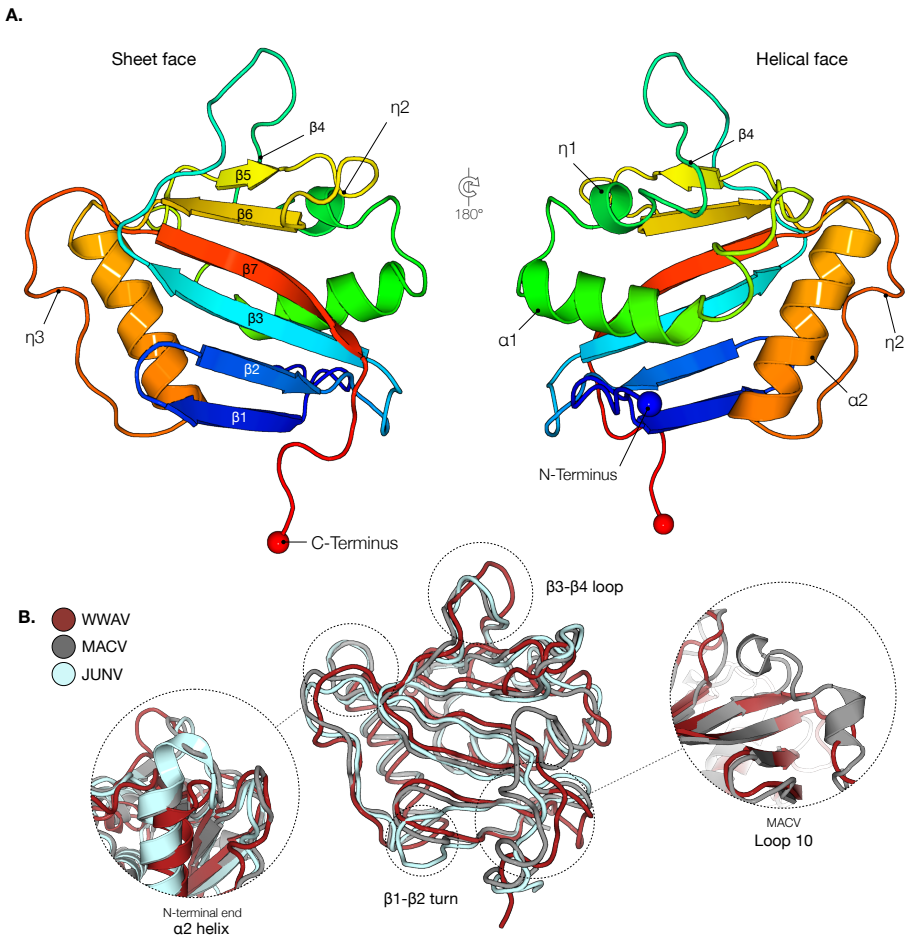


Figure 4.6.1: The crystal structure of WWAV GP1. (A) Cartoon representation of WWAV GP1 coloured blue to red, from N to C terminus, respectively. Secondary structural elements were defined with DSSP (Kabsch and Sander, 1983) are labelled in order of appearance in the primary sequence. Alpha helices are denoted α, 3_{10} helcies, η, and beta strands, β. The N and C termini are shown as blue and red spheres, respectively. GP1 is oriented to display the sheet face (left) and rotated 180 degrees, illustrating the helical face (right). (B) Structural superimposition of GP1s from WWAV (red), MACV (grey), and JUNV (cyan). Regions of notable structural deviation are highlighted with dashed circles, labelled and expanded where necessary. The structure of JUNV is removed in the 'loop 10' zoomed panel, for clarity. Overlaid structures are displayed as cartoon tubes for clarity.

Table 4.6.1: Arenavirus GP1 sequence and structure similarity.

	WWAV	JUNV	MACV	LCMV	LASV
WWAV		48.8% / 23.4% (12.7%)	47.4% / 25.6% (15.8%)	32.9% / 18.6% (29.9%)	36.7% / 19.9% (22.2%)
JUNV	2.5 Å (128)		57.6% / 44.4% (5.9%)	26.6% / 15.1% (53.5%)	34.2% / 22.1% (34.2%)
MACV	2.3 Å (119)	1.5 Å (139)		34.1% / 20.3% (33.7%)	36.0% / 18.6% (31.0%)
LCMV	2.2 Å (105)	2.5 Å (108)	2.1 Å (104)		66.5% / 48.8% (5.7%)
LASV	2.0 Å (106)	2.8 Å (103)	2.7 Å (103)	1.5 Å (129)	

Structure comparison

Sequence comparison

Table 4.6.1: Amino acid and structure conservation of arenaviruses. Pairwise amino acid sequence and structural comparison of NW and OW arenavirus GP1s. RMSD between aligned GP1 structures were calculated using secondary structure matching alignment in COOT (Emsley and Cowtan, 2004) and are shown in red boxes. The number of successfully aligned alpha carbons, from which the RMSD was calculated, is shown in parentheses. Pairwise alignment of GP1 protein sequences was performed with EMBOSS Needle (Needleman and Wunsch, 1970). Comparisons are presented in the format: similarity / identity, and the number of gaps is shown in parentheses. Sequences and structures used in the analysis: WWAV (NCBI ref: YP001911113.1), JUNV (PDB: 5NUZ, GenBank: ACS12870.1), MACV (PDB: 2WFO, GenBank: AAS77641.1), LCMV (PDB: 5INE, GenBank: BAM08941.1), LASV (PDB: 5VK2, NCBI ref: NP_694870.1).

4.6.2 Disulphide bonding within GP1

The structure of WWAV GP1 is stabilised by three disulphide bonds that surround the perimeter of the sheet face. Two disulphide bridges (C89–C224^{WWAV} and C136–C167 ^{WWAV}) are fully conserved with other NW arenaviruses, and each cysteine from the third (C185–C205^{WWAV}) is conserved with either NW or OW GPls (**Fig 4.6.2**).

C89–C224^{WWAV} is conserved across MACV, JUNV and the prototypic OW arenavirus, Lymphocytic Choriomeningitis virus (LCMV). C167 participates in a disulphide bond in all three viruses, however the counterpart, C136, is conserved only in the NW arenaviruses. Similarly, C205 is conserved in MACV, JUNV, and LCMV but C185 is only present in LCMV. Interestingly, while the C185–C205 disulphide bond stabilises the loop preceeding the β7 strand in MACV, in WWAV it bridges two helical segments: η3 and α2. The gross position and number of the disulphide bridges is similar in all aforementioned cases, except MACV that contains a fourth

within the C-terminal insertion, 'loop 10'. Spatial conservation, despite primary sequence position variability, highlights the importance of disulphide bonds in maintaining the core architecture of the GP1 fold.

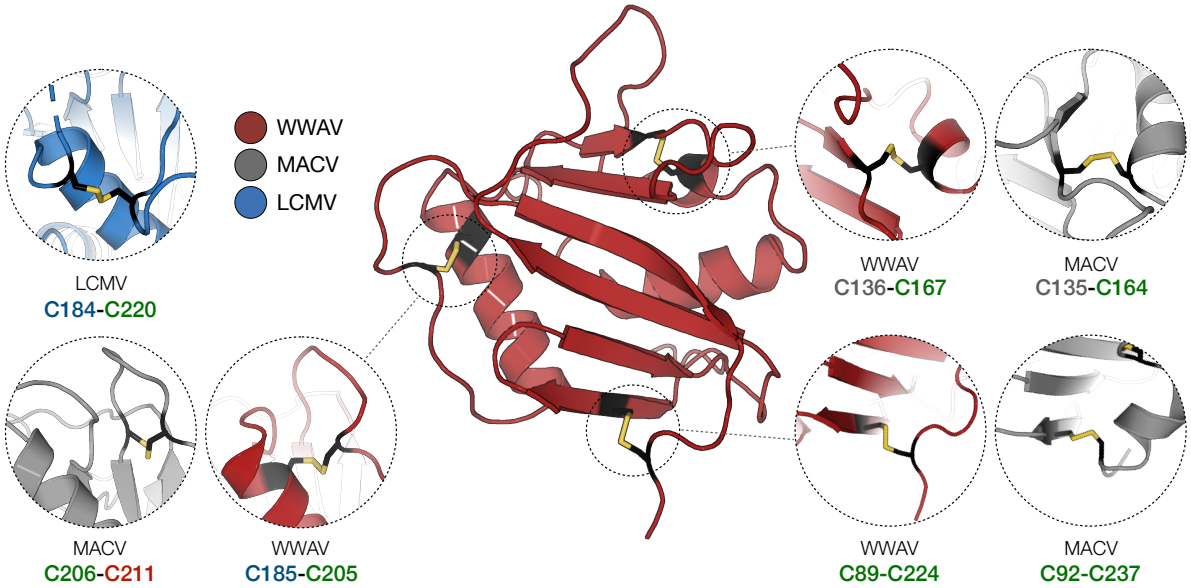


Table 4.6.2: Disulphide bonding within WWAV GP1. Conserved disulphide bonding stabilises the core architecture of the arenavirus GP1. WWAV GP1 is shown as a cartoon coloured red. Disulphide linkage are shown as sticks with carbon atoms coloured red and sulphur coloured yellow. Each linkage is shown as a zoom in panel with the equivalent disulphide bond from MACV and LCMV displayed in the equivalent orientation. For zoom panels, LCMV is coloured blue and MACV grey. Participating cysteine residues are numbered according to their virus of origin and coloured according to sequence conservation: green denotes residues conserved across MACV, LCMV and WWAV; grey denotes cysteines conserved between MACV and WWAV; blue denotes cysteine that is conserved in sequence position between WWAV and LCMV; and red denotes cysteines with no equivalent residue within the viruses analysed.

4.6.3 Comparison with the recently reported WWAV GP1

The structure of WWAV GP1, derived from a heterologous insect cell expression system, was recently reported to 2.25 Å resolution (Shimon et al., 2017), hereafter this structure will be referred to according to its protein data bank (PDB) accession code, 5NSJ. The authors modelled 142 residues (P79–C220) of the 157 (L72–M228) GP1 residues included in the construct, 138 of which align against our WWAV GP1 structure with an overall RMDS of 0.8 Å.

Structural differences map exclusively to peripheral loop regions, with the largest RMSDs observed at the C-terminal end of the α1 helix; in the loop immediately following the C-

terminal end of the η3 helix; and the greatest deviation, in the loop connecting helices η1 and α1 (**Fig 4.6.3**). Given that both GP1s derive from the type WWAV strain, thus exhibiting identical primary sequences, the observed structural differences reflect inherent variability of the secondary and tertiary structure (**Fig 4.6.3**). Within our WWAV GP1 structure, all of the aforementioned regions of above average RMSD, aside from the α1 helix C-terminal loop, exhibit above average *B*-factors, indicative of inherent flexibility. Notably, three samarium atoms co-localise at the η1-α1 loop of the 5NSJ structure. Accommodating the binding of these heavy atoms likely accounts for the large perturbation observed between the two structures in this region. Indeed, the conformation of the equivalent loop in our structure would not be compatible with samarium binding, owing to numerous steric clashes. The two structures of WWAV GP1 highlight the structural plasticity within loop regions, many of which likely require the quaternary architecture of the trimeric spike complex for stabilisation.

Shimon et al. observed electron density for two occupied N-linked glycosylation sites, N88 and N126, both of which are occupied in our GP1 structure (**Fig 4.6.3**). We observe an additional glycosylation site at N176, which, if occupied in the 5NSJ structure, would likely preclude crystal formation due to steric overlap between the sugar moiety and the η2 helix of the second GP1 molecule in the crystallographic asymmetric unit. Thus, the 5NSJ crystal form and glycosylation at N176 appear to be mutually exclusive, highlighting the role of crystallogenesis as a purification process, and limiting the possibility of having obtained this crystal form ourselves. Furthermore, these data indicate that glycosylation at N176 is not required for the correct folding of WWAV GP1.

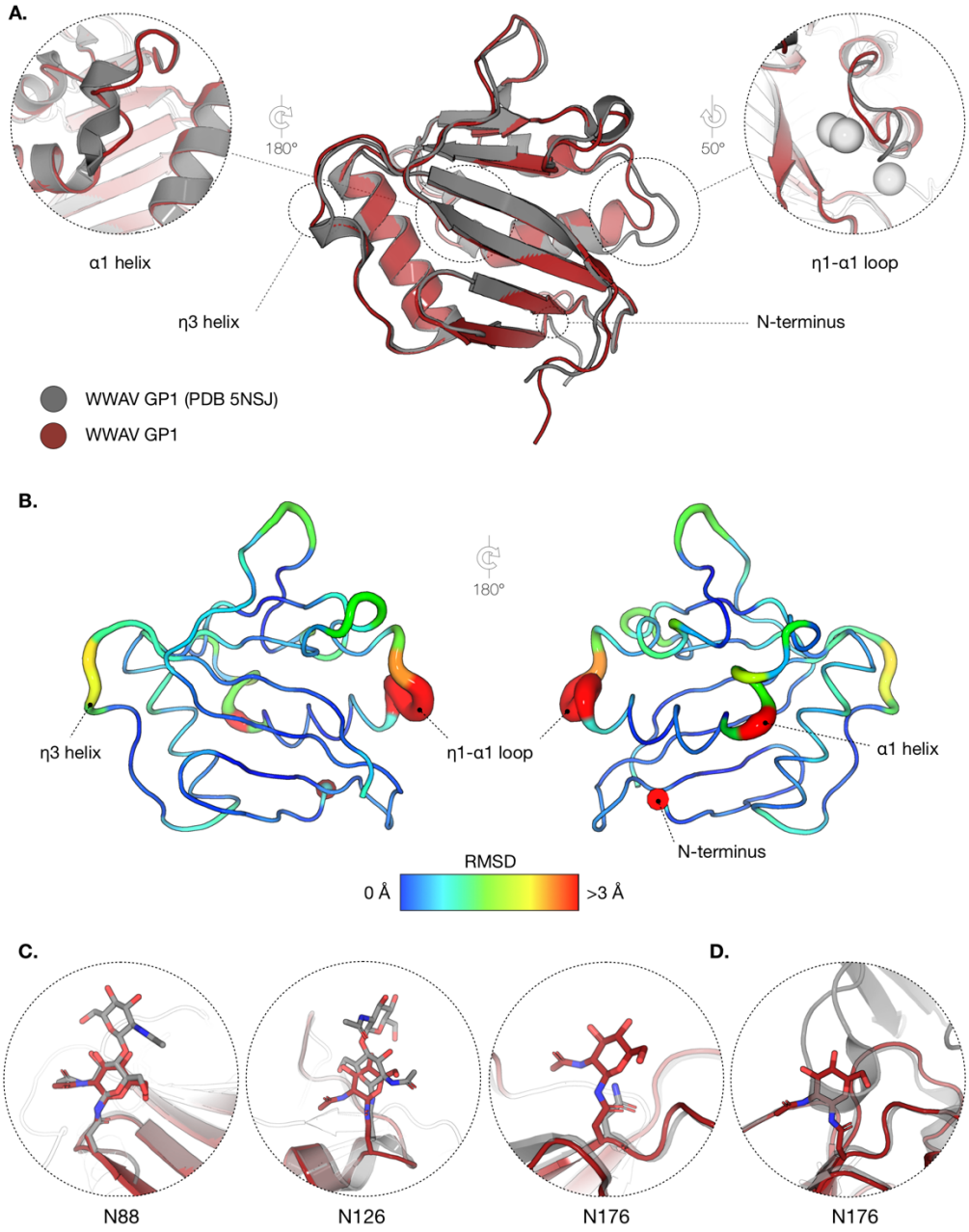


Figure 4.6.3: Comparison of WWAV GP1 crystal structures. Comparison of two independently reported WWAV GP1 structures reveals limited conformational variability that is primarily confined to loop regions. (A) Superimposition of the recently published structure of WWAV GP1 (grey), PDB accession code 5NSJ (Shimon et al., 2017), and the WWAV GP1 reported in this thesis (red). GP1 molecules are shown in cartoon representation with regions of greatest structural deviations highlighted. Rotations of the highlighted areas relative to the central panel are displayed. Bound samarium atoms in the 5NSJ are illustrated as white spheres. (B) Illustration of local RMSD between the two GP1 structures. GP1 is oriented as in A (left) and rotated 180 degrees anticlockwise about the y axis (right). RMSD is represented by cartoon width and colour, ramped from thin to thick and blue to red with increasing RMSD, respectively. (C) Comparison of differently occupied glycosylation sites. Asparagine side chains and their bound N-acetyl glucosamine moieties are shown as sticks and coloured by element: carbon is grey for 5NSJ, or dark red for our GP1, nitrogen is blue, and oxygen bright red. Glycosylation sites are labelled according to the sequence position of the modified asparagine residue. (D) Modelled steric clash between a potential glycan at N176 the η2 helix of a non-crystallographic symmetry related GP1 in 5NSJ. The symmetry related GP1 molecule is shown as partially transparent for clarity.

4.6.4 Glycosylation and a model for trimerisation of WWAV GP1

WWWAV GP1 encodes six N-linked glycosylation sequons (N-X-S/T where X≠P), of which five are encoded within the crystallised construct, at positions N88, N126, N165, N176 and N215. Unambiguous electron density corresponding to single asparagine-linked GlcNAc moieties was observed at sites N88, N126, and N176. The absence of visible electron density beyond single GlcNAc moieties at each of these sites is consistent with 'complete' endoF_1 deglycosylation (**Fig 4.6.4i**). The apparently non-glycosylated sites N165 and N215 exhibit no clear electron density in the vicinity of the asparagine residue, despite unambiguous density for the residue itself. It is unlikely that glycosylation at either of these sites would preclude crystallogenesis, given that neither is in close proximity to crystal packing interfaces. Glycans present at either site could access large solvent channels and would therefore likely be disordered, thus accounting for the lack of observable electron density. Whilst it is possible that these sites may not be glycosylated, the presence of five distinct species in the SDS-PAGE analysis of WWAV GP1 prior to and post-endoF_1 treatment evidences at least four occupied sites (**Fig 4.3.2**).

The sequons at N88, N176, and N215 are highly conserved and appear in arenaviruses belonging to all NW clades as well as OW arenaviruses. The strict conservation of N215 is suggestive of functional importance, and thus we consider it likely to be glycosylated. N126 is conserved in some NW clade B viruses such as the hemorrhagic-fever causing agents, Guanarito (GTOV), Sabia (SABV), and Chapare (CAPV), as well as in clade D arenaviruses. N165 is conserved only in JUNV and MACV. N215 is not conserved in the clade B viruses. It is plausible that the highly conserved glycosylation sites, such as N88, N176, and N215, may be crucial to the folding of the GP1 scaffold (Vigerust and Shepherd, 2007, Bonhomme et al., 2011), with less conserved sites arising from functionally diverse requirements, such as immune evasion and receptor-binding.

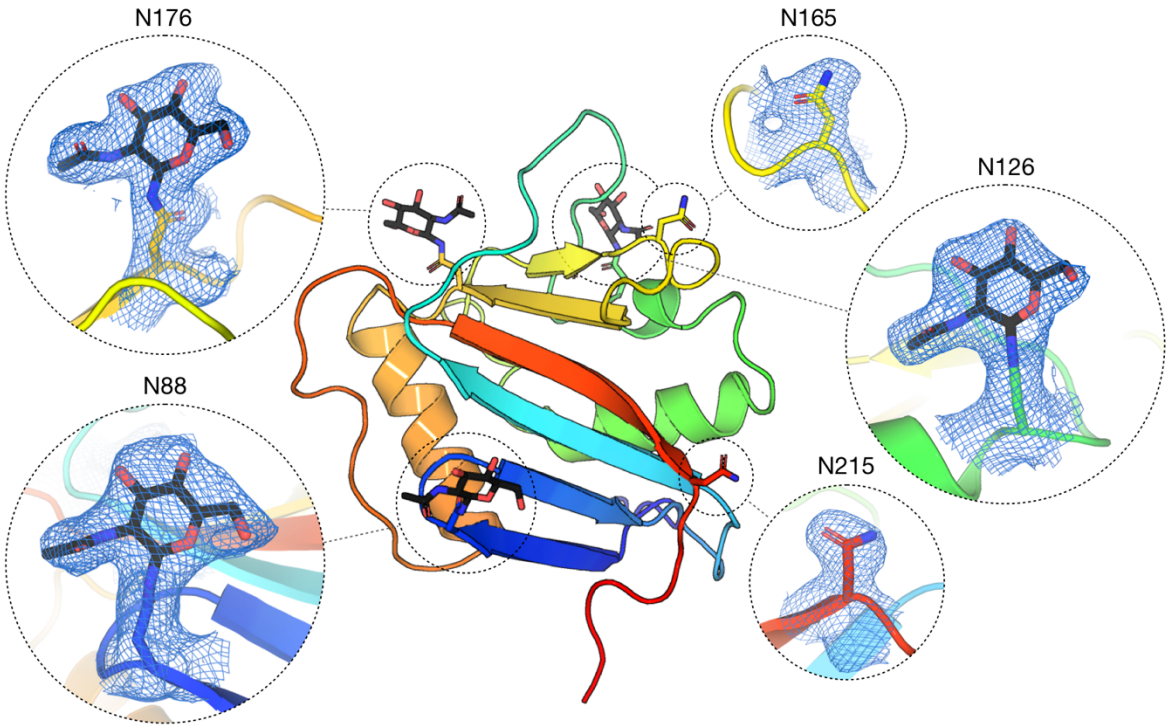


Figure 4.6.4i: Glycosylation of WWAV GP1. Asparagine side chains and their bound N-acetyl glucosamine moieties are shown as sticks on a cartoon representation of GP1. GP1 is coloured from blue to red from C to N terminus, respectively, with the relevant asparagine side chain carbon atoms coloured according to sequence position. Carbon atoms of the sugar moieties are coloured black, nitrogen in blue and oxygen in red. Expanded views of modelled glycans and asparagine residues from canonical sequons are indicated with dashed circles. Refined *2Fo-Fc* electron density maps for sites of interest are contoured to 1σ and displayed as blue mesh.

Of the five potential glycosylation sites observed within WWAV GP1, four extend from the sheet face of the molecule, primarily located around the periphery. The fifth, N126, is situated at the 'top' of the helical face behind the β3-β4 loop (**Fig 4.6.4i**). The position of all glycan sites is consistent with a model for GP1 trimerisation, as part of mature GPC spike complexes, based upon the crystal structure of LASV GPC (Hastie et al., 2017) (**Fig 4.6.4ii**). When superimposed onto the trimeric LASV GPC, WWAV GP1 exhibits no clear steric incompatibilities that would preclude a similar mode of self-association. Furthermore, the observed glycan sites do not exist at or near protein interfaces, rather they extend outward, away from both the apex and 'side' of the spike. In the context of the modelled trimer, the sheet face of WWAV GP1 is positioned outward, away from the trimeric axis with the strands of the central sheet running approximately perpendicular. Inter-protomer and GP1-GP2 interactions would

almost exclusively be mediated by the helical face and surrounding loops, leaving the receptor-binding sheet face sterically accessible for TfR1 recognition.

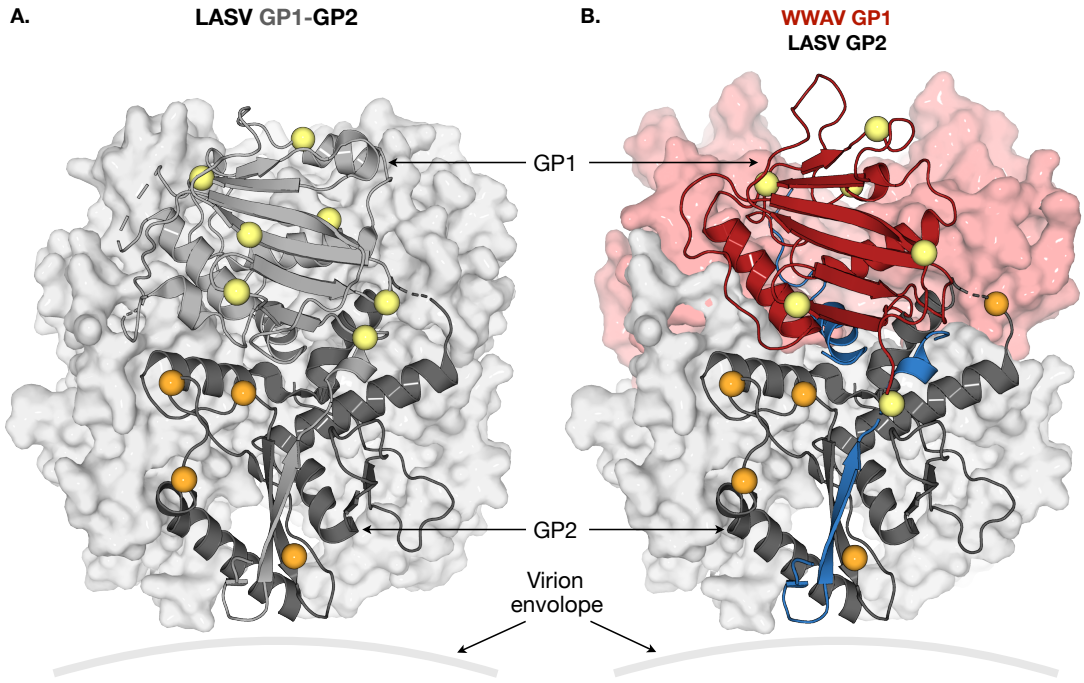


Figure 4.6.4ii: Glycosylation distribution supports a model for the trimeric assembly of WWAV GPC. (A) The trimeric ectodomain of LASV GPC (PDB: 5VK2). For clarity, a single protomer is shown in cartoon representation with GP1 and GP2 coloured light and dark grey, respectively. The remaining protomers are shown as white surfaces. The position of N-linked glycosylation sequons (N-X-S/T where X≠P) are shown as yellow and orange spheres, in GP1 and GP2, respectively. (B) WWAV GP1 was superimposed onto LASV GPC and is shown in red, the LASV GP1 to which it was aligned is removed for clarity. Regions of LASV GP1 for which the equivalent residues were truncated in the WWAV construct are shown in blue. The positions of N-linked glycosylation sequons in WWAV are displayed as in A. In both instances, the position of the virion envelope is indicated for orientation purposes.

Analysis of epitopes within GPC has revealed sites of immunological vulnerability, and conversely, protection, on the arenavirus spike complex (Robinson et al., 2016, Zeltina et al., 2017). To this end, of 113 patient-derived monoclonal antibodies against LASV GPC, 75% targeted LASV GPC at either GP2 or mixed GP1-GP2 epitopes, predominantly in membrane proximal regions of the spike. Furthermore, 13 of the 16 of those that were neutralising *in vitro* required mixed GP1-GP2 epitopes (Robinson et al., 2016). As such, quaternary or GP2 epitopes appear to be immunodominant in LASV, which, in light of recent evidence of the presence of a

dense glycan 'shield' that occludes large areas of the glycoprotein spike, supports a model for glycan-mediated immune evasion in LASV (Sommerstein et al., 2015, Hastie et al., 2017, Watanabe et al., 2018). In contrast, a number of studies have reported anti-GP1 antibodies capable of neutralising JUNV utilising different, albeit overlapping, epitopes (Mahmutovic et al., 2015, Zeltina et al., 2017, Clark et al., 2018). These data highlight the sheet face of NW, but not OW, GP1s as a site of immunological vulnerability. WWAV, like other TfR1-tropic NW arenaviruses, exhibits fewer N-linked glycosylation sites within GP1 than OW viruses. It seems plausible that the evolutionary constraint placed upon NW GP1s to maintain a relatively large protein-binding interface, as opposed to the more localised glycan-receptor binding site of the OW viruses, would limit the antigenic variability permitted within GP1, particularly the inclusion of glycans that could represent a steric barrier to TfR1 binding.

4.6.5 The receptor binding surface of WWAV GP1

WWAV, like other known clade D and B arenaviruses, utilises the TfR1 orthologue of its respective host species, *Neotoma albigula*, to mediate host-cell entry (Zong et al., 2014). The ability to additionally utilise the human orthologue of TfR1 appears to delineate zoonotic, hemorrhagic fever-causing arenaviruses such as those in clade B (Radoshitzky et al., 2007). To date, the crystal structure of MACV GP1 in complex with the hTfR1 ectodomain provides the only structural data regarding TfR1 recognition by arenaviral GP1s (Abraham et al., 2010)(**Fig 4.6.5i**).

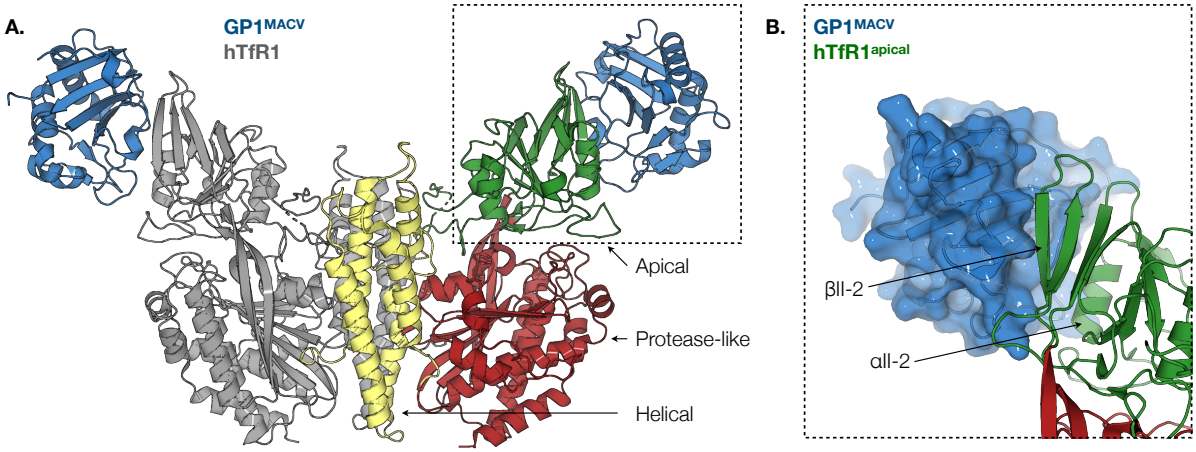


Figure 4.6.5i: Structure of the MACV GP1-hTfR1 complex. The crystal structure of the MACV GP1-hTfR1 ectodomain (PDB: 3KAS). MACV GP1 interacts exclusively with the apical domain of hTfR1, that has no known endogenous function. (A) Overall architecture of the MACV GP1-hTfR1 complex. Both proteins are shown as cartoons. MACV GP1 is coloured blue and a protomer of TfR1 grey. The other hTfR1 protomer is coloured according to the constituent structural domains: Apical green, protease-like red, and helical yellow. (B) A close-up view of the region indicated with the dashed box in A, rotated 70° about a vertical axis. Proteins are coloured as in A and the surface of MACV is shown for clarity. The interacting elements of hTfR1 are labelled.

When compared with other TfR1-tropic arenaviruses, MACV exhibits a C-terminal insertion of approximately 11 amino acids (**Fig 4.6.5ii**). This insertion, commonly referred to as 'loop 10', follows the structurally conserved β7 strand and comprises two loops and a short helical segment (η3) that sit atop the β1, 2, 3 and 7 strands of the sheet face. Thus, the surface topology of the sheet face of MACV is substantially different to that observed in WWAV and JUNV (**Fig 4.6.5ii**). Three residues within the inserted loop 10 are the sole constituents of 'motif 3', which encodes a number of residues that form a series of hydrophobic interactions and a hydrogen bond with the βII-2 strand of hTfR1 (Abraham et al., 2010). In the absence of any structurally equivalent residues in any known arenaviruses, it is unclear whether motif 3 is compensated by a subtly altered binding mode, or whether such interactions are simply not present. Indeed, the presence of the loop 10 insertion may define the MACV-hTfR1 interface as being non-canonical with respect to all other GP1-TfR1 interactions. However, the widespread utilisation of TfR1 as a host-cell receptor and, to a small extent, the conservation of molecular

properties within other GP1 motifs, make it unlikely that the overall location of binding is significantly different in any GP1-TfR1 pair.

Amino acid conservation within the GP1 of TfR1 tropic arenaviruses is low (**Table 4.6.1**), with two closely related hTfR1-utilising viruses, JUNV and MACV, exhibiting only 44% identity within GP1, and WWAV strain AV96010151 exhibiting 24% identity with MACV, despite common utilisation of hTfR1. A structural mapping approach was performed to assess the spatial relationship between conserved amino acids within the GP1s of WWAV and other arenaviruses, within the TfR1 binding site. Mapping the positions of amino acids that are absolutely conserved, or physiochemically similar, with respect to MACV, for the GP1s of JUNV, WWAV, and for comparison with a non-TfR1 tropic virus, LCMV, revealed surprising diversity across the receptor-binding surface of GP1s (**Fig 4.6.5iii**). In JUNV GP1, amino acid conservation appears to be reduced within the putative receptor binding site (44%), relative to GP1 as a whole (49%). Interestingly, WWAV, though not able to use hTfR1, has a more conserved receptor binding site (31%) than AV96010151 (25%) which is capable of binding hTfR1, despite being similarly divergent overall (26% and 23% identity to MACV, respectively). Though seemingly counter intuitive, and in contrast to observations of binding site conservation within henipaviruses (**Chapter 3**), such differences may arise from the selective pressure resulting from by co-evolution with the TfR1 orthologue of their respective host species (Bowen et al., 1997), which are more divergent than the ephrin ligands utilised by henipaviruses (**Table 4.6.6**).

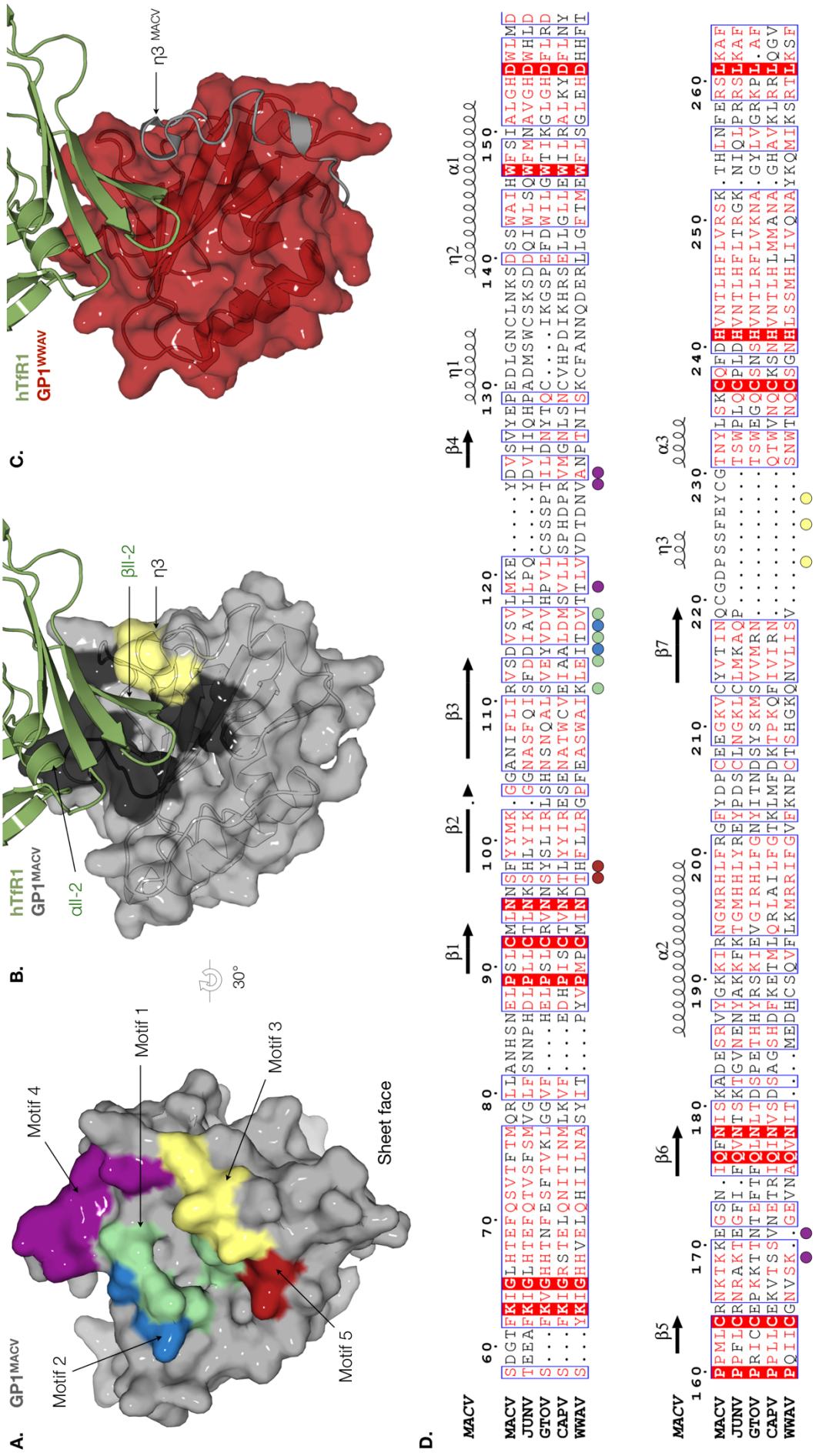

A.
GP1MACV
Motif 4
Motif 1
Motif 2
Motif 3
Motif 5
Sheet face
B.
hTfR1
GP1MACV
αII-2
βII-2
η3
30°
C.
hTfR1
GP1WWAV
η3 MACV
D.
MACV
JUNV
GTOV
CAPV
WWAV

Figure 4.6.5ii: The hTfR1 binding footprint of MACV GP1. (A) MACV GP1 is shown in surface representation (grey), oriented to show the entire sheet face with the β -1 strand at the bottom, with residues that contact the hTfR1 apical domain coloured according to the five motifs defined by (Abraham et al., 2010). (B) The GP1-interacting region of the hTfR1 apical domain is shown as a green cartoon. For clarity, the interaction footprint on GP1 is coloured black with the exception of motif 3, which is left yellow. (C) WWAV GP1 was superimposed upon the MACV GP1 component of the GP1-hTfR1 complex and displayed as a red surface. The C-terminal MACV insertion containing the η3 helix is shown as a grey cartoon, the rest of MACV GP1 is removed for clarity. (D) Sequence alignment of the GP1 glycoproteins from MACV, JUNV, GTOV, CAPV and WWAV. Absolutely conserved residues are highlighted in red, partially conserved residues are coloured red, and non-conserved residues are coloured black. Putative hTfR1 interacting motifs are denoted with coloured circles below the sequence, coloured according to A. Secondary structure elements of the MACV crystal structure are annotated above the sequence and labelled according to type: β strands are shown with arrows and labelled β1-7; α helices are illustrated as coils and labelled α1-3; 3_{10} helices are also labelled with a coil and designated η1-3. Sequence alignments were determined by MultAlin (Corpet, 1988) and plotted using ESPript (Gouet et al., 2003).

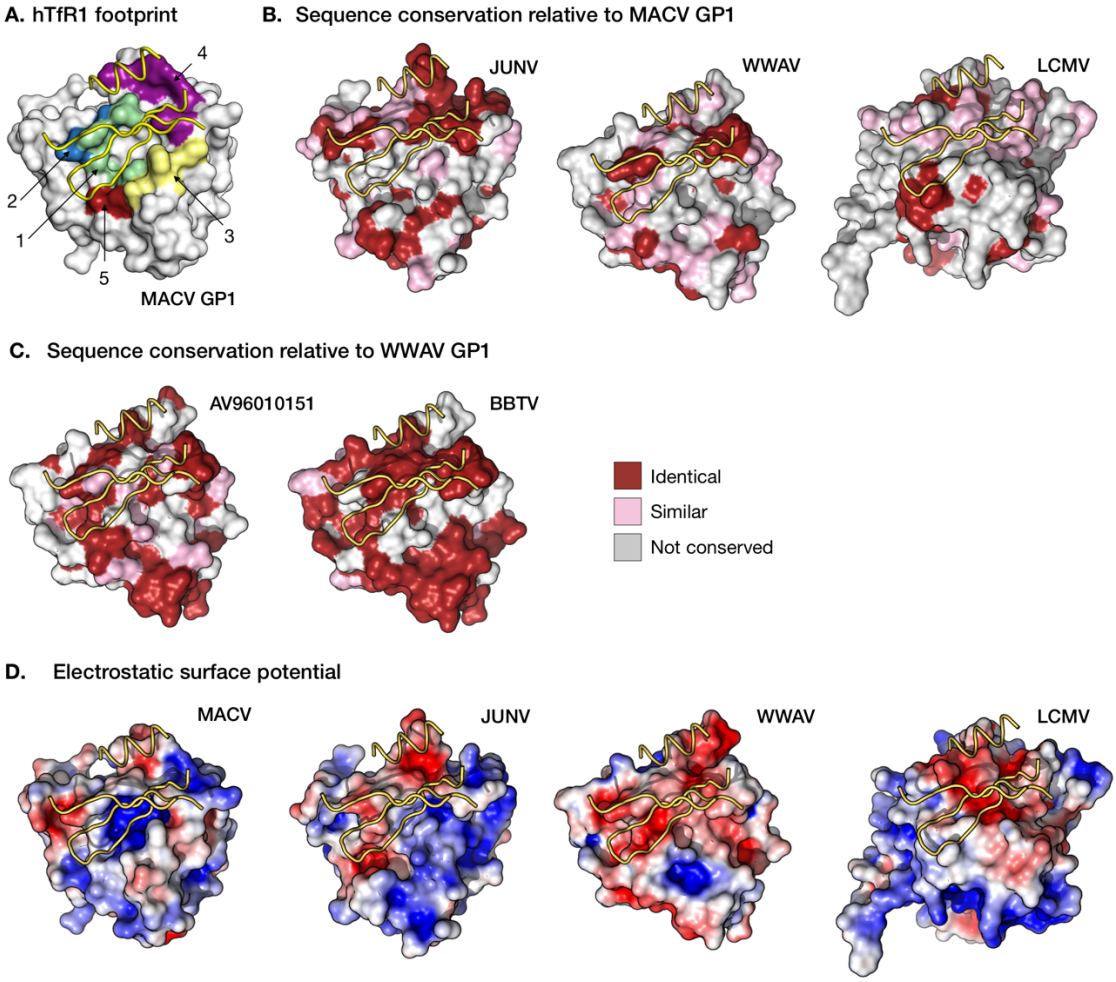
Sequence conservation within clade D arenavirus GP1s also tends to be reduced in the TfR1 binding site relative to the domain as a whole, despite an overall similarity of up to 70%. Of these, the GP1 from Big Brushy Tank virus (BBTV), which shares a common rodent reservoir with WWAV, exhibits elevated sequence conservation within the putative TfR1 binding site. The elevated sequence conservation within the putative receptor binding site of viruses that utilise the same receptor orthologues may therefore be indicative species-specific adaption to TfR1s.

Concomitant with the observed sequence divergence, the electrostatic properties of the putative receptor binding surface of WWAV GP1 has a predominantly negative electrochemical potential (**Fig 4.6.5iii**), contrasting observations of MACV and JUNV, which both exhibit mixed positive and negative character with an approximately similar distribution. Taken together, these data suggest that hTfR1 recognition may be achieved via numerous distinct interactions within a relatively divergent interface. However, WWAV GP1 presents putative receptor-binding surfaces that are well conserved within clade B viruses. These regions localise to motifs 1, 2, and 4 and may therefore represent key determinants of TfR1 recognition which are less tolerant to change.

When compared with WWAV GP1, LCMV GP1, which utilises α-dystroglycan as a functional entry receptor, is similarly divergent from MACV GP1 (19% identical across the entire protein) (**Fig 4.6.5iii**). Despite bearing similar levels of sequence conservation, the equivalent

'TfR1-binding' surface of LCMV GP1 is only 6% identical, indicative of its inability to bind TfR1. It is plausible that the asymmetric distribution of conserved surface residues within LCMV has arisen as a consequence of the functional constraint to maintain inter-protomer and GP1-GP2 interaction. The region of LCMV equivalent to the 'TfR1-binding' interface would not be restrained by the necessity to bind TfR1 and thus permitted to diversify, possibly under selective immune pressure or the requirement for altered receptor usage, accounting for its low local conservation relative to MACV. The stark reduction in similarity within the equivalent 'TfR1-binding' region of LCMV GP1 highlights that, despite the relatively diverse nature of the TfR1 binding site, NW GP1s nevertheless appear to demonstrate the constraint placed upon them by the requirement to utilise TfR1.

Of note is that receptor binding site identity was calculated inclusive of the three residues within motif 3 of MACV that are not present in any other arenavirus. The efficacy of such analyses that account for motif 3 when numerically assessing binding site conservations, therefore remains unclear. However, in the absence of structural data that may delineate a more representative mode of TfR1 recognition, it is assumed that TfR1 would dock with a similar overall binding mode in all TfR1-GP1 complexes. Indeed, though relatively diverse in primary sequence, the topological features, and localised regions of physiochemical similarity within WWAV GP1 and JUNV GP1 appear compatible with a highly similar mode of TfR1 binding (**Fig 4.6.5ii**). Nevertheless, the universal motifs 1, 2, 4 and 5 are still relatively highly diversified across NW arenaviruses, despite conserved receptor tropism characteristics. Thus, future studies are required to elucidate the varied GP1-TfR1 recognition interfaces.



E. Amino acid conservation relative to MACV GP1

	NW clade B	NW clade D		OW
Virus	**JUNV**	**WWAV**	**AV96010151**	**LCMV**
Seq identity	49%	26%	23%	19%
Binding site identity	44%	31%	25%	6%
hTfR1 binding	yes	no	yes	no

Figure 4.6.5iii: Conservation within the transferrin receptor (TfR1) binding surface of GP1. Analysis of the arenavirus GP1 surface reveals a poorly conserved putative TfR1 binding site. Human TfR1 (hTfR1) was aligned to all structures other than MACV, for which the complex structure is available. Regions of hTfR1 recognised by GP1 are shown as a backbone ribbon in yellow, with non-interacting regions removed for clarity. (A) The experimentally determined hTfR1 binding interface of MACV GP1 is coloured according to the interaction motifs defined by (Abraham et al., 2010). (B) Surface residue conservation relative to MACV GP1, red amino acids are absolutely conserved, and pink are similar. LCMV, which does not bind TfR1, is included as a negative control. (C) Surface conservation of clade D arenavirus GP1s, relative to WWAV, are mapped onto the structure of WWAV GP1 and coloured as in B. Big Brushy Tank virus (BBTV) and the hTfR1 utilising WWAV-like virus AV96010151 are shown. (D) Surface electrostatic potential was calculated using APBS (Baker et al., 2001) and residues are coloured blue and red to denote regions of positive and negative potential, respectively. (E) Table of amino acid conservation relative to MACV GP1. Pairwise alignment of GP1 protein sequences for percentage identity calculations was performed with EMBOSS Needle (Needleman and Wunsch, 1970). Binding site identities are expressed as a percentage of the total residues known to interact with hTfR1, as defined by (Abraham et al., 2010).

4.6.6 Sequence conservation of reservoir species' TfR1 orthologues

The observed variability within the receptor binding interfaces of NW GP1s (**Fig 4.6.5ii**) may be a consequence of arenavirus co-evolution with reservoir species (Bowen et al., 1997). While it is tempting to assume that sequence diversification within GP1 occurs as a consequence of the necessity to recognise similarly divergent TfR1 orthologues, sequence analysis of the TfR1s from arenaviral reservoir species suggests that TfR1 diversity alone is not sufficient to account for the observed sequence variation within GP1 (**Table 4.6.6**).

Table 4.6.6: Amino acid conservation amongst TfR1 proteins.

	Homo sapien	*Calomys callosus*	*Calomys musculinus*	*Neotoma albigula*	*Neotoma mexicana*	*Mus musculus*
		MACV	JUNV	WWAV	AV96010151	non-host
Homo sapien		77.0%[a]/ 88.1%[b]	77.0%/ 87.8%	77.9%/ 87.9%	78.6%/ 88.9%	76.8%/ 87.4%
Calomys callosus	73.7%/ 84.9%		98.0%/ 98.6%	87.6%/ 92.2%	87.6%/ 92.6%	83.2%/ 91.0%
Calomys musculinus	72.7%/ 83.4%	95.1%/ 96.1%		87.3%/ 92.1%	87.4%/ 92.5%	83.0%/ 91.1%
Neotoma albigula	76.0%/ 84.8%	82.3%/ 89.7%	81.4%/ 88.7%		97.9%/ 98.9%	83.9%/ 90.4%
Neotoma mexicana	76.5%/ 86.3%	82.3%/ 89.7%	81.9%/ 88.7%	97.5%/ 98.0%		83.6%/ 91.0%
Mus musculus	73.5%/ 84.3%	78.3%/ 89.2%	78.4%/ 89.7%	78.8%/ 85.7%	78.3%/ 85.7%	

Apical domain

Entire protein

Table 4.6.6: Pairwise amino acid sequence comparison of human and rodent reservoir TfR1 orthologues. *Mus musculus* is not a known host of any NW arenavirus and was included for illustrative purposes. Pairwise alignment of TfR1 protein sequences were performed with EMBOSS Needle (Needleman and Wunsch, 1970). Comparisons are presented in the format: similarity[a] / identity[b]. Comparisons are displayed for the entire protein coding region (upper right half, blue) and the GP1-interacting apical domains only (lower left half, red). The viral reservoir species for which the indicated rodents are hosts is indicated beneath the column titles.

Whilst MACV GP1 and WWAV GP1 are only 26% identical (**Table 4.6.1**), the TfR1 orthologues of their respective hosts, *C. callosus* and *N. albigula*, are 88% identical (82% within the GP1-interacting apical domain, **Table 4.6.6**). Furthermore, The GP1s from WWAV and BBTV, which

share the host *Neotoma albigula*, are only 66% identical despite common receptor usage. Interestingly, it has been demonstrated that in some instances, North American arenaviruses appear not to preferentially utilise the TfR1 orthologue of their respective host reservoir for cellular entry (Zong et al., 2014). The ability of arenavirus GP1s to promiscuously bind TfR1 receptors from different species highlights similarities between TfR1 orthologues, or flexible modes of TfR1 recognition. We suggest that the diversification of GP1s is not solely a consequence of adaptation to unique features of a given TfR1, which would otherwise represent a species barrier, but rather, likely arose in conjunction with the influence of additional pressures, such as those imposed by the hosts immune response. This analysis highlights the diversity with which NW GP1s are able to recognise relatively conserved receptor binding sites, underpinning the requirement for an enhanced molecular understanding of diverse modes with which TfR1s are engaged.

4.6.7 Molecular determinants of differential TfR1 tropism amongst WWAV strains

WWAV strain AV96010151 (hereafter called AV96) is a putative human pathogen that is able to utilise hTfR1, as well as nmTfR1 from its host species *Neotoma mexicana* (Milazzo et al., 2008, Zong et al., 2014). Given the similarities between WWAV and AV96, these viruses serve as a useful model for dissecting the molecular determinants of hTfR1 usage. The amino acid sequences of GP1 from WWAV and AV96 are 70% identical (82% similar), and their natural host species' TfR1s are 98% identical (**Table 4.6.6**), enabling a comparative analysis of the potential determinants that underpin the altered receptor tropism characteristics of these two viruses.

In order to study the GP1 from AV96, identical constructs to those tested for WWAV GP1 were made using a synthetic AV96 GPC gene template (GenBank accession EU123330.1) (**Section 4.3.1**). Despite the similarity of the two proteins, AV96 GP1 constructs expressed

poorly and upon upscaling, the most suitable candidate did not yield a suitable quantity of protein to permit crystallisation trials.

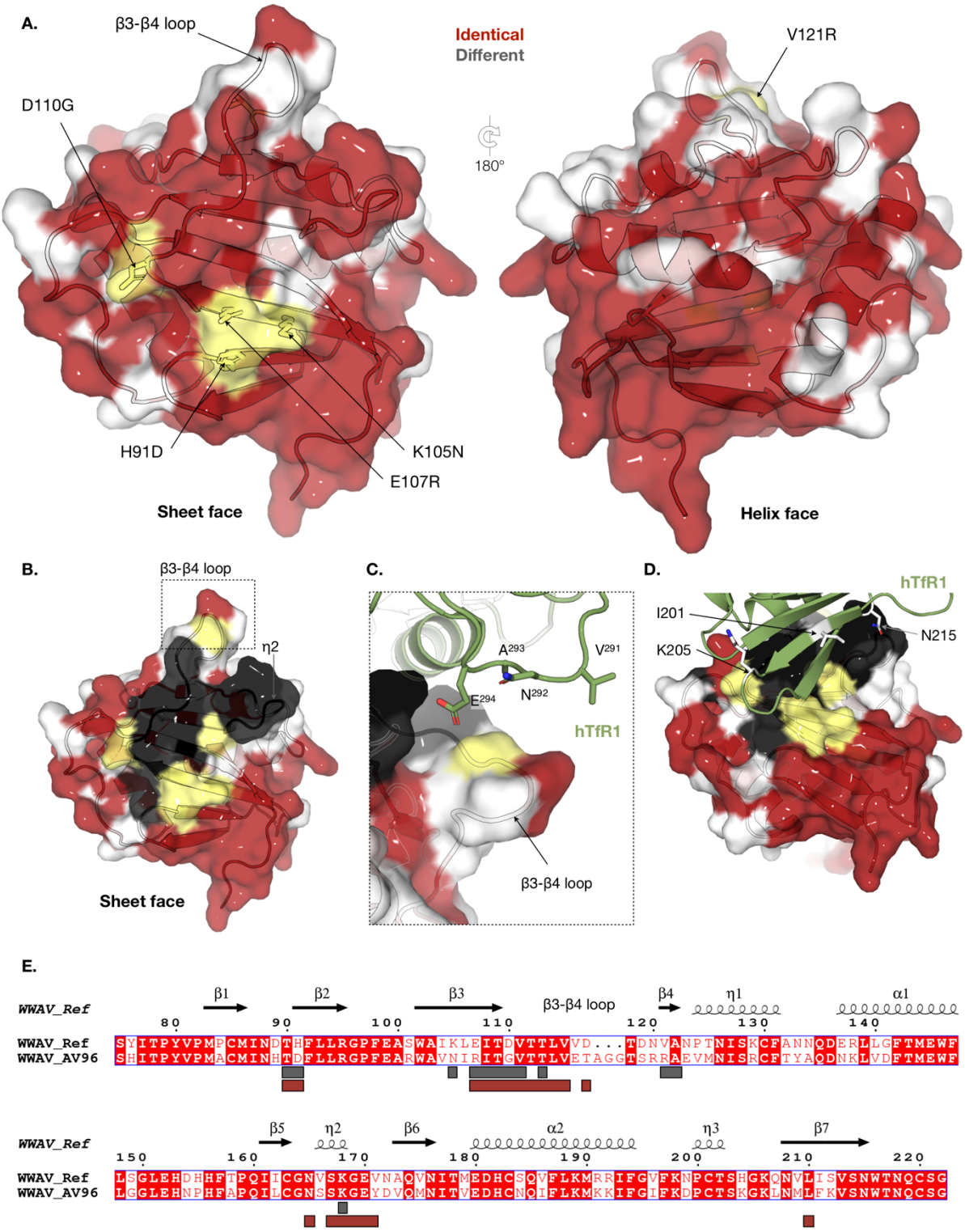


Figure 4.6.7: Mapping sequence conservation of WWAV strains with differential TfR1 tropism. Mapping sequence conservation of WWAV and AV96010151 (AV96) reveals potential hotspots for differential TfR1 tropism. (A) The crystal structure of WWAV GP1 is shown as a transparent surface. Sequence conservation between the type strain WWAV GP1 and the putative human pathogenic strain AV96 is mapped and displayed according to level of conservation: red regions are absolutely conserved, and white are not. Residue differences that align with known hTfR1 interacting residues of MACV GP1, based on sequence alignment, are coloured yellow. (B) WWAV GP1 is shown orientated and coloured as in panel A. Residues predicted to contact TfR1, calculated from a superimposition of WWAV GP1 on the MACV GP1-hTfR1 complex, using PDBe PISA (Krissinel and Henrick, 2007), are coloured black, and those of which that differ between the viral strains are coloured yellow. (C) Expanded view of the β3-β4 loop with the superposed hTfR1 shown as cartoon. (D) Location of amino acids that differ between the WWAV type strain and AV96 hosts, *Neotoma albigula* and *Neotoma mexicana* (respectively). Residues are mapped onto the structure of hTfR1, with the corresponding hTfR1 residues shown as white sticks and labelled according to hTfR1 numbering. (E) Sequence alignment of the GP1 from the WWAV type strain (WWAV_Ref) and strain AV96 (WWAV_AV96). Absolutely conserved residues are highlighted in red, partially conserved residues are coloured red, and non-conserved residues are coloured black. Putative hTfR1 interacting motifs, inferred from alignment with MACV GP1, are grey boxes below the sequence. Interacting residues determined by PDBe PISA are indicated by red boxes beneath the sequence. Secondary structure elements of the WWAV GP1 crystal structure are annotated above the sequence and labelled according to type: β strands are shown with arrows and labelled β1-7; α helices are illustrated as coils and labelled α1-3; 3_{10} helices are also labelled with a coil and designated η1-3. Sequence alignments were determined by MultAlin (Corpet, 1988) and plotted using ESPript (Gouet et al., 2003).

In the absence of a crystal structure of AV96 GP1, a structural mapping approach was employed to assess the spatial relationship between amino acid differences observed in both WWAV strains (**Fig 4.6.7**). Amino acid substitutions within WWAV and AV96 appear disperse in the primary sequence, and when mapped onto the structure of WWAV GP1, no contiguous regions of highly conserved or diverse amino acids are apparent. Indeed, even the putative receptor binding site (as predicted by sequence alignment with MACV GP1 **Fig 4.6.5i**) does not appear more highly conserved than the domain as a whole (66% identity in the RBS versus 70% overall), despite the identity of their cognate TfR1 receptors. Five WWAV residues (H91, K105, E107, D110 and V121) that align with the MACV GP1 TfR1 binding site are not conserved between the WWAV strains, and thus may represent sites with the potential to directly modulate the TfR1 interaction. Amino acids at positions 91 ^{WWAV}, 105 ^{WWAV}, 107 ^{WWAV} and 110 ^{WWAV} are all located within or very near the central β sheet, which is structurally conserved in WWAV and MACV (**Fig 4.6.1**), thus making it probable that these residues also participate in receptor binding in WWAV. Interestingly, when mapped onto the structure of WWAV GP1, V121 ^{WWAV} is

present on the 'back' of the β3-β4 loop, on the helix face of the molecule, and is therefore unlikely to mediate direct contact with TfR1. This analysis highlights the importance of integrating structural data with sequence-based alignments for determining the likely functional equivalence of amino acids within a protein.

In the absence of an atomic model of the WWAV GP1-TfR1 complex, an approximate TfR1 footprint was generated using structural superimposition of the MACV GP1-hTfR1 complex (**Fig 4.6.7**). Interestingly, the resultant hTfR1-WWAV GP1 interface exhibited no obvious steric clashes that would preclude TfR1 binding in the superposed binding mode, despite WWAV not being able to utilise hTfR1 (Zong et al., 2014). The superimposed complex was analysed with the online PDBe PISA server (Krissinel and Henrick, 2007), to ascertain residues in WWAV GP1 that were within distances that would permit non-covalent interaction with hTfR1 in the superimposed complex. Although imperfect, such superimposition analyses provide a model for predicting the molecular aspects of the GP1-TfR1 interaction. Indeed, these data support previous hypotheses that TfR1s are engaged by diverse GP1s with an overall similar binding mode, centered around a conserved tyrosine (Y^{211} in hTfR1) residue (Radoshitzky et al., 2008, Abraham et al., 2010, Shimon et al., 2017, Clark et al., 2018). Though the superimposition analysis is unlikely to recapitulate the fine specificity of the WWAV GP1-TfR1 interaction, it does support a conserved mode for TfR1 recognition, mediated by diverse amino acids.

Analysis of the hTfR1 footprint on WWAV GP1, derived from the superimposition analysis, suggests the involvement of residues not predicted by sequence-alignment alone (**Fig 4.6.7**). Indeed, the majority of residues that were identified by both approaches as being TfR1-interacting fall within the structurally conserved region comprising the first three β-strands. Following this region is the structurally variable β3-β4 loop, which in WWAV is extended by five amino acids when compared to both MACV and JUNV; when compared to WWAV, AV96 exhibits an additional insertion of three amino acids in the same region. The variability in the β3-β4 loop likely accounts for the failure of sequence alignment to accurately predict interacting

residues, as is the case for V121^{WWAV}, discussed previously. The β3-β4 loop is predicted by superimposition to form contacts with TfR1 that are not predicted by sequence alignment alone. The additional insertion in AV96 within the β3-β4 loop would likely result in an altered loop conformation relative to WWAV, that could potentially further modulate the TfR1 interface and contribute additional contacts to the TfR1 interaction.

The respective TfR1 orthologues of WWAV and AV96 host species do not exhibit amino acid variation within the region likely to contact the β3-β4 loop, though both differ from hTfR1 (**Fig 4.6.7**). As such, the observed variation in AV96 is unlikely to be a consequence of adaptation to specific determinants within nmTfR1, nonetheless such structural variation may permit more effective receptor usage within the natural host and even contribute to the ability to additionally utilise hTfR1.

To assess the impact that differences in TfR1 from *Neotoma mexicana* (AV96 host) and *Neotoma albigula* (WWAV host), might have on the receptor binding surface of AV96 and WWAV GP1, respectively, the location of amino acid differences were mapped onto TfR1 (**Fig 4.6.7d).** Only one difference (Y/N, in WWAV/AV96), at position 215 (hTfR1 numbering) is close enough to the docked WWAV GP1 to mediate any direct contacts. Interestingly, however, the nearest amino acid within GP1, G169^{WWAV}, and the i+1 and i-1 residues (E170^{WWAV} and K168 ^{WWAV}), are conserved across both WWAV strains. These data further indicate that the divergence of these two GPIs is unlikely to be derived from a requirement to recognise specific determinants within either naTfR1 or nmTfR1.

Consistent with the structure based-hypothesis that the differences within naTfR1 and nmTfR1 are unlikely to differentially preclude the binding of WWAV or AV96 GP1 (**Fig 4.6.7D**), AV96 GP1 has been shown to bind naTfR1 *in vitro* (Shimon et al., 2017). Furthermore, WWAV and AV96 are both able to enter HEK293T cells expressing either naTfR1 or nmTfR1 (Zong et al., 2014). Together, these data indicate that observed differences within the putative TfR1 binding interface of GP1 are not necessarily driven by specific adaptation to unique

determinants of host TfR1s, but nevertheless do reflect a constraint on maintaining TfR1 as a cellular entry receptor. Furthermore, structural analyses reveal that WWAV GP1 does not exhibit any features that represent an obvious barrier to hTfR1 recognition, despite its inability to do so. Comparison with AV96 suggests that subtle variation within the putative hTfR1 binding site, and the presence of additional hTfR1 contacts, contributed by the β3-β4 loop insertion in AV96, are crucial to the additional ability to utilise hTfR1. In light of such modest differences between WWAV and AV96 GP1 within the putative hTfR1 interface, it is likely that only minor changes in WWAV GP1 would enable it to utilise hTfR1 and emerge as a human pathogen.

Interestingly, preliminary studies performed in the laboratory of Prof. Stefan Kunz (Institut de microbiologie, CHUV, Switzerland) demonstrate that the clade D TAMV, is able to rapidly acquire the abilitiy to utilise hTfR1 for functional cellular entry. The study suggests that only a small number of mutational changes within GP1 are required to engender functional hTfR1 utilisation in hTfR1-incompetent viruses. To this end, these data evidence the zoonotic potential of clade D arenaviruses, such as WWAV, which likely only require minor modifications to their GP1 to enable hTfR1 utilisation.

4.7 Discussion

The crystal structure of the attachment glycoprotein from WWAV reveals that, despite low sequence conservation with pathogenic clade B arenaviruses, the clade D GP1 maintains a conserved overall architecture. However, concomitant with amino acid sequence diversity across NW areanaviruses, the GP1s of MACV, JUNV, and WWAV possess regions that are structurally dissimilar. Interestingly, a number of these diverse loop regions participate in the formation of the putative TfR1 binding site, underscoring the potential for varied modes of TfR1 recognition amongst arenaviruses.

Sequence variation amongst NW arenaviral GP1s is high, and not reflected in the magnitude of diversity observed within the TfR1 orthologues of the known host species. Though the notion that viruses exhibit greater and more rapidly acquired genetic variability than mammals is unsurprising (Emonet et al., 2009)(**Fig 1.1.1**), these observations highlight that TfR1 variation alone is unlikely to account for the diversity within the putative receptor binding sites of GP1s. It is likely that additional environmental influences and host factors, such as host immune pressure or alternate or co-receptor usage (Flanagan et al., 2008, Reignier et al., 2008), are drivers of change within NW GP1s, that in part account for the now widely accepted theory of arenavirus-host co-evolution (Gonzalez et al., 2007). Indeed, given the requirement to retain a complementary interface to the more highly conserved cognate GP2 protein and the antibody shielding effect afforded by N-linked glycans, the TfR1 binding site appears to be one of the few large and contiguous proteinaceous surfaces accessible to antibodies, and thus a site likely under increased selective pressure. Together, these influences have likely driven the diversification of GP1, resulting in varied modes of TfR1 recognition.

Consistent with observations that TfR1 orthologues from rodent hosts alone do not exhibit many features that likely represent a species barrier to a given virus, a number of studies have shown that TfR1-tropic viruses appear to promiscuously utilise the TfR1 from numerous rodent species for cellular entry, sometimes even more effectively than the virus for which it is the natural host (Zong et al., 2014, Shimon et al., 2017). These data, and our analysis of the diverse receptor-binding surfaces of NW GP1s, suggest that the collective contributions of an increased number of non-covalent interactions between GP1 and hTfR1 permit its utilisation as an entry receptor. Indeed, both MACV and AV96 GP1 appear to exhibit higher affinity for naTfR1 than its reservoir species, WWAV (Zong et al., 2014, Shimon et al., 2017). Furthermore, both MACV GP1 and AV96 GP1 bind hTfR1, but with an affinity orders of magnitude lower than naTfR1 (Shimon et al., 2017). Therefore, it is likely that WWAV, which possesses a lower affinity for even its host TfR1 orthologue (naTfR1), than MACV or AV96, does not exhibit sufficient

affinity for hTfR1 to permit utilisation as a functional entry receptor. However, given the relatively limited number of receptor-binding site mutations in hTfR1-utilising variants of WWAV (AV96), it is likely only small changes in GP1 would allow functional hTfR1 utilisation, resulting in a virus with the potential to emerge as a zoonotic pathogen.

The established role of hTfR1 as a critical determinant of pathogenic potential in humans (Flanagan et al., 2008) makes understanding the molecular interactions that determine GP1-hTfR1 recognition of critical importance. Given the existence of clade D arenaviruses with the ability to utilise hTfR1, and others that likely require only minor mutational changes to obtain the ability to use hTfR1, this group of viruses represents a tangible risk to human health. It is anticipated that an increased understanding of the molecular determinants of TfR1 tropism, provided by structural data like those presented herein, may provide a blueprint for defining a 'molecular code' that could be utilised to assess the threat posed by any given NW arenavirus, based upon the GP1 sequence alone.

4.8 Conclusion

Here, we report the first crystal structure of the GP1 from a non-clade B NW arenavirus. The structural insights obtained highlight the structural variability within the canonical arenavirus GP1 fold and provide a basis for understanding the determinants of the differential receptor tropism characteristics observed within clade D arenaviruses. Structural data, such as those presented herein, pave the way for predictive tools that may be used to assess the pathogenic potential of newly sequenced arenaviruses, without the need for time- and cost- intensive virus isolation or experimentation. Such data enhance our ability to determine the risk posed to human health by the ever-increasing number of identified arenaviruses, many of which have the potential to rapidly adapt to human hosts and therefore emerge as pathogens.

Chapter 5:

Discrete structural and functional states of arenavirus glycoproteins

Chapter 5:

Discrete structural and functional states of arenavirus glycoproteins

5.1 Preface

This chapter addresses discrete structural and functional states adopted by the arenavirus glycoprotein complex (GPC) and its sub-components. Structural transitions within mature trimeric GPC complexes, their environmental triggers, and binding of the receptor lysosome associated membrane protein 1 (LAMP1) were investigated using Lassa virus-like particles (VLP) (Schlie et al., 2010). Cryo-electron microscopy and VLP production was performed exclusively by Dr. Sai Li as previously described (Li et al., 2016). Data presented in this chapter are also included in the publication 'Acidic pH-induced conformations and LAMP1 binding of the lassa virus glycoprotein spike' (Li et al., 2016). Glycan composition analyses were carried out in collaboration with Mr. Yasunori Watanabe (Crispin Laboratory, University of Southampton).

The structures of isolated GP1 sub-components of the arenavirus GPC complex, from Old World (OW) and New World (NW) arenaviruses, are utilised to define the discrete conformational states adopted by arenaviral GP1s. Conformational plasticity is discussed with reference to the structures of the GP1s from an Asiatic rodent-borne OW arenavirus, Loei River (LORV) and Whitewater Arroyo virus (WWAV), first presented in Chapter 4. The crystal structure of LORV GP1 was solved by Weng Ng, a fellow DPhil student, and therefore expansive methodological details are not provided, though they largely adhere to those previously described for WWAV GP1. LORV GP1 utilised in circular dichroism spectroscopy experiments was expressed and purified by myself, adhering to the protocol detailed for WWAV GP1 (**Chapter 4**). Data and analyses presented herein are also presented in a published manuscript entitled 'Structure-Based Classification Defines the Discrete Conformational Classes Adopted by the Arenaviral GP1' (Pryce et al., 2018).

5.2 Summary of findings

Biochemical and cryo-electron microscopy analyses indicate that the OW Lassa virus (LASV) GPC complex undergoes pH-induced conformational changes that appear to accommodate the acidic pH-dependent binding of the intracellular receptor, LAMP1. Intriguingly, we demonstrate that sialylation of LAMP1 appears dispensable for binding to LASV GPC, contrary to initial reports (Jae et al., 2014, Israeli et al., 2017).

The crystal structures of the GP1 sub-component of the arenavirus GPC, from WWAV and LORV, solved at both acidic and near-neutral pH, permit the first direct structural analysis of the effect of pH on isolated GP1s. X-ray crystal structures and solution-state circular dichroism measurements reveal that pH does not alter the conformational state occupied by isolated GP1 proteins. Furthermore, structural phylogenetic analysis reveals WWAV GP1 to be equidistant from known OW and NW GP1s, highlighting the architectural variation observed within the GP1 scaffold, particularly amongst transferring receptor-1 (TfR1) tropic viruses. Using the same structural phylogenetic approach, we define discrete structural and functional states adopted by arenavirus GP1 glycoproteins. We suggest the nomenclature 'GP2-free' and 'GP2-associated' to distinguish the structural classes of OW arenavirus GP1 proteins, owing to their apparent dependence upon a cognate GP2 for maintaining the canonical architecture observed within trimeric glycoprotein complexes presented on mature virions prior to cellular-attachment (Li et al., 2016, Hastie et al., 2017). We demonstrate that NW GP1s adopt the GP2-associated state independent of quaternary associations and ambient pH, and thus are unlikely to exhibit the conformational duality of OW GP1s.

Taken together, these data allow us to propose a structure-based model of arenavirus GPC-mediated host cell entry. Intriguingly, the structural states defined herein appear not to be universal within the *Arenaviridae* family and may represent adaptations to independent cellular entry pathways and novel immune-evasion mechanisms.

5.3 Introduction

The tripartite arenavirus spike complex, GPC, is the sole protein antigen presented on the virion surface. GPC orchestrates host recognition and subsequent fusion of the viral and host-cell envelopes in a process that couples environmental pH sensation to both subtle and large-scale structural rearrangements (Li et al., 2016). GPC is expressed as a polyprotein precursor that is extensively post-translationally modified. The polyprotein precursor is proteolytically processed into the retained and myristoylated stable signal peptide (SSP), and the extensively glycosylated fusion (GP2) and attachment (GP1) glycoproteins (Burri et al., 2012). The resultant non-covalently associated SSP-GP1-GP2 protomers assemble into a non-covalently associated trimer of trimers that constitute the mature GPC.

Cryo-electron microscopy-derived tomographic reconstructions of *bona fide* LASV virions and VLPs, and crystallographic analyses of engineered, soluble trimeric GP1-GP2 ectodomains, have revealed the supramolecular organisation of the arenavirus spike complex (Li et al., 2016, Hastie et al., 2017) (**Fig 5.3i**). The globular extra-viral region of GPC comprises the entire GP1 and the majority of GP2. GP1 is positioned atop GP2 at the apex of the trimer, while GP2 traverses the viral membrane in an unusual tripodal arrangement. The structure of the functionally indispensable SSP and its interactions with the transmembrane domain of GP2 remain undetermined.

The arenaviral GP2 is classified as a class I fusion protein that, reflective of such a classification, undergoes pronounced fusogenic structural rearrangements that culminate in the formation of a post-fusion six-helix bundle (Igonet et al., 2011, Parsy et al., 2013, Hastie et al., 2016). Following primary receptor engagement at the cell surface and subsequent internalisation, the acidic environment of endocytic compartments is necessary and sufficient to trigger the fusogenic rearrangements of GP2 (Di Simone et al., 1994, Cohen-Dvashi et al., 2016). The class I fusion protein architype, the influenza HA protein, has been extensively studied (Stegmann, 2000) and many parallels drawn between it and the arenavirus GP2. While

large-scale structural rearrangements are essential to the functionality of fusion proteins, conformational modulation of attachment glycoproteins is not absolutely required for receptor-binding and thus not shared amongst diverse viral families to the extent of fusogenic rearrangements. As such, structural modulation of the arenaviral GP1, which does not possess clear analogy with any other viral attachment protein, as fusion proteins do, is relatively poorly understood.

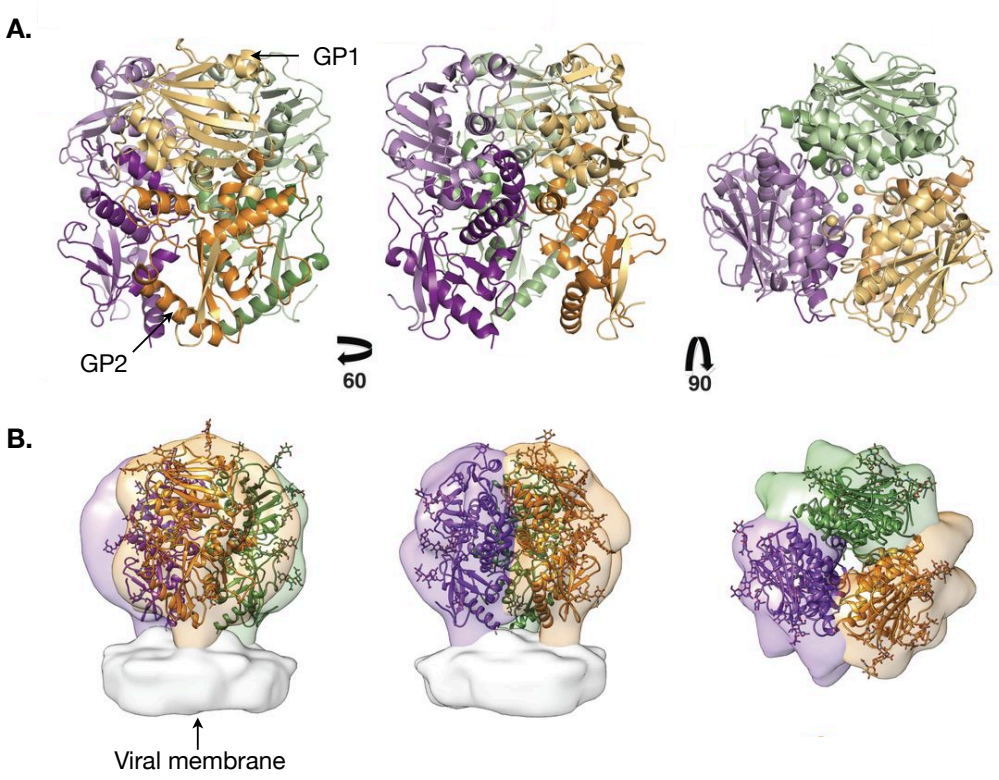


Figure 5.3i: Supramolecular organisation of LASV GPC. (A) The crystal structure of soluble trimeric LASV GP1-GP2 (PDB: 5VK2) reveals the compact globular architecture of the arenavirus GPC. The structure is displayed as cartoons with protomers coloured purple, orange, and green. In each instance, the GP1 from each protomer is coloured a lighter shade than GP2. (B) Fitting the crystal structure of LASV GP1-GP2 into the cryo-electron tomography reconstruction of GPC from fixed LASV virions (EMD-3290), shown as a surface. Panels are orientated with respect to one another. Figure adapted from Hastie et al. (2017).

Intriguingly, LASV requires the lysosome-resident protein LAMP1 for cellular entry (Jae et al., 2014) in addition to the primary receptor, α-dystroglycan (α-DG) (Cao et al., 1998). Recognition of the secondary receptor LAMP1, which appears to be a unique feature of LASV entry (Israeli et al., 2017), is pH-dependent and mediated by GP1 (Cohen-Dvashi et al., 2015).

Furthermore, the sialyltransferase, ST3GAL4, has been shown to act as critical modifier of LAMP1, adding a terminal sialic acid to the N-linked glycan present on N76, that permits its use as an entry factor for LASV infection (Jae et al., 2014). Following receptor engagement and cellular internalisation, LASV GP1 has been shown to shed from GPC in a low pH-dependent process in which LAMP1 is non-essential (Li et al., 2016), likely allowing the fusogenic rearrangements of GP2. In addition to detachment mediated by cellular-entry, isolated LASV GP1 has been reported to be actively secreted *in vitro,* from GPC expressing cells (Branco and Garry, 2009), and has been detected in patient sera during acute infection (Branco et al., 2010). Taken together, these observations highlight the potential virological and immunological importance of characterising isolated arenaviral GP1s.

Crystallographic studies of isolated GP1 glycoproteins from both OW and NW arenaviruses have revealed that all NW GP1s, and OW GP1s bound to a cognate GP2 molecule, adopt a shared α/β fold (Bowden et al., 2009b, Mahmutovic et al., 2015, Shimon et al., 2017, Hastie et al., 2016, Hastie et al., 2017). However, OW GP1s expressed in the absence of a cognate GP2 exhibit substantial structural deviations relative to their GP2-associated state, primarily within the helical regions of the molecule (Cohen-Dvashi et al., 2015, Israeli et al., 2017) (**Fig 5.3.i and ii**).

Following the publication of the structure of the GP1 from Machupo virus (MACV) (Bowden et al., 2009b), the crystal structure of an equivalent, truncated LASV GP1 (Cohen-Dvashi et al., 2015) revealed a surprising level of structural difference between the two molecules. While the central β-sheet is structurally conserved, the number and orientation of the helical segments of the molecules are diverse. The structures of the Lymphocytic choriomeningitis virus (LCMV) GP1 (Hastie et al., 2016) and LASV GP1 (Hastie et al., 2017) bound to cognate GP2 molecules, in the latter case as part of an assembled trimeric GPC complex, revealed that the observed structural deviations within the GP2-free $GP1^{LASV}$ do not solely reflect the sequence diversification between OW and NW GP1s. Indeed, GP2-associated

GP1^{LASV} more closely resembles NW GPIs than it does the GP2-free state of GP1^{LASV}. These data suggest that LASV GP1 is capable of adopting at least two distinct conformational states. The structure of the GP1 from the OW Morogoro virus (MORV) revealed that the observed conformational rearrangements are not unique to LASV and might be a more widely conserved feature of OW arenaviruses (Israeli et al., 2017).

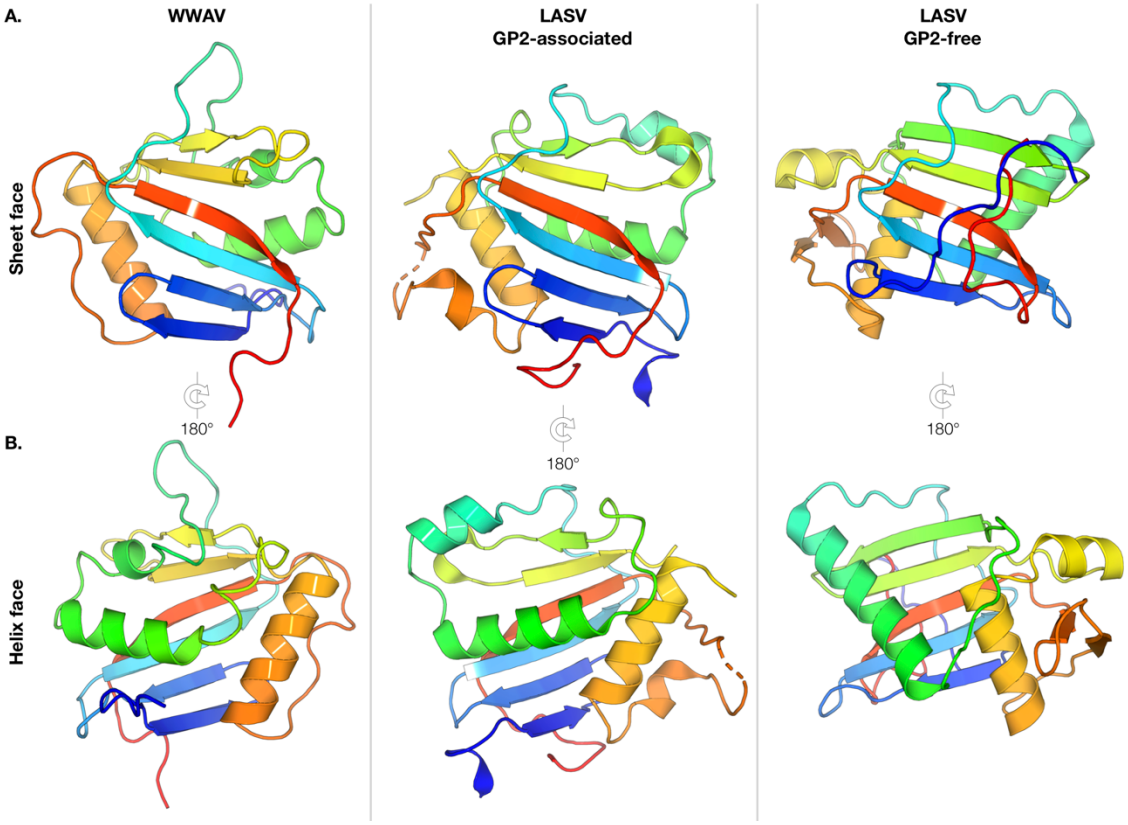


Figure 5.3i: Structural variation of arenavirus GP1s. LASV GP1 adopts two distinct conformations in the absence ('GP2-free' PDB: 4ZJF) or presence ('GP2-associated' PDB: 5VK2) of cognate LASV GP2. In the GP2-associated state, LASV resembles the GP1s from NW arenaviruses, such as WWAV. The structures of WWAV and LASV GP1 are shown as a cartoon coloured in a gradient from blue to red, from N- to C-terminus. All structures are shown to scale and in equivalent, aligned orientations for ease of comparison. The GP2-associated LASV GP1 was truncated to the same length as the GP2-free structure for illustrative purposes. (A) GP1s are orientated showing the sheet face. (B) Structures are rotated 180° about a vertical axis relative to A, displaying the helix face.

The GP2-free and GP2-associated structures of LASV GP1 permit a direct comparison of the structural rearrangements within OW GPIs. Upon superimposition, only the central β-sheet of the two structural states align, with some connecting loops exhibiting similar conformations

(**Fig 5.3ii**). While the sheet face appears to be the most structurally conserved portion of the molecule, a dramatic reorientation of both termini results in a topologically distinct surface in each of the two states.

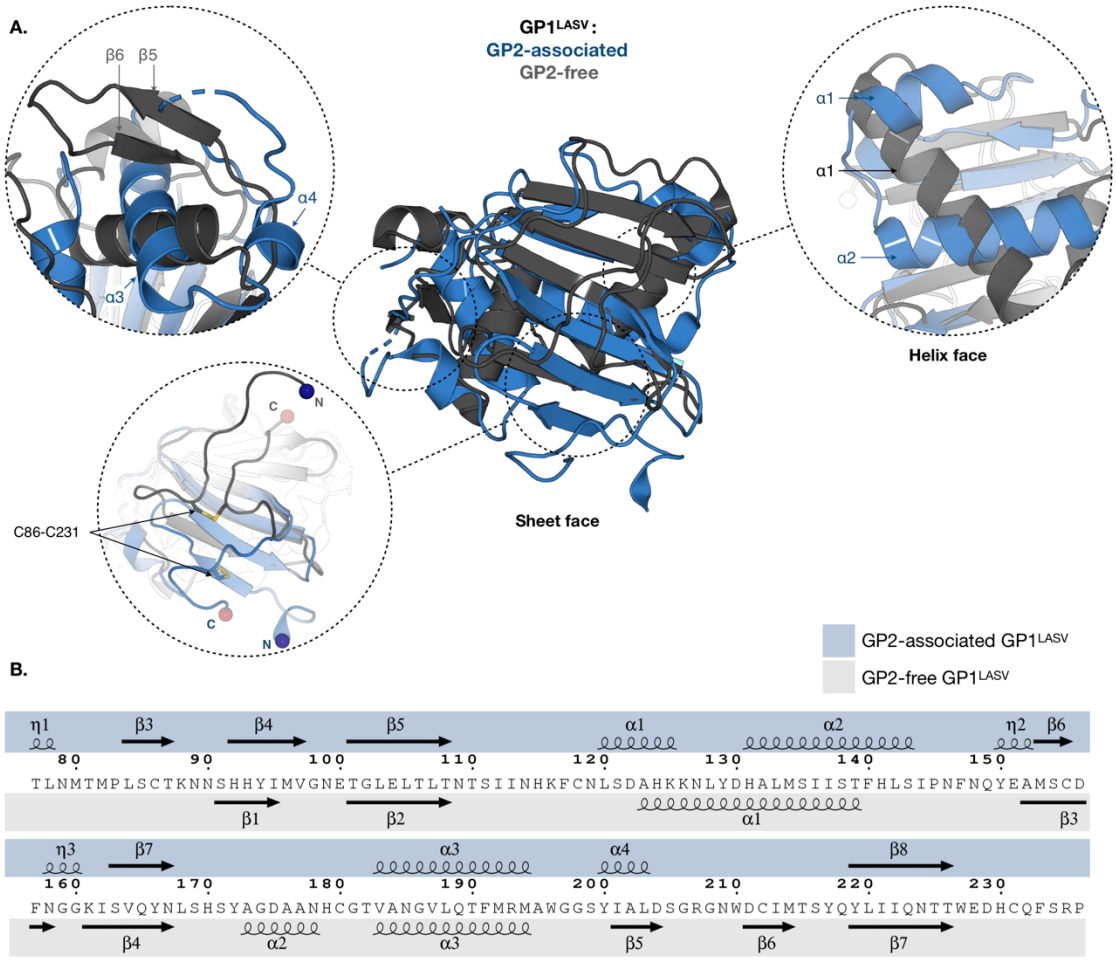


Figure 5.3ii: Structural variation of LASV GP1. (A) The two known LASV GP1 structures are shown as cartoon superimposed on one another. GP2-asscoiated GP1 (PDB: 5VK2) is coloured blue and GP2-free GP1 (PDB: 4ZJF) is coloured dark grey. Regions of interest are highlighted with zoomed panels (dotted lines) and labelled accordingly. The C86-C231 disulphide bond in shown in sticks with sulphur atoms coloured yellow. Where relevant, the C- and N-termini are illustrated with red and blue spheres, respectively. (B) Secondary structure annotation of the LASV GP1 sequence, the secondary structural elements of the GP2-associated and GP2-free conformations are shown above (highlighted blue) and below (highlighted grey) the sequence, respectively. Sequence annotation was performed using ESPript (Gouet et al., 2003).

The N terminus of the GP2-free GP1 is displaced by around 35 Å relative to the GP2-associated GP1 and in so doing, the first strand of the central sheet is repositioned to form a loop structure (residues 84–88). Similarly, the C-termini in both states are separated by around 31 Å. Both termini reposition from a downward (towards the virion surface) orientation, to an

opposing position, facing toward the apex of the trimeric GPC, occluding the upper portion of the β-sheet. Interestingly, the C- and N-terminal segments are linked by a long range disulphide bridge, from the C terminal loop to the β3-strand (in GP2-assciated GP1), which likely couples the movement of the two termini, since repositioning the disulphide would require concerted movement of both regions of the polypeptide.

The majority of structural rearrangements map to the helix face of the molecule, wherein all helical segments undergo either spatial or compositional rearrangements. The α1 and α2 helices of the GP2-associated state combine to form a single extended helix, termed α1 in the GP2-free state, that is repositioned relative to both α1 and α2. Whilst compositionally identical in both states, the α3 helix is repositioned, presumably to accommodate the newly formed β-hairpin. The β-hairpin in the GP2-free state comprises two antiparallel β-strands (β5 and β6), which form a short helix (α4) and partially disordered loop in the GP2-associated state. Furthermore, the unmodeled, and presumably disordered, loop comprising residues 171-179 forms an α helix (α2) in GP2-free state which is unlikely to form in the GP2-associated state, given its observed disorder.

The trigger for the large-scale conformational changes within LASV GP1, and other OW GPls, is unknown. However, based upon the observation that all current GP2-free GP1 structures were crystallised in acidic conditions, it was postulated that pH may be a bidirectional trigger for the transition from the one state to another (Cohen-Dvashi et al., 2015). Supporting a role for pH as a potential trigger for structural changes within GP1, arenaviruses are known to fuse in the acidic environment of endosomes (Nunberg and York, 2012), wherein GP1 is shed from GPC, and in the case of LASV, undergoes a pH-dependent receptor switch (Jae et al., 2014).

Here, we sought to characterise the discrete structural states adopted by the conformationally labile arenavirus GPC and its GP1 sub-component. Furthermore, we characterise the interaction of LASV GPC with the intracellular receptor, LAMP1.

5.4 Structural transitions and LAMP1 binding of LASV GPC

5.4.1 Construct design and small-scale expression testing of LAMP1

Constructs derived from codon-optimised synthetic genes encoding the entire reading frame of human LAMP1 (UniProt: P11279) and a mutant possessing an amino acid substitution at position 76 (N76S), which mimics chicken LAMP1, that is non-permissible to LASV entry. Constructs were generated for the full-length lumenal domain (S27–M382), termed $LAMP1^{FL}$; first lumenal domain (S27–D194), termed $LAMP1^{D1}$; and second lumenal domain (N228–M382), termed $LAMP1^{D2}$ (**Fig 5.4.1**). Analogous constructs harboring the N76S substitution were also produced for $LAMP1^{FL}$ and $LAMP1^{D1}$, termed $LAMP1^{FLS76}$ and $LAMP1^{D1S76}$, respectively. Domain boundaries were primarily devised based upon UniProt domain assignments.

$LAMP1^{D1}$, $LAMP1^{D2}$, $LAMP1^{FL}$, $LAMP1^{D1S76}$, and $LAMP1^{FLS76}$ were all secreted from HEK293T cells, indicative of successful folding (**Fig 5.4.1**). In an attempt to produce an un-glycosylated sample for biochemical analysis, LAMP1 expression was trialed in a bacterial expression system (*E. coli* Rosetta-gami™ 2(DE3)), optimised for the production of mammalian proteins. $LAMP1^{D1}$ (S27–D194) and $LAMP1^{D1S76}$ were cloned into the pOPINE and pOPINS3C vectors that encode an N-terminal histidine tag, the expression chaperone SUMO (in the case of pOPINS3C), and a 3C protease cleavage site (Bird et al., 2014). Not unexpectedly, expression did not yield folded protein, likely due to a requirement of the extensive glycosylation for efficient folding and stabilisation of the protein.

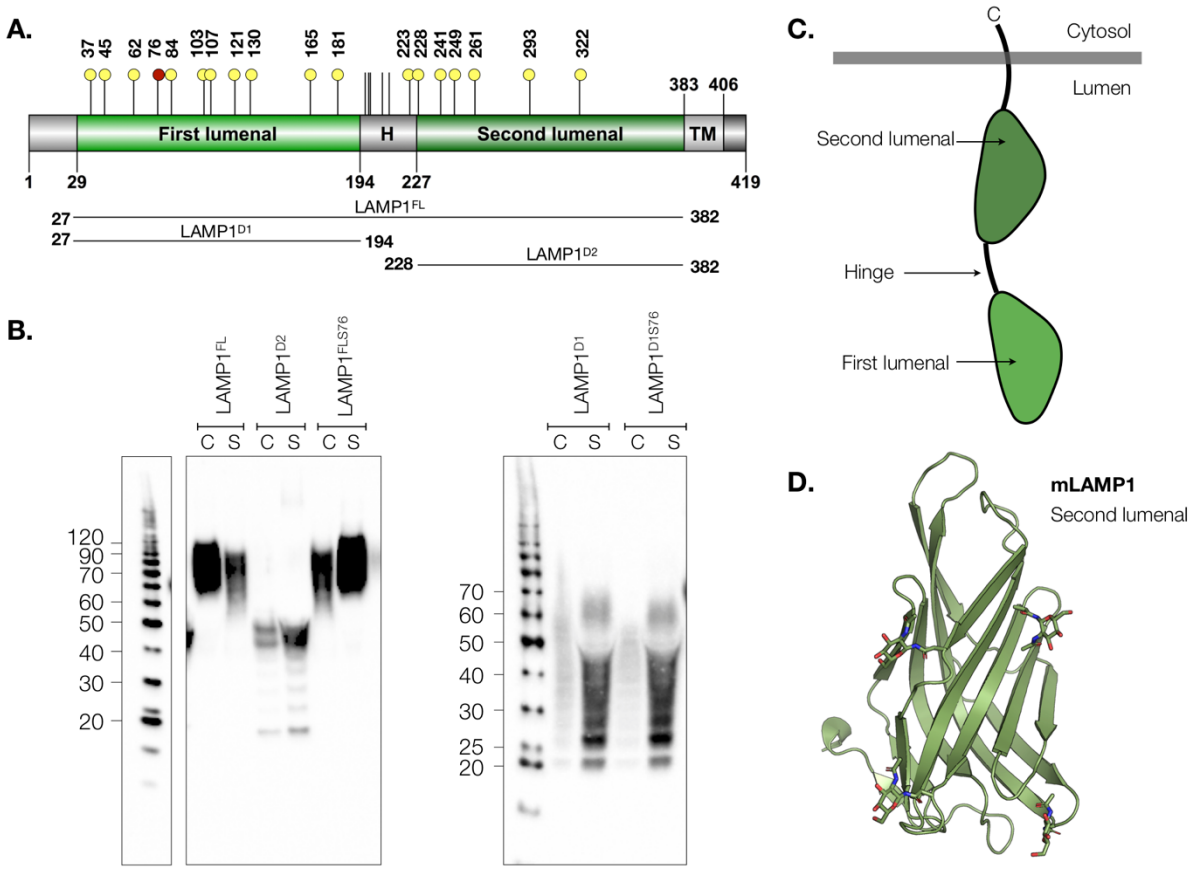


Figure 5.4.1: Domain organisation, construct design, and expression testing of LAMP1. (A) Domain organisation and post-translational modifications of human LAMP1 (UniProt: P11279). The structural domains of LAMP1, the first lumenal and second lumenal, are coloured light and dark green, respectively. The connecting mucin-like hinge domain is designated with 'H'. The C-terminal transmembrane region is labelled as TM. Putative N-linked glycosylation sites are labelled with yellow pins, and the site identified as crucial to LASV GPC binding is coloured red (N76) (Jae et al., 2014). Putative O-linked glycosylation sites are shown as vertical lines and left unlabelled for clarity. Expression construct boundaries are indicated as lines below the sequence. (B) Small-scale expression testing of LAMP1 in HEK293T cells. Western blot analysis of cellular (C) and supernatant (S) fractions from small-scale transient transfections with LAMP1. (C) Schematic representation of LAMP1 coloured as in A. (D) The crystal structure of the second lumenal domain, which is a homologue of the first lumenal domain, from mouse LAMP1 reveals a unique β-prism fold (Terasawa et al., 2016). The protein is coloured green and shown as a cartoon with N-linked glycans shown as sticks.

5.4.2 Large-scale expression and purification of LAMP1

$LAMP1^{D1}$, $LAMP1^{FL}$, and $LAMP1^{D1S76}$ constructs were used to transiently transfect HEK293T cells. The α-mannosidase inhibitor, kifunensine, was utilised to modulate the resultant glycoforms presented by LAMP1 (Chang et al., 2007). Untreated cell cultures would likely result in LAMP1 constructs with near-native glycoforms, and kifunensine treated samples with high-mannose glycoforms ($Man_{6-9}GlcNAc_2$) that could be further processed by enzymatic deglycosylation with

endoF$_1$ to bear only single N-acetylglucosamine glycoforms, thereby permitting assessment of the glycan dependence of the LASV GPC-LAMP1 interaction.

Cell culture supernatants were clarified by centrifugation and filtration, and subsequently buffer exchanged into 10 mM Tris pH 8.0, 150 mM NaCl (TBS), prior to being subjected to combined nickel-affinity and size-exclusion chromatography (SEC) schedule. LAMP1 proteins eluted from SEC in symmetric peaks, indicative of a monodisperse samples (**Fig 5.4.2i**). When assessed by SDS-PAGE, LAMP1^{D1} appeared pure, though the sample exhibited a heterogeneous population with a broad range of molecular masses (20–40 kDa), likely due to the large number of glycosylation sites present on the protein being differentially occupied (**Fig 5.4.2i**).

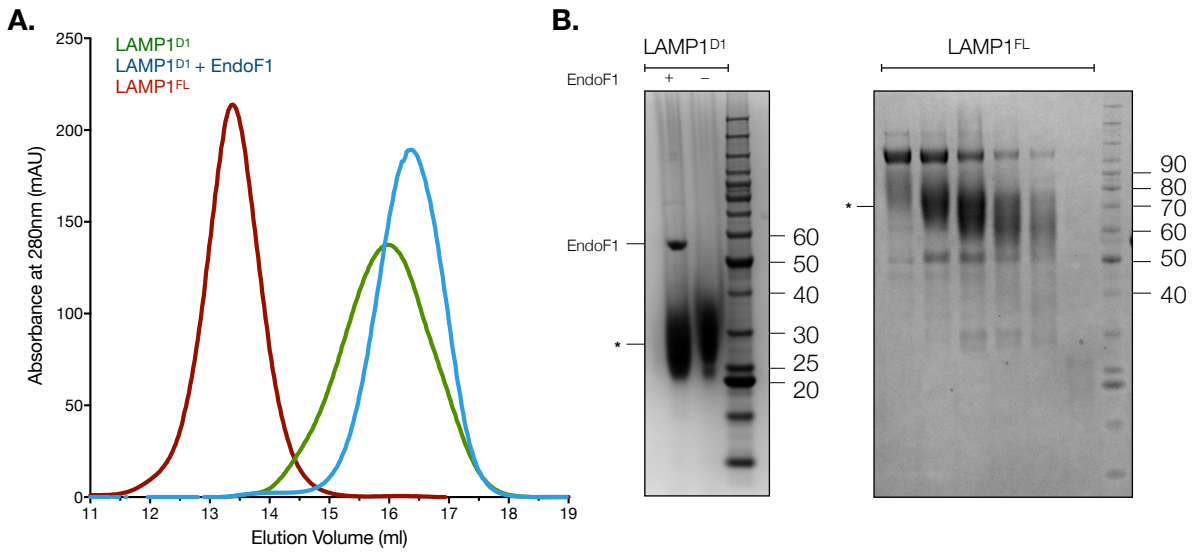


Figure 5.4.2i: Purification of LAMP1. (A) Representative SEC chromatograms of LAMP1^{D1} (green and blue) and LAMP1^{FL} (red) constructs. EndoF$_1$ treated LAMP1^{D1} (blue) elutes in a narrower peak with a mass shift, relative to the untreated sample (green). (B) Representative reducing SDS-PAGE analysis of LAMP1 constructs. LAMP1^{D1} appears largely resistant to endoF1 mediated deglycosylation. LAMP1^{FL} co-elutes with a number of common contaminants. Putative LAMP1 bands are indicated with an asterisk in each case. Molecular weight standards are displayed and the approximate molecular masses shown in kDa.

5.4.3 The pH and glycan dependence of LAMP1 binding to LASV VLPs

To determine the minimal structural domain of LAMP1 required for binding to LASV VLPs, pure LAMP1^{D1} was utilised in enzyme-linked immunosorbent assays (ELISA) (**Fig 5.4.3i**). LAMP1^{D1}

encodes the first lumenal domain of the protein (**Fig 5.4.1**), which has been suggested to be important for the LASV interaction (Jae et al., 2014). ELISA analyses revealed that LAMP1^{D1} binds LASV VLPs in a pH dependent manner (**Fig 5.4.1**). Apparent binding of LAMP1 at pH 8, which is significantly above the corresponding pH 8 control (p=0.0002), is likely a consequence of non-specific interactions, not fully accounted for by the negative control. Supportive of this, the LAMP1 concentration series at pH 8 (**Fig 5.4.3i B**) can not be fitted with a non-linear model for protein-protein binding, whilst the pH 5 data fit with an R^2 of 0.93, supportive of a specific interaction between LAMP1 and LASV at pH 5 but not pH 8.

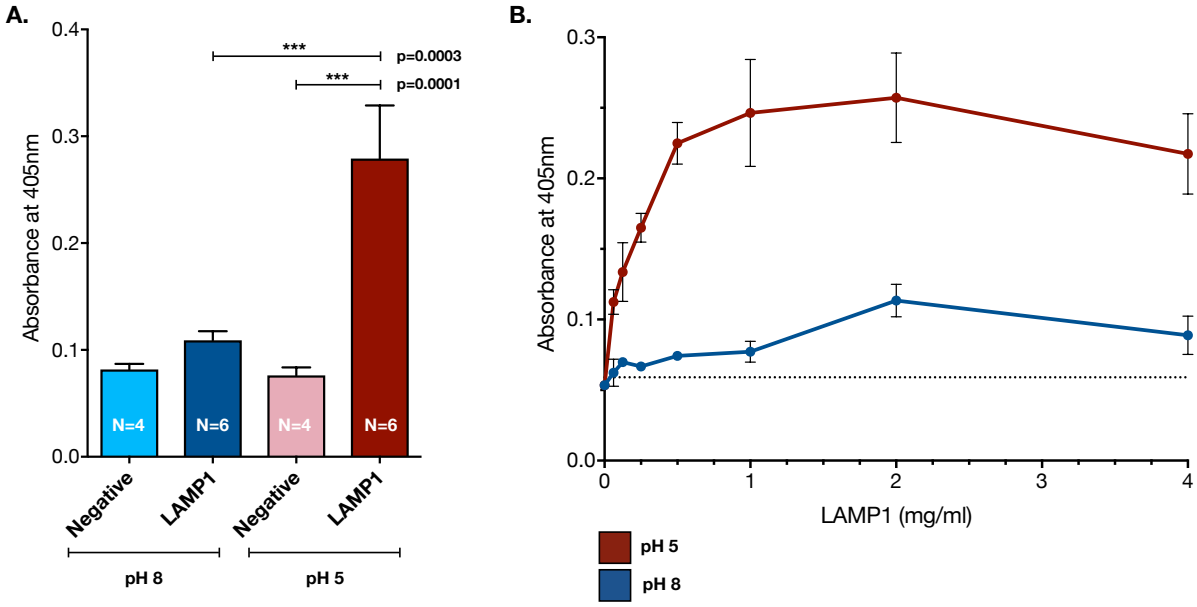


Figure 5.4.3i: The first lumenal domain of LAMP1 binds LASV VLPs in a pH dependent manner. Enzyme-linked immunosorbent assays were performed in 96-well Nunc-Immuno™ PolySorp® plates (Sigma) to which LASV-VLPs were bound at a concentration of approximately 10 µg/ml. Plates were blocked with PBS containing 5% milk powder and incubated with 50 µl LAMP1^{D1} at the indicated pH. LAMP1 detection was mediated using a rabbit derived anti-6His primary (abcam), and goat-anti rabbit-HRP (abcam) antibody combination. Binding was assessed on the basis of optical absorbance at 405nm following HRP substrate addition. (A) Binding of LAMP1^{D1} at 1 mg/ml is significantly greater at pH 5 than at pH 8. The number of biological replicates is indicated with 'N' on each bar. Error bars denote standard deviation. Statistical significance was calculated using an unpaired T-test with Welch's correction. Significance of P<0.001 is denoted with ***. (B) LAMP1^{D1} binding to LASV VLPs at pH 5 (red) and pH 8 (blue) is plotted as a function of LAMP1^{D1} concentration. Error bars denote the standard deviation of three independent measurements. Error bars are not shown in instances where they are smaller than the data point symbol. The average absorbance value for mock wells (PBS only) is shown as a dotted line.

Interestingly, the LAMP1^{D1} construct used for the ELISA binding analysis was expressed in the presence of kifunensine, and would thus be expected to bear only oligomannose

glycoforms, not the complex-type sialylated glycans that have been previously postulated to be critical for the LAMP1-LASV GPC interaction (Jae et al., 2014, Israeli et al., 2017). Compositional glycan analysis of LAMP1^{D1}, expressed in the absence of kifunensine, was performed to assess whether the presence of sialylated glycans on LAMP1 is a pre-requisite for the LASV GPC interaction *in vivo* (**Fig 5.4.3ii**).

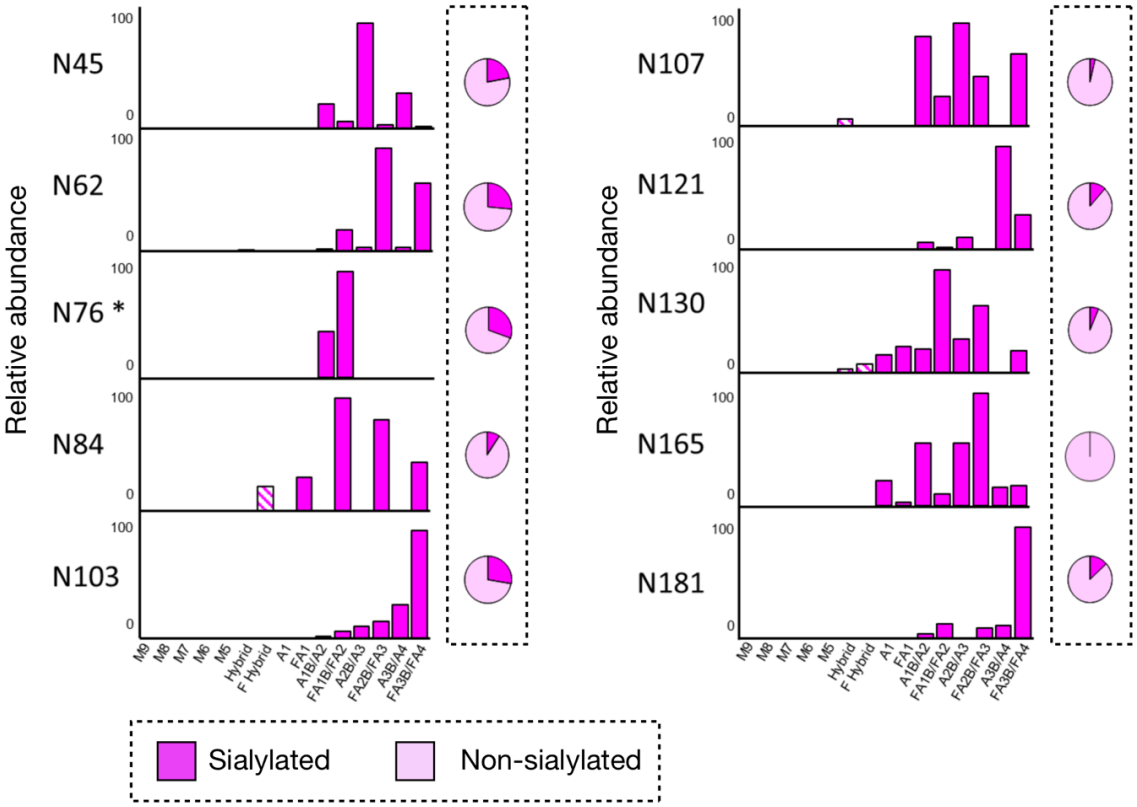


Figure 5.4.3ii: Position N76 on LAMP1 is incompletely sialylated. Quantitative site-specific analysis of N-linked glycans on recombinant LAMP1^{D1}. Bar graphs represent the relative quantities of glycan species: the oligomannose-type glycan series (M9 to M5; $Man_9GlcNAc_2$ to $Man_5GlcNAc_2$), afucosylated and fucosylated hybrid glycans (Hybrid and F Hybrid, respectively) shown as dashed pink bars, and complex glycans grouped according to fucosylation and number of antennae (A1 to FA4) coloured solid pink. Pie charts summarise the quantification of the complex glycans that are sialylated (dark pink) and non-sialylated (light pink). Position N76 which bears a glycan postulated to be critical for the LAMP1-LASV GPC interaction is denoted with an asterisk. For mass spectrometry (MS) of glycopeptides, 50 µg aliquots of recombinant LAMP1 were proteolytically digested using trypsin and chymotrypsin (Promega). C18 ziptips (Merck Millipore) were used to purify glycopeptides for LC-ESI MS. For site specific analysis, glycopeptides were analysed on an Orbitrap Fusion™ Tribrid™ mass spectrometer (Thermo Fisher Scientific) coupled to an EASY-Spray™ nano-LC system (Thermo Fisher Scientific). MS data were acquired with XCaliburTM 4.0 (Thermo Fisher Scientific). Glycopeptides were fragmented using both higher energy collisional dissociation (HCD) and energy transfer dissociation (ETD). Data analysis and glycopeptide identification were performed using Byonic and Byologic software (Protein Metrics).

Recombinant LAMP1^{D1}, expressed in the absence of kifunensine, presents complex-type glycans at all detected sites (N45, N62, N76, N84, N103, N107, N212, N130, N165 and N181). Quantification of the glycoforms revealed that only approximately 30% of the glycans present at N76 are sialylated. Combined with ELISA analyses that reveal sialylated glycans are dispensible for the LAMP1-LASV GPC interaction (**Fig 5.4.3i**), glycan composition determination reveals that recombinant LAMP1 presents only a low abundance of terminally sialylated glycoforms.

5.4.4 pH-induced structural rearrangements of LASV GPC accommodate LAMP1

Having established that the membrane distal domain of LAMP1 alone is sufficient for binding to LASV GPC (**Fig 5.4.3i**), and that sialylated glycans at position N76 appear to be dispensable for the interaction, we sought to determine the LAMP1^{D1} binding interface on the LASV GPC, alongside an ongoing cryo-electron microscopy study. Tomographic reconstructions of the LASV GPC at three physiologically relevant pHs (pH 7, 5, and 3) revealed that the spike complex undergoes pH-dependent structural changes (**Fig 5.4.4**) (Li et al., 2016).

The loss of density attributed to shedding of GP1, confirmed by western blot analysis, permitted putative assignation of the GPC subunits (**Fig 5.4.4**). GP1 likely occupies the membrane distal region of the spike complex, with GP2 positioned beneath, providing the tripodal anchor to the viral envelope. This hypothesis was later confirmed by the high-resolution crystal structure of LASV GP1-GP2 (Hastie et al., 2017).

At pH 5, and in the absence of LAMP1, LASV GPC undergoes structural rearrangements within the membrane-proximal transmembrane-anchoring 'legs' and a subtle opening of the trimer, when viewed from its apex (**Fig 5.4.4**). The observed loss of density at the membrane proximal region of LASV GPC could be attributed to the repositioning of the N-terminus of GP1 that forms a β-strand which extends down the face of GP2 in the neutral pH crystal structure of trimeric LASV GPC (Hastie et al., 2017), but is repositioned to the opposing direction in the

structure of the isolated LASV GP1 (Cohen-Dvashi et al., 2015). Given the moderate resolution (16.5 Å) of the pH 5 reconstruction, it is difficult to further interpret the nature of the opening of the spike complex.

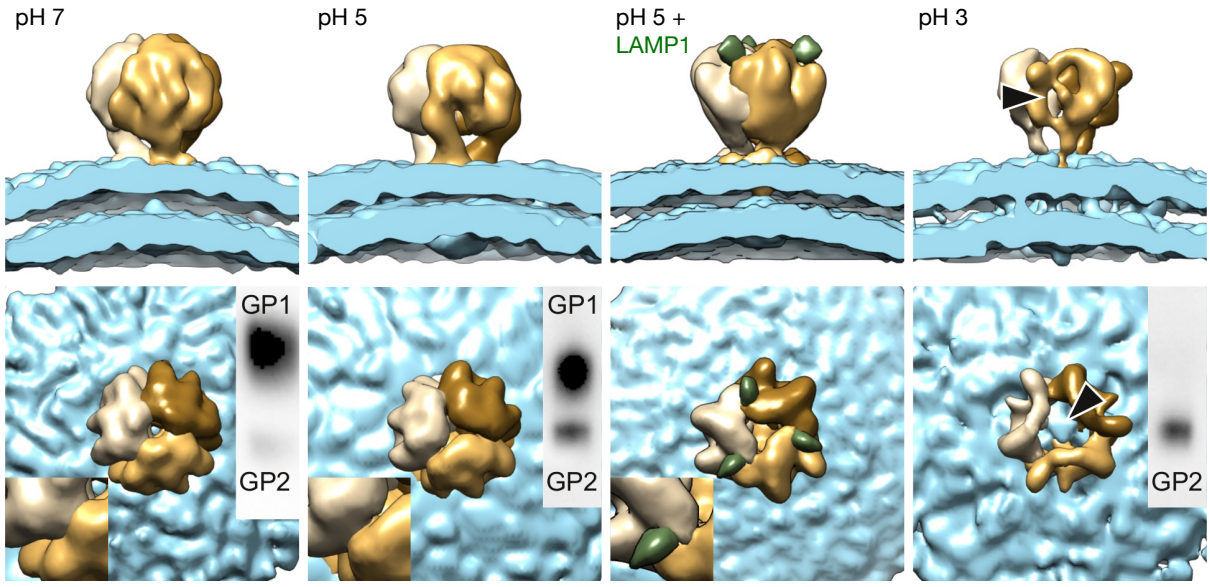


Figure 5.4.4: Acidic pH-induced structural changes and LAMP1 binding of GPC. Cryo-electron tomography derived reconstructions of LASV GPC structures at different pHs are shown in side (top row) and top (bottom row) views. All volumes were filtered to 17 Å resolution. LASV GP1-GP2-SSP protomers are coloured shades of brown and the residual density attributed to LAMP1 is coloured in green. Inserts in the lower left corners show close-up views of the inter-protomer interfaces. Insets in the top right corners show Western blot analysis of GP1 and GP2. Arrowheads indicated missing density, attributed to the loss of GP1, within the remaining spike complex. Figure taken from (Li et al., 2016).

When incubated with LAMP1^{D1}, LASV GPC exhibited additional density not apparent in the reconstruction lacking the receptor. Interestingly, LAMP1 appeared to bind at the newly-formed inter-protomer crevices, suggestive that the pH-dependent structural rearrangement of GPC is independent of, but important for, LAMP1 binding. The putative LAMP1 density is not sufficient to account for the predicted molecular mass of the construct, suggesting LAMP1 binding is either sub-stoichiometric or flexible.

Intriguingly, at pH 3, the remaining spike complex, that likely comprises GP2, and SSP which is unresolved at this resolution, does not resemble the canonical post-fusion six-helix

bundle architecture of class I fusion proteins, adopted by the both OW and NW arenaviruses (Igonet et al., 2011, Parsy et al., 2013). As such, the structural and functional features of this structure are unclear and may represent either a non-productive conformation formed in the absence of a target membrane or a stable fusion intermediate not yet observed by other methods.

Taken together, tomographic reconstructions of the LASV GPC, at pHs encountered during LASV cellular entry *in vivo*, reveal at least three distinct, and presumable stable, pH-inducible structural states. Furthermore, these data corroborate studies that suggest GP1 is shed from the complex during viral entry at pH 4.5 and below (Jae et al., 2014).

5.5 The pH dependence of isolated WWAV and LORV GP1s

5.5.1 Crystallisation of WWAV and LORV GP1

To assess the structural and functional features of the GP1 sub-component of the arenaviral GPC, that is likely analogous to GP1 that is shed during retrograde endosomal trafficking, we performed comparative crystallographic and solution state analyses of isolated NW and OW GP1s.

The crystal structures of WWAV and LORV GP1 were both solved at near-neutral and acidic pH. For simplicity, in subsequent discussion 'neutral' pH is defined in the range of 7.5 to 8.0 and 'acidic' in the range 5.0 to 5.6, reflective of the pH at which each structure was solved. Methodological details pertaining to all processes relevant to obtaining the crystal structures of WWAV GP1 are described in Chapter 4. Briefly, crystals from which the structures derive were formed in two conditions: 10.1 mg/ml protein, 0.2 M potassium sodium tartrate, 2 M ammonium sulphate and 0.1 M tri-sodium citrate pH 5.6; and 11.0 mg/ml protein, 0.1 M HEPES-Na pH 7.5, and 0.05 M cadmium sulphate. Crystals from which the structures of LORV GP1 derive were similarly formed in two conditions: 4.5 mg/ml protein, 30% w/v PEG 6000 and 0.1M citrate pH 5.0; and 5.3 mg/ml protein, 10% w/v ethanol, and 1.5 M NaCl. In all instances,

proteins were in a solution comprising 10 mM Tris pH 8.0 and 150 mM NaCl. One LORV condition was unbuffered by the crystallisation screen, thus the final pH was dictated by Tris (5mM final concentration in the drop) derived from the protein solution, confirmed by pH measurement.

5.5.2 The crystal structures of WWAV and LORV GP1 at neutral and acidic pH

Comparison of the neutral and acidic pH crystal structures of WWAV GP1 reveals the protein to be essentially identical under both conditions, with the two structures exhibiting an RMSD of 0.4 Å across all common Cα positions (**Fig 5.5.2i**). Given that both crystals formed in the space group $P6_322$, with nearly identical unit cell dimensions, the similarity of the respective crystal environments likely accounts for the high degree of similarity to one another. Structural differences between the two pH structures are minimal and limited primarily to the β3-β4 loop and the η3 helix. Both regions exhibit similar variation, with RMSDs of approximately equal magnitude, when compared with the published structure of WWAV GP1 (**Fig 4.6.3i**), indicative of inherent flexibility. Supportive of such inherent flexibility, the β3-β4 loop and the η3 helix of both structures exhibit above average *B*-factors. Taken together, these data suggest that pH does not modulate the conformation of WWAV GP1.

As was observed for WWAV GP1, the structures of LORV GP1 at neutral and acidic pH are essentially identical, exhibiting a RMSD of 0.7 Å across all common Cα positions (**Fig 5.5.2i**). Secondary structure elements are in close agreement between the two structures; variations of >1 Å are only observed in loop regions of the molecule, with the C-terminal and β1-β2 loops exhibit locally elevated RMSDs, likely due to inherent flexibility. LORV GP1 crystallised in two distinct space groups, $P4_32_12$ at pH 5.0 and $P4_12_12$ at pH 8.0, and thus the observed variation may also be a consequence of the local crystal environments, which were likely more diverse than those experienced by WWAV GP1.

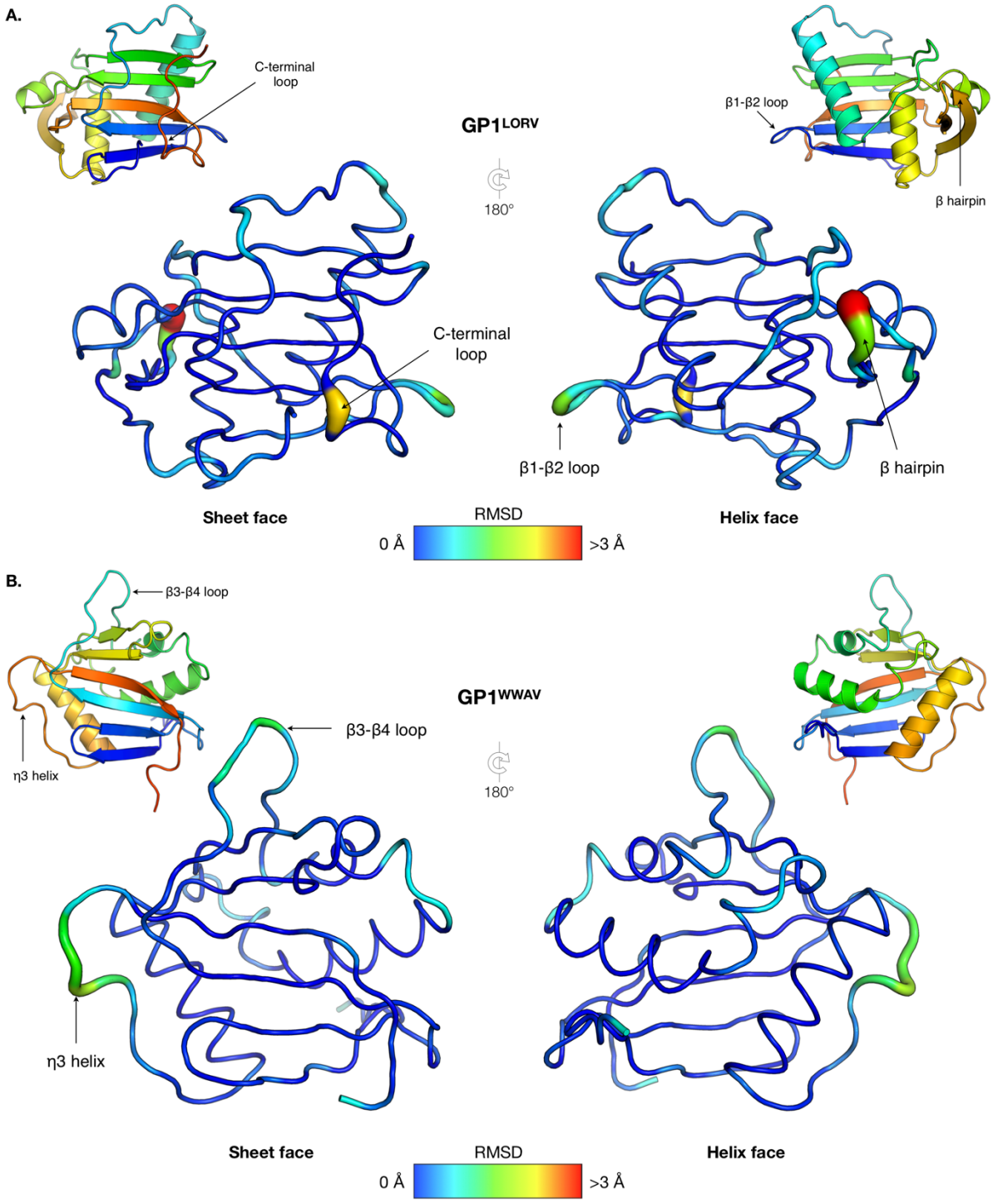


Figure 5.5.2i: Comparison of the neutral and acidic pH structures of WWAV and LORV GP1. (A) The structural differences between the neutral and acidic pH structures of LORV GP1 are illustrated on the molecule displayed as a tube. Root-mean-square deviation (RMSD) between the Cα positions of equivalent residues is represented by both colour (ramped from blue to red with increasing RMSD) and tube width (ramped from thin to thick with increasing RMSD). Regions of greatest variability are labelled. In each instance the higher resolution structure of the acidic pH LORV GP1 is shown as a cartoon in an identical orientation, coloured blue to red, from N- to C-terminus for ease of interpretation. (B) The structural differences between the neutral and acidic pH structures of WWAV GP1 are illustrated as in part A. RMSD illustrations are displayed on the same scale and thus can be directly compared.

The crystal structures of the GPIs from OW arenaviruses, expressed in the absence of GP2, are remarkably similar. When compared with LASV GP1 and MORV GP1, the higher resolution acidic pH LORV GP1 structure exhibits RMSDs of 0.8 Å and 0.7 Å, and amino acid sequence identities of 67% and 68%, respectively (**Fig 5.5.2ii**). These data indicate that, for the majority of the molecule, the GP2-free state of the GPIs from LASV, MORV and LORV are essentially indistinguishable in terms of Cα position and thus main chain conformations. The GP2-free state is therefore unlikely to constitute a crystallisation or expression artefact.

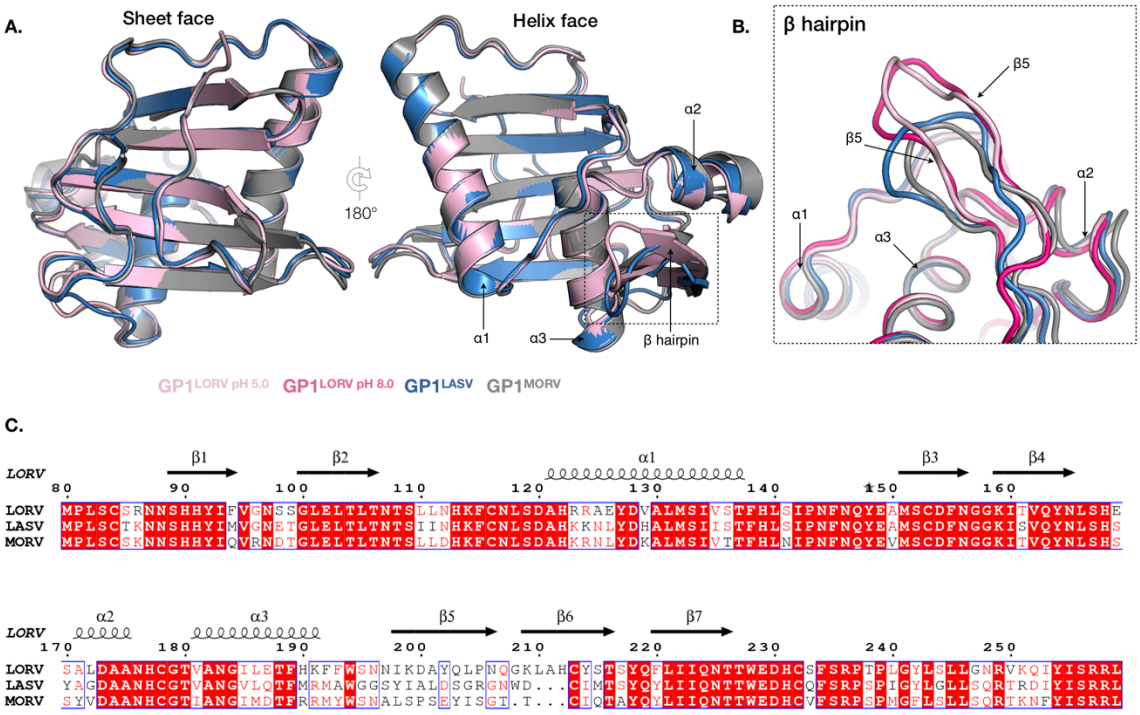


Figure 5.5.2ii: Comparison of GP2-free OW GP1s reveals striking structural similarity. (A) Superimposition of known OW GP1 structures expressed in the absence of a cognate GP2. Structures of the GP1s from LORV at pH5 (light pink), LASV (blue) and MORV (grey) are shown as cartoons with salient features labelled. (B) Zoom in view of the structurally variable β hairpin motif. Structures are displayed as tubes for clarity and the structure of LORV GP1 at pH 8 is added to illustrate the intra- as well as inter-protein variation. (C) Sequence alignment of the crystallised constructs of LORV, LASV and MORV. Identical residues are highlighted red, similar residues are coloured red and dissimilar residues are black. Secondary structure elements of the LORV crystal structure are annotated above the sequence and labelled according to type: β strands are shown with arrows and labelled β1-7; α helices are illustrated as coils and labelled α1-3. Sequence alignments were determined by MultAlin (Corpet, 1988) and plotted using ESPript (Gouet et al., 2003).

The greatest structural variation within LORV GP1 is observed in the loop of the 'β-hairpin' motif (Israeli et al., 2017), formed by the β5 and β6 strands. Interestingly, this region not only presents the most variability within LORV GP1 but is also the most diverse region across the known OW GP1 structures, both structurally and at the primary sequence level (**Fig 5.5.2ii**). It is likely that the β-hairpin represents a site of enhanced variability across OW arenaviruses, both in terms of sequence diversification and conformational plasticity. Concomitant with structural variability, the β-hairpin has been demonstrated to mediate a functionally distinct role in LASV, where it is implicated in binding the secondary receptor, LAMP1 (Israeli et al., 2017).

The crystal structures of GP1 from the NW WWAV and OW LORV, at both neutral and acidic pH, provide the first direct structural evidence that pH does not modulate the structures of isolated arenaviral GP1s. In light of the recent structure of a mature OW arenavirus GPC complex (Hastie et al., 2017), it is highly likely that the pH-independent 'GP2-free' conformation observed for OW GP1s is distinct from the conformation adopted when comprising mature GPC complexes. NW arenaviruses, on the other hand, appear not to occupy the 'GP2-free' conformational state observed in OW viruses and thus isolated NW GP1s are likely antigenically and functionally equivalent to GP1 on mature virions.

5.5.3 Structure-based classification of arenavirus GP1s

Structure-based phylogenetic analysis has utility in demonstrating the functional and evolutionary relationships of proteins and can add insight beyond that afforded by sequence-based methods alone. Structural phylogenies have been exploited to rationalise the relationships and functional characteristics of both viral (Bamford et al., 2005, Rissanen et al., 2017) and cellular proteins (Bowden et al., 2009a). Here, the Structural Hology Programme (Stuart et al., 1979) was employed to delineate the structural and functional relationships between arenavirus GP1 glycoproteins.

Concomitant with sequence-based phylogenetic analysis of arenaviral GPCs, structure-based phylogenetic analysis, using only the structures of the GP1 subcomponent, divides arenaviruses into the OW and NW lineages (**Figs 5.5.3 & 1.3.1i**). Furthermore, NW GP1s appear to segregate according to their respective clades, in accordance with genetic analysis of GP1. Interestingly, despite not exhibiting the same degree of genetic divergence within GPC as is seen within the analysed NW GP1s, OW GP1s form two distinct and distant clusters.

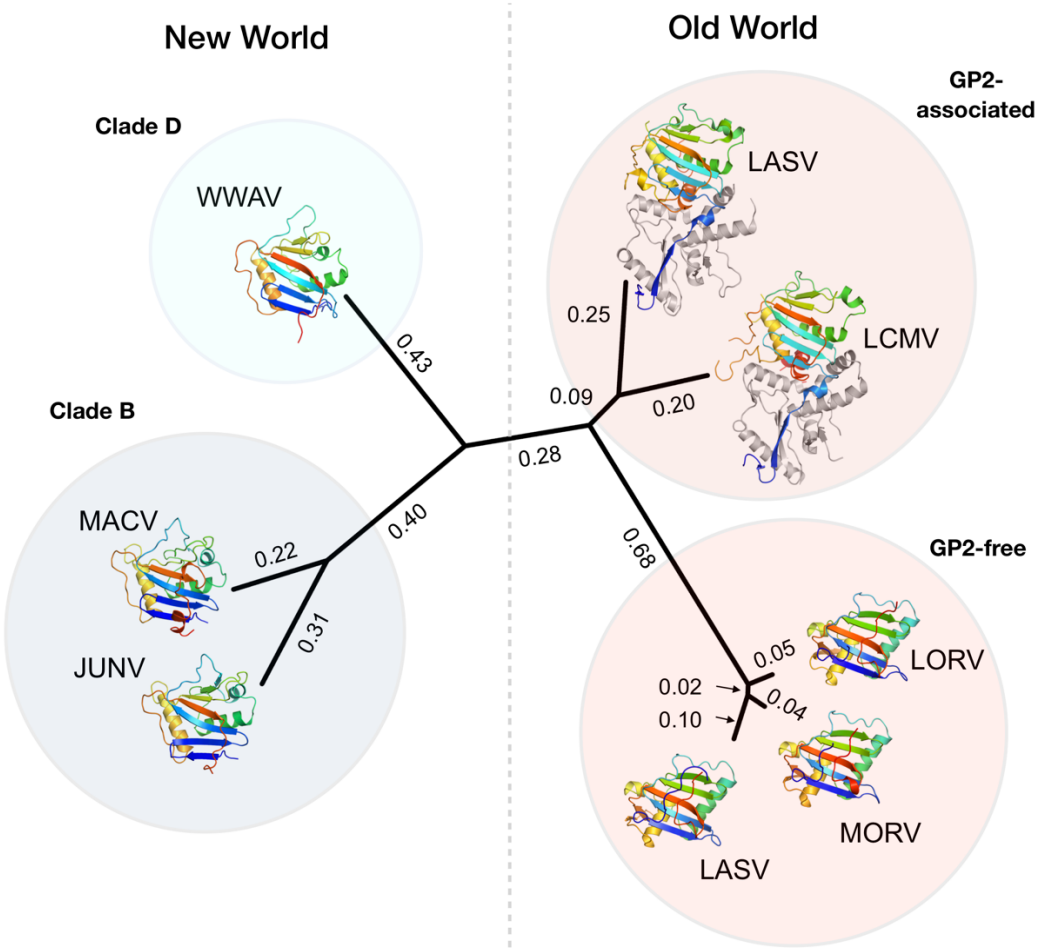


Figure 5.5.3: Structural phylogeny of arenaviral GP1s. Pairwise distance matrices were calculated using SHP (Stuart et al., 1979) and plotted as an unrooted tree with PHYLIP (Felsenstein, 1989). The crystal structures of the GP1 proteins from LASV (4ZJF & 5VK2), LCMV (5INE), LORV (pH 5.0), MORV (5NFF), JUNV (5NUZ), MACV (2WFO) and WWAV (pH 5.6) were used for phylogenetic analysis. For 5NUZ, 5VK2 and 5INE, all chains comprising non GP1 molecules (e.g. GP2 or antibody fragments) were removed prior to structure alignment. GP1 structures are displayed in equivalent orientations and shown cartoons coloured from N- to C-terminus. GP2 molecules are shown as grey cartoons where relevant. Evolutionary distances are annotated on each branch. OW arenaviruses are highlighted in red, NW clade D in cyan and NW clade B in blue. Coloured circles highlight the groupings discussed in the accompanying text.

Structural divergence of proteins within the two OW GP1 clusters is exemplified by LASV GP1, which resides in both groups. The bifurcation of LASV GP1 highlights the value of structure-based phylogenetic analyses in delineating discrete structural and functional states of proteins that possess identical primary sequence. Indeed, the GP1 proteins from LASV and LCMV, in their GP2-associated states, are more similar to one another than the two known conformations of LASV. Structures of the GP1 from LASV, along with other OW arenaviruses, highlight that OW GP1s can exist in at least two discrete conformational states, defined by substantial structural rearrangements within the helical and loop regions of the molecule. While the trigger for the conformational switching of OW GP1s is unknown, our crystallographic analyses suggest that pH does not trigger the conversion of GP2-free GP1s into their GP2-associated state. The two conformational groups within the OW lineage therefore appear to be defined by their association, or lack thereof, with a cognate GP2 molecule.

Another striking feature of our structure-based phylogeny is that although the GP1s from WWAV, MACV, and JUNV all utilise TfR1 as a receptor, WWAV is approximately equidistant from the clade B cluster and GP2-associated OW arenavirus GP1 structures (**Fig 5.5.3**). The pronounced structural variation of the NW GP1 glycoprotein scaffold within TfR1 tropic viruses highlights the variable modes with which GP1-TfR1 interactions are likely mediated. Furthermore, the absence of a 'GP2-free conformation' within WWAV, as well as MACV and JUNV GP1, despite all being solved in the absence of cognate GP2 molecules, and at both neutral and acidic pH, suggests that NW GP1s do not adopt the two discrete structural classes defined for OW GP1s. The ability of isolated WWAV and MACV GP1s to bind orthologues of the primary cellular receptor, TfR1 (Abraham et al., 2010, Shimon et al., 2017), and vaccine-elicited patient-derived monoclonal antibodies (Clark et al., 2018) further evidences the structural equivalence of isolated and GP2-bound NW GP1s. As such, isolated NW GP1s are likely reflective of mature virions.

5.5.4 Circular dichroism analysis of WWAV and LORV GP1

Whilst unlikely, it is possible that the crystal structures of LORV and WWAV GP1 represent a preferred conformational state selected for by the crystallisation process. In solution, the GP1 glycoproteins may therefore still be able interchangeably adopt the conformations that define the different structural classes, in a pH dependent manner. To assess whether the observed pH independence of the structures of WWAV and LORV GP1, obtained by X-ray crystallography, is reflective of the behavior of these proteins in a non-crystalline environment, solution-state circular dichroism (CD) spectroscopy analyses were performed. CD spectra are predominantly influenced by the secondary structure of a protein (Kelly et al., 2005) and thus suitable to examine to different conformational stated of arenaviral GPIs, that are distinct in their secondary structure content and organisation.

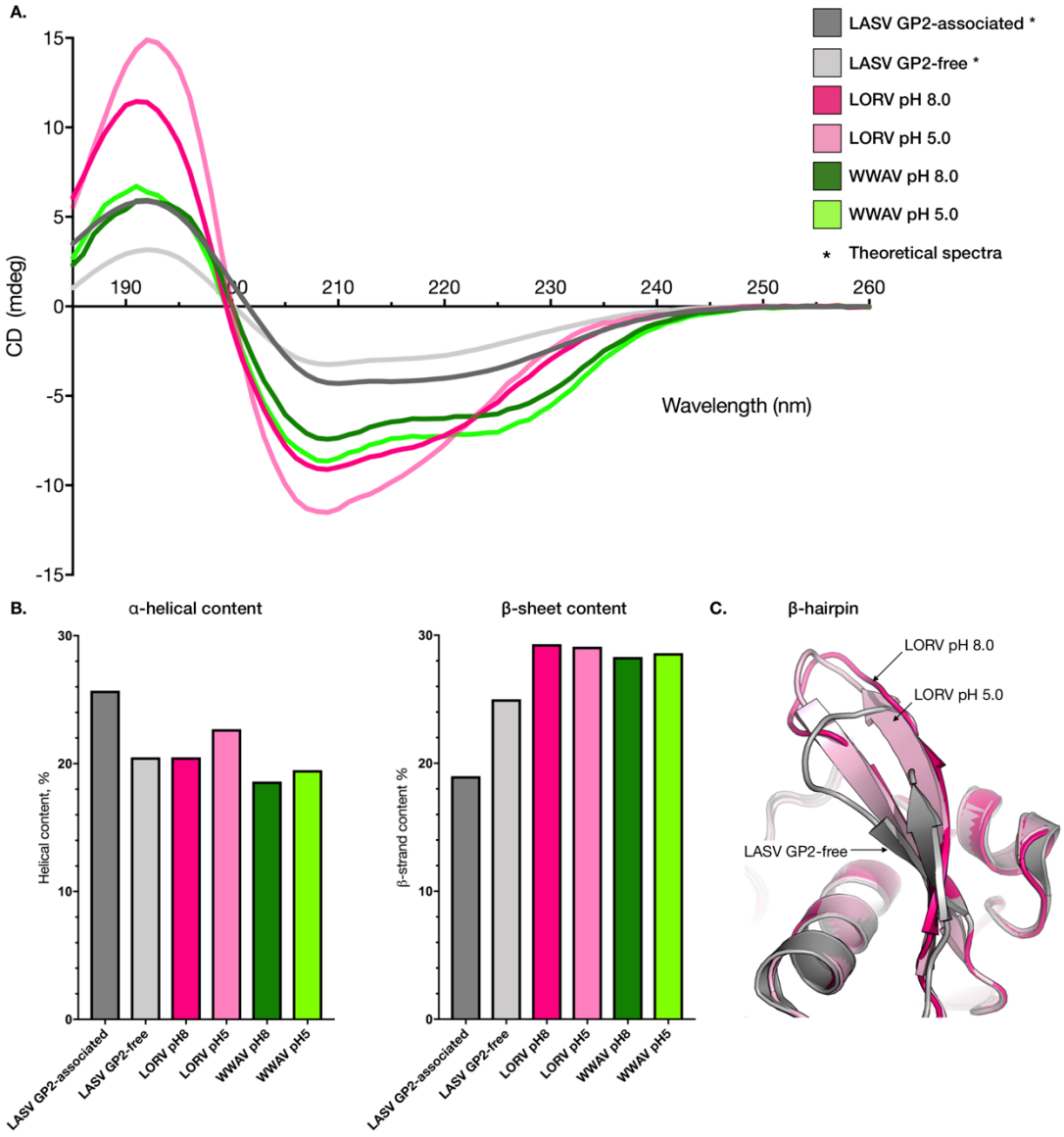
The CD spectra obtained for WWAV GP1 at pH 8.0 and 5.0 are very similar, both in terms of CD magnitude and the qualitative profile of the spectrum (**Fig 5.5.4i**). Consistent with these similarities, and supportive of our crystallographic analysis, the secondary structure content within the protein is essentially unaltered by ambient pH. Furthermore, the CD spectra in the near-UV region are qualitatively and quantitatively similar, indicative of identical tertiary structure (**Fig 5.5.4ii**). These data further support the hypothesis that WWAV GP1 is structurally unaltered by pH.

The CD spectra of LORV GP1 at pH 8.0 and 5.0 display greater variability than WWAV GP1, exhibiting deviations in magnitude in the 190-199 nm and 203-222 nm regions, indicative of secondary structure changes (**Fig 5.5.4i**). Analysis of the total secondary structure content reveals an increase in helical content and loss of disorder associated with a transition from pH 8.0 to 5.0. In the absence of LASV GP1 protein for experimental comparison, theoretical CD spectra calculations were performed using the PDB2CD server (Mavridis and Janes, 2017) and secondary structure content was calculated based upon DSSP analyses (Kabsch and Sander,

1983), for the GP2-associated and GP2-free states of LASV, using the respective crystal structures (5VK2 & 4ZJF).

The experimental and theoretical spectra for LORV GP1 and LASV GP1 (respectively), both in the GP2-free state, though qualitatively similar, differ greatly in magnitude (**Fig 5.5.4i**). Given the similarity of the crystal structures, it is unlikely that such CD differences would be reflective of large-scale structural differences of LORV in solution, when compared with the crystalline environment. It is plausible that the theoretical CD spectra are inaccurate. Indeed, the spectra calculations rely on known experimental spectra (Mavridis and Janes, 2017), of which none exist for GP1s in the online CD data repository, the PCDDB. The inaccuracy is likely compounded by an absence of known structural homologues to the GP1 fold (Bowden et al., 2009b). Using the theoretical spectra as a qualitative comparison, it may be tentatively noted that the spectral changes associated with the transition of LORV GP1 from pH 8.0 to 5.0 are roughly opposite to those observed in the GP2-associated to GP2-free transition, supporting the hypothesis that, once formed, the GP-free conformational state does not revert to the GP2-associated state in a pH dependent manner.

A comparison of the measured total secondary structure content with that observed in the crystal structures of LASV GP1 was performed, to negate inaccuracies incurred when using the theoretical LASV spectra. For LASV GP1, the transition from the GP2-associated to -free state is characterised by a decrease and increase in α-helical and β-sheet content, respectively. The α-helical content of LORV at pH 8.0 closely resembles that of LASV in its GP2-free state, however, the greater total β-sheet content of LORV is likely due to the elongated β1 strand and β-hairpin motif therein (**Fig 5.5.4i**). These data support the crystallographic data, demonstrating that at pH 8.0 LORV GP1 adopts a GP2-free conformation that is unaltered by the crystallisation process. The secondary structure content measured by CD is in good agreement with that determined from the crystal structures: 29.2% sheet and 19.2% helical; and and 27.4% sheet and 19.8% helical, in LORV GP1 at pH 5.0 and 8.0, respectively.



D.

Scan	α-helix	β-strand	Turns	Unordered	Helix segments	Strand Segments	Helix Length	Strand Length	Standard Deviation
WWAV pH8	0.186	0.283	0.227	0.303	2.43	4.959	7.663	5.716	0.036
WWAV pH5	0.195	0.286	0.225	0.295	2.532	4.864	7.684	5.871	0.039
LORV pH8	0.205	0.293	0.211	0.290	2.527	5.382	8.124	5.454	0.026
LORV pH5	0.227	0.291	0.224	0.258	2.787	5.720	8.140	5.086	0.035

Figure 5.5.4i: Far-UV Circular dichroism analysis of WWAV and LORV GP1s. (A) Experimentally determined CD spectra obtained from WWAV and LORV GP1 both at pH 8.0 and pH 5.0. The theoretical CD spectra of LASV GP1 in its GP2-associated and -free states are shown in shades of grey, calculated using the PDB2CD server (Mavridis and Janes, 2017) utilising the structures of LASV GP1 from PDB entries 5VK2 and 4ZJF. (B) Percentage α-helical and β-sheet content calculated from the CD spectra are displayed for LORV and WWAV GP1. The α-helical and β-sheet content of LASV GP1 was calculated using DSSP (Kabsch and Sander, 1983). (C) Superposition of the β-hairpin in LASV and LORV, coloured according to A. (D) Table of CD derived secondary structure content calculations, values represent 'fraction of folding' contributed by each element.

Interestingly, while CD indicated that the β-sheet content appears essentially unaltered by pH, helical content subtly increases at pH 5.0 with a concomitant loss of unordered residues. The structural basis of this change is unclear and can not be rationalised using our crystal structures. As WWAV GP1 also exhibits an increase in helical content at pH 5.0, though lower in magnitude, it is possible that the vector derived residues that are not observed in the crystal structures, common to both recombinant proteins, may account for these observations.

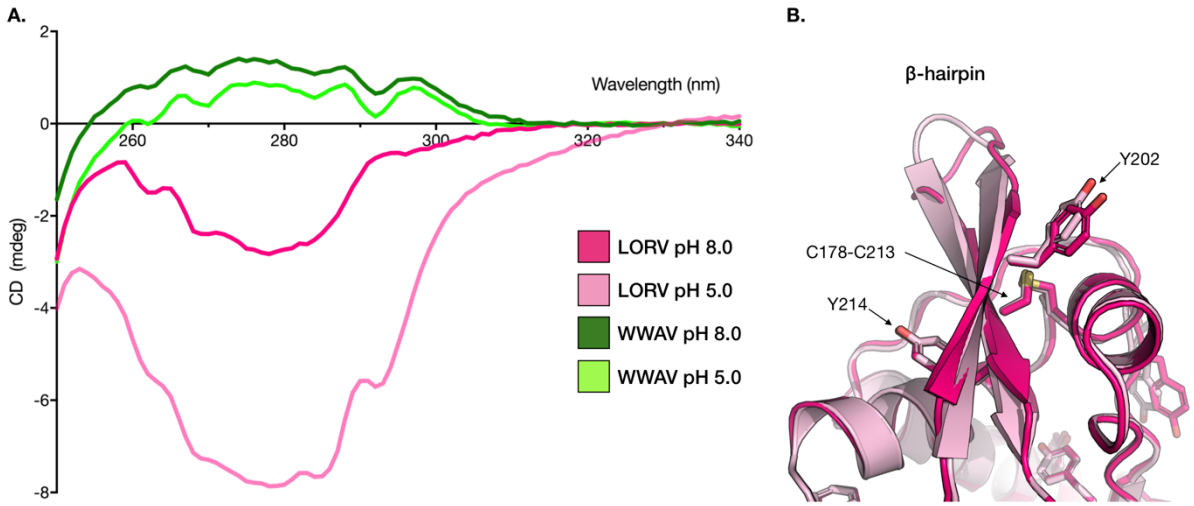


Figure 5.5.4ii: Near-UV circular dichroism analysis of WWAV and LORV GP1s. (A) Near-UV CD spectra of WWAV and LORV GP1, both at pH 8.0 and pH 5.0. (B) Superposition of the β-hairpin region of the LORV GP1 pH 5.0 and pH 8.0 structures, coloured according to A. Structures are displayed as cartoons with local optically active groups, tyrosine and disulphide bonds, shown as sticks, coloured according to constituent elements: nitrogen in red, sulphur in yellow and carbons pink.

Tertiary structure information is predominantly encoded within the near-UV region of the CD spectrum, and whilst quantitative analyses are not yet as reliable as for far-UV data, they nevertheless provide a qualitative structural 'fingerprint' of a protein (Kelly et al., 2005). In the near-UV region, the tertiary structure of WWAV GP1 appears unaltered by pH, consistent with the hypothesis that it does not change conformational states. LORV GP1, like WWAV, yielded qualitatively similar CD spectra, though with much larger differences in absolute CD intensity (**Fig 5.5.4ii**). Spectral differences in the 250-300 nm region are indicative of changes within the

local environment of aromatic side chains (F/Y/W), and differences at wavelengths greater than 310 nm result from dihedral angle change within disulphide bonds (Kelly et al., 2005). Variation within both regions of the spectrum engender the LORV GP1 different tertiary structure fingerprints at pH 8.0 and 5.0. The most structurally labile region within the crystal structures of LORV GP1 is the β-hairpin (**Fig 5.5.2i**), which therefore may also account for the observed variability of LORV GP1 in solution. The presence of two tyrosine residues and a disulphide bond within the β-hairpin may therefore account for the different tertiary structure fingerprints of LORV GP1 (**Fig 5.5.4ii**).

Taken together, the CD spectra reveal that isolated WWAV GP1 is essentially inert towards pH change, exhibiting little to no secondary or tertiary structure modification as a function of pH. Similarly, isolated LORV GP1 exhibited only minor secondary and tertiary structure changes at neutral and acidic pH, that are more likely to be a consequence of localised flexibility than conformational switching. CD is a low-resolution technique, and thus the nature of the observed changes is difficult to interpret. However, these data support our crystallographic investigation and indicate that LORV GP1 is incapable of reversibly adopting a GP2-associated state with a pH-induced mechanism. Interestingly, it is also plausible that the subtle increase in order observed upon acidification of LORV GP1, if a feature shared amongst other OW arenavirus GPIs, may account for the apparent propensity of OW GPIs to crystallise under acidic conditions.

5.6 Expression and purification of Lujo virus GP1

5.6.1 Rationale

To assess the extent to which the formation of a GP2-free state of OW arenaviral GPIs is conserved, we sought to determine the structure of the isolated GP1 from Lujo virus (LUJV), an emergent human pathogen (Briese et al., 2009). Phylogenetic analyses demonstrate that LUJV does not cluster with any know arenaviruses, though is most closely related to the prototypic

OW virus LCMV. Coincident with constituting a unique arenaviral lineage, LUJV entry has also been shown to be independent of both of the established arenaviral primary receptors, α-DG and TfR1 (Raaben et al., 2017). LUJV GP1 engages neuropilin-2 at the cell surface and requires an endosomal protein, the tetraspanin CD63, for cellular entry, in an apparently low-pH dependent receptor switch that is analogous to the LAMP1 receptor switching necessary for LASV entry. Taken together, the genetic divergence of LUJV, and the idiosyncratic receptor utilisation and switching, renders LUJV GP1 an intriguing target by which to assess the conservation of the GP2-free state across disparate OW GP1s. Indeed, the structure of LUJV GP1 may provide insight as to whether GP1 conformational class changes are a conserved feature necessary for receptor switching or possible immunomodulation.

5.6.2 Construct design, expression, purification, and crystallisation of LUJV GP1

Constructs of LUJV GP1 derived from a codon-optimised synthetic gene encoding the entire reading frame of the LUJV GPC (NCBI reference sequence: YP_002929490.1). Six constructs were generated utilising all possible combinations of three N-terminal start sites (L67, M74, and P81) and two C-terminal end sites (H199 and E209) (**Fig. 5.6.2**). Domain boundaries were devised based upon sequence alignment with MACV GP1, the only GP1 for which a structure was available at the time (PDB: 2WFO).

Small-scale transient expression testing in HEK293T cells revealed all constructs beginning with L67 failed to express, while those that did not were secreted, indicative of correct folding (**Fig. 5.6.2**). Of the secreted constructs, those that terminated with H199 were present in a greater quantity than constructs terminating at E209. Given the apparent equivalence in secreted expression of M74–H199 and P81–H199, the shorter construct (P81–H199), hereafter referred to as LUJV GP1, was selected for large-scale expression in order to minimise the presence of flexible regions that may be refractory to crystallisation.

Large scale expression of LUJV GP1 in HEK293T cells, in the presence of kifunensine, yielded milligram (per litre) quantities of highly pure protein (**Fig. 5.6.2**). LUVJ GP1 appeared to be monomeric and monodisperse in solution, as assessed by SEC. Deglycosylation with endoF_1 yielded an appreciable peak shift when assessed by SEC, which was confirmed by SDS-PAGE to be due to a reduction in the number of sugar moieties present on the protein (**Fig. 5.6.2**). Consistent with a predicted proteinaceous mass of 15.3 kDa, SDS-PAGE analysis of 'glycosylated' LUJV GP1 revealed five distinct species ranging in mass from approximately 15 to 35 kDa, indicating the presence of at least four occupied glycan sites (of 5 predicted). Intriguingly, the presence of a species at around 15 kDa, if comprised of un-glycosylated LUJV GP1, may indicate that glycosylation is dispensable for correct folding of the protein. Alternatively, the lowest molecular weight species may be a singly glycosylated glycoform, that would be expected to have a molecular mass of approximately 17 kDa. Enzymatic deglycosylation successfully reduced the sample compositional heterogeneity, reducing the number of observed species to three, with a predominant species at around 20 kDa, that likely represented GP1 presenting a single high-mannose site that is resistant to deglycosylation, under the conditions tested. The resultant sample was highly pure and composed primarily of one distinct glycoform, and as such was deemed suitable for crystallisation experiments.

Purified, deglycosylated LUJV GP1 was subjected to crystallisation trials. Over 3,400 distinct crystallisation conditions were screened, including a range of protein concentrations between 12.4 and 4.1 mg/ml and temperature (25°C and 4°C). In the absence of observable crystals, *in situ* proteolysis with 1:1000 (by mass) trypsin was also performed. In all instances no crystalline growth was observed, over a time period of at least one year. The apparent resistance to crystallisation, despite suitable purity and solution behavior, suggests further construct optimisation, or stabilisation by complexation with a receptor, is warranted.

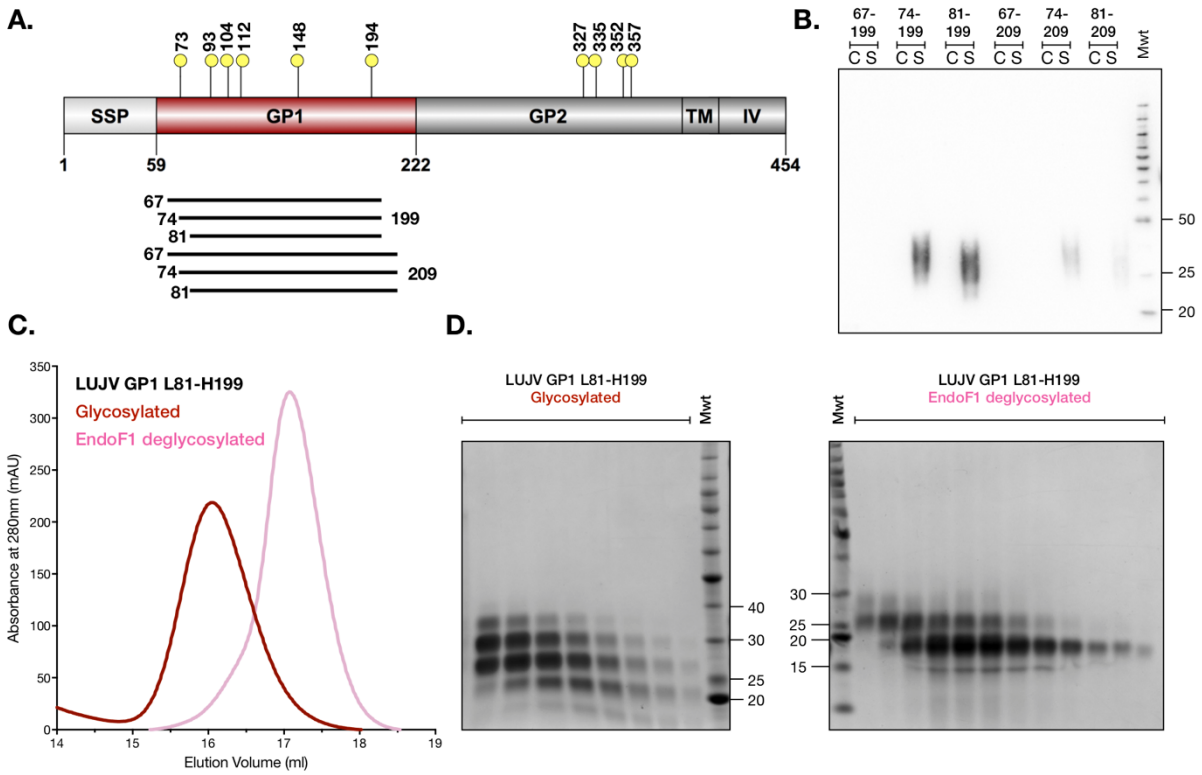


Figure 5.6.2: Construct design, expression and purification of LUJV GP1. (A) Domain organisation of the GPC gene from LORV. The stable signal peptide (SSP), GP1, GP2, transmembrane (TM) and intra-virion domains are annotated. Putative N-linked glycosylation sites are illustrated with pins. Cloned construct regions are illustrated as bars below the sequence. (B) Western-blot analysis of transiently transfected LUJV GP1 constructs indicated in A. Cellular (C) and supernatant (S) fractions were subject to reducing SDS-PAGE analysis and subsequent western blotting with an anti-his tag probe. (C) Size exclusion chromatograms of kifunensine treated LUJV GP1 (L81-H199) prior to (red) and after enzymatic deglycosylation with endoF1 (pink). (D) Reducing SDS-PAGE analysis of the SEC purifications shown in C. Representative fractions spanning the peaks in C are shown. Molecular weight standards are labelled 'Mwt' with selected bands labelled with the approximate masses, in kDa.

5.7 Discussion

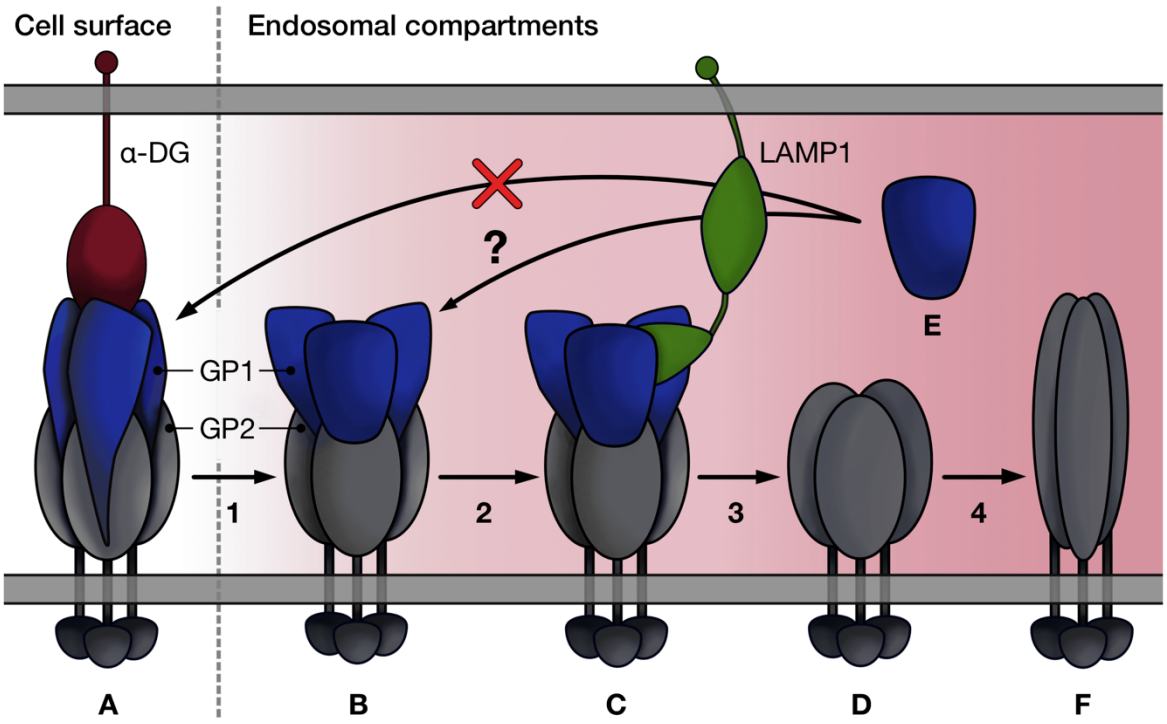
Here, we exploit structural, biophysical and biochemical methods to elucidate the structural and functional states of the arenaviral GPC during host-cell entry. These data delineate the discrete structures adopted by the dynamic and multi-functional GPC, therein providing a template for guiding the rational development of therapeutic interventions and vaccination strategies.

As the sole protein antigen presented on the arenaviral surface, GPC is responsible for host-cell attachment and membrane fusion, and therefore must couple the sensation of

environmental stimuli with structural rearrangements within the protein. Conformational rearrangements are critical to the function of fusion proteins, and thus are commonly associated with fusogens, such as GP2, for which structures have been extensively studied (Stegmann, 2000). Attachment glycoproteins, on the other hand, do not paradigmatically undergo substantial structural rearrangements. Here, we delineate discrete states of the arenavirus GP1, both in isolation and as part of mature GPC complexes, that have distinct structural and functional properties. Furthermore, these data permit the construction of an integrated model for the differential cellular entry processes of arenaviruses (**Fig 5.7i**).

The neutral pH crystal structure of the LASV GPC was recently solved (Hastie et al., 2017), which provides high resolution validation of the of the cryo-electron tomography reconstructions presented in section 5.4.4 (Li et al., 2016). These data reveal GP1 to present an α-DG competent binding surface at the apex of the spike complex (Acciani et al., 2017). Following primary receptor engagement and subsequent internalisation, the increasingly acidic environment within the endocytic pathway stimulates α-DG dissociation and a series of structural transitions within GPC.

Upon acidification GPC appears to adopt an 'open' conformation wherein the inter-protomer interface in the vicinity of GP1 forms a crevice, and a loss or redistribution of density within the membrane-proximal region of the spike is observed (Li et al., 2016). While high resolution data do not exist to support a detailed analysis of this state, functional characteristics of the acidified GPC and isolated GP1 sub-components suggest functional, and therefore structural similarity. Both LASV GP1 in the GP2-free state and the acidified GPC are able to bind the intracellular receptor LAMP1, usually encountered in the endocytic compartments during cellular entry, but not α-DG (**Fig 5.4.3i & 5.4.4**) (Jae et al., 2014, Cohen-Dvashi et al., 2015, Li et al., 2016).



Transitions

1. Internalisation & acidification to < pH 6
2. LAMP1 binding at < pH 6
3. GP1 detachment at < pH 4
4. Fusogenic rearrangements & membrane fusion

Structural states

A. Neutral pH complex
B. Acidified complex
C. LAMP1-bound
D. SSP-GP2 only
E. Detached GP1 in GP2-free state
F. Post-fusion hairpin conformation

Figure 5.7i: An integrated model for LASV GPC-mediated entry. A schematic representation of an integrated model for LASV entry derived from cryo-electron microscopy, crystallography and biochemical studies. The arenaviral GP1 and GP2 subunits of the trimeric arenaviral GPC are coloured blue and grey, respectively. SSP is not shown due to a lack of structural data. At neutral pH (>pH 6) LASV GP1 engages the cell-surface receptor α-dystroglycan (α-DG, red) at the apex of the spike (A) and is subsequently internalised. Acidification of GPC (<pH 6) causes an opening of the GP1 protomer interface at the apex of the spike, as well as a loss of mass from the membrane proximal region, here attributed to reorientation of the GP1 N-terminus (state B). The acidified complex (B) is LAMP1 (green) binding competent and engages the membrane distal β-prism domain at the newly formed inter-protomer crevices, in an interaction that is not critically dependent upon terminal sialic acid (C). At pH values of below four, GP1 detaches leaving a trimeric SSP-GP2 complex (D), presumably primed for membrane fusion. Detached GP1 likely forms the GP2-free state (E) observed in all known OW GP1 structures. The GP2-free state of the OW GP1 can not reform the GP2-associated state observed in (A) with a simple pH-switch. The structural equivalence of the GP2-free state of GP1 (E) and the acidified GP1 as part of GPC (B & C) is undetermined, however subtle biochemical differences exist. The residual SSP-GP2 complex mediates membrane fusion, culminating in the canonical class I post-fusion hairpin conformation (F). In all instances, receptor binding stoichiometry is undefined and thus one receptor is shown for clarity. It is unknown whether other OW viruses that do not depend upon LAMP1 for entry utilise a similar mechanism. The conservation of state E in OW arenaviruses suggests such viruses likely adopt states A, B, D, E and F. NW viruses, which are LAMP1 independent, and do not possess a GP2-free state of GP1 (E), likely utilise only states A, D and F (wherein detached GP1 resembles GP1 in state A).

Despite functional similarities that suggest a structural equivalence between the GP2-free state of the OW LASV GP1 and GP1 as part of the acidified spike complex—in which structural transition result in an 'open' LAMP1-binding competent state— both exhibit subtly different receptor-binding properties. Contrary to previous studies that show isolated LASV GP1 fails to bind kifunensine-treated LAMP1 (Israeli et al., 2017), cryo-electron microscopy reconstructions and ELISAs demonstrate that acidified LASV VLPs, containing authentic trimeric GPC, bind LAMP1 (**Figs 5.4.3i & 5.4.4**). Given the localisation of the LAMP1 binding site to the pH-induced inter-protomer crevices, it is possible that LAMP1 binding involves quaternary interactions that bridge two GP1 molecules. Whilst mutagenesis has shown the β-hairpin of LASV to be crucial to LAMP1 binding, and sufficient to confer LAMP1 binding upon a LAMP1-independent virus (Israeli et al., 2017), these studies utilised isolated or Fc-tagged proteins that would not recapitulate the quaternary architecture of GPC. It is therefore plausible that the different LAMP1 binding properties of GP1 are a consequence of the presence or absence of quaternary binding sites in these studies. Indeed, a quaternary binding site, afforded by the more biologically relevant VLP, may provide sufficiently increased binding affinity to permit a detectable interaction between non-sialylated LAMP1 (**Fig 5.4.3i**) and GP1, that is otherwise undetectable using isolated GP1s. Alternatively, the GP2-free state of LASV GP1 and GP1 in the acidified GPC may not be structurally identical, though both LAMP1-binding competent, suggesting that detachment from GP2 may further modulate the structure of GP1. Higher resolution structural studies of the acidified GPC will be required to confirm the structural equivalence of GP1 in the acidified complex and in the detached GP2-free state.

Interestingly, site-specific glycan analysis suggests that that LAMP1 is not fully sialylated at N76 (**Fig 5.4.3ii**). Recombinant LAMP1 utilised in kinetic binding studies with isolated LASV GP1 was also produced in HEK293T cells and therefore likely possessed a highly similar glycan profile to that determined herein (Israeli et al., 2017). If sialylation is critical to the interaction of LAMP1 with isolated LASV GP1, the calculated interaction kinetics therefore do not accurately

account for approximately 70% of the analyte being binding incompetent. Alternatively, if, as our data suggest, sialic acid is not critical, a variety of LAMP1 glycoforms may interact, possibly with varying kinetics. The discrepancies pertaining to the glycan dependence of the LAMP1 interaction therefore require further experimental clarification.

Based upon mutagenic mapping of the LAMP1 binding site and a fitting of the GP2-free state of LASV GP1 into the reconstruction of the acidified GPC (Li et al., 2016), Israeli et al. (2017) proposed a model for LAMP1 binding at the trimer apex, distal to the observed density attributed to LAMP1, which the authors suggest to be an unidentified component (**Fig 5.7ii**). As a feature that appeared after the addition of highly-pure recombinant LAMP1, the volume attributed thereto is likely a more reliable feature than the GP1 fitting used rationalise the newly proposed binding mode (**Figs 5.4.4 & 5.7ii A**). As such, GP2-free LASV GP1 is likely wrongly or poorly fitted or does not accurately reflect the conformation of GP1 present in acidified GPC. Indeed, attempts at generating a new fitting that rationalise both structural and mutagenic data failed to yield any plausible solutions.

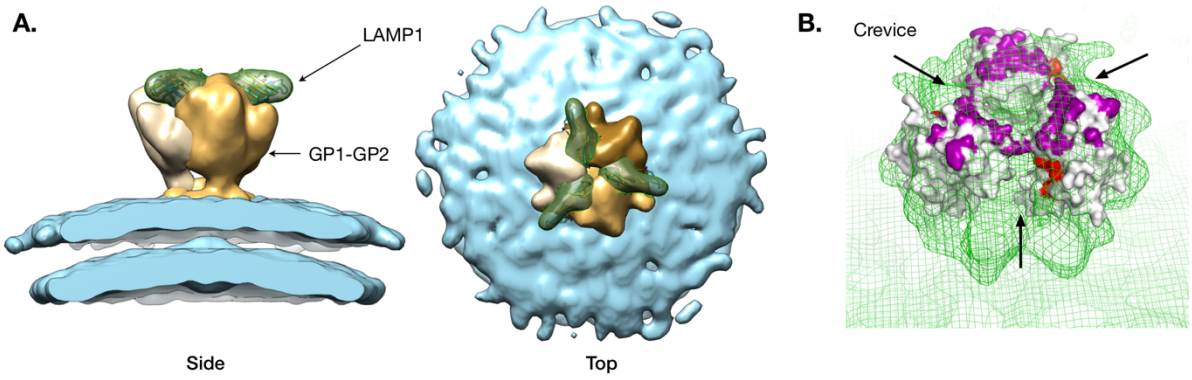


Figure 5.7ii: Rationalising cryo-EM and mutagenic data to define the LAMP1 binding site. (A) Cryo-EM reconstruction of LASV GPC at pH 5 in the presence of LAMP1. The structure of the β-prism domain of LAMP3 (PDB 4AKM) is fitted into the density attributed to its homologue, LAMP1. GP1-GP2 protomer volumes are coloured shades of brown, putative LAMP1 density in green, and the membrane in blue. (B) A model of the acidified GPC utilising GP2-free LASV GP1 (white surface) fitted into the EM density of the acidified spike (green mesh), taken from Israeli et al. (2017). The histidine triad and residues that impact LAMP1 binding are coloured red and purple, respectively. Inter-protomer crevices are shown with arrows. The proposed LAMP1 binding surface appears to decorate the top of the spike complex, not the crevices, as suggested by panel A. Panel A and B are not to scale.

The crystal structure of the isolated LASV GP1, and the GP1 from the OW virus MORV, reveal a conformation distinct from that presented by the neutral pH LASV GPC (Cohen-Dvashi et al., 2015, Hastie et al., 2017). To assess the pH-dependence of the structure of isolated GP1, we performed comparative crystallographic and CD analyses of the GP1 glycoprotein from the closely related OW LORV, at pH 8.0 and 5.0. Isolated LORV GP1 forms the GP2-free state observed in all other isolated OW GP1s solved to date (Cohen-Dvashi et al., 2015, Israeli et al., 2017) and is conformationally independent of ambient pH change. Intriguingly, the GP2-free state is conserved and highly similar (RMSD <0.9 Å) across the OW GP1s of known structure, independent of a requirement for LAMP1-binding. The conservation of the GP2-free conformation of GP1 is therefore suggestive of functionality that in not linked to LAMP1 utilisation.

The crystal structures of the NW WWAV GP1, solved at neutral and acidic pH, reveal that isolated NW GP1, like their OW counterparts, are conformationally unaltered by pH. Intriguingly, however, the conformation adopted by NW GP1s, unlike OW GP1s, likely reflects the protein as it is displayed on mature virions. Indeed, NW GP1s have been shown to bind and be structurally unaltered by both monoclonal antibodies and the cellular receptor TfR1 (Abraham et al., 2010, Mahmutovic et al., 2015, Zeltina et al., 2017, Clark et al., 2018). The GP1s of NW arenaviruses are therefore unlikely to possess the conformational multiplicity of OW GP1s. As such, it seems plausible that the NW GPC does not possess a distinct 'acidified' conformation, at least within GP1, that is seen in the OW LASV (**Fig 5.7i**). While GP1 detachment during cellular entry is yet to be demonstrated for NW viruses, it is likely a conserved requirement necessary to permit fusogenic rearrangement of GP2.

It remains undetermined whether ectopically expressed GP1s, having never encountered a cognate GP2, are equivalent to *bona fide* detached GP1 expressed and folded alongside GP2 as part of the GPC polyprotein, prior to pH mediated detachment. However, given the striking structural conservation of the OW GP2-free state, the structural transitions identified within

acidified LASV VLPs (Li et al., 2016), and the observation that analogous NW GPIs constructs adopt a primary receptor-competent state likely presented on mature virions (GP2-associated state), it is likely that the GP2-free state is of biological importance.

Given the apparent dichotomy between OW and NW viruses in their ability to adopt a distinct GP2-free state, determining the factors that dictate whether a given arenaviral GP exhibits a GP2-free state are of interest. Understanding the predictors of GP2-free state formation may enable elucidation of biological function. Indeed, all GPIs solved to date use either TfR1 or α-DG as primary receptors and concomitantly bifurcate into GP2-free incapable and capable groups, respectively. The association between state switching and receptor utilisation however, is undetermined. Of interest in this regard are the Clade C NW viruses, Latino and Oliveros virus, that utilise α-DG (Spiropoulou et al., 2002) and the OW LUJV that utilises NRP-2 for cellular entry. Determination of the classes adopted by viral GPIs that are genetically diverse and display idiosyncratic receptor utilisation is warranted to probe the determinants and functional implications of GP1 conformational duality.

The GP2-free state of GP1, formed as a consequence of acidification-mediated detachment or dissociation from GP2, and active shedding, by as yet undetermined means (Branco and Garry, 2009, Branco et al., 2010) may, like in the case of Ebola virus (Dolnik et al., 2004), have immunomodulatory functions. Indeed, the pronounced conformational difference between the GP2-associated and GP2-free states of the OW GP1 likely render them antigenically distinct. The sheet face of GP1 , which is the most structurally conserved region in both conformational states of OW GPIs, is largely occluded by a dense glycan shield (Watanabe et al., 2018) and the reorientated N- and C-termini (**Fig. 5.3ii**). Furthermore, the comparatively 'unprotected' helix face undergoes large-scale structural rearrangements that, in combination with the aforementioned features of the sheet-face, likely render antibodies raised against the GP2-free state ineffective at binding authentic virus comprising GP1 in its GP2-associated state. Presentation of dramatically different epitopes by virion-displayed (GP2-attached) and shed

(GP2-free) GP1 may contribute to the lack of neutralising antibodies observed during LASV infection (Robinson et al., 2016).

The *Arenaviridae* contains a number of known lethal human pathogens that are both endemic, in the OW and NW, and cause sporadic outbreaks of hemorrhagic fever. Given a paucity of therapeutic interventions and lack of a licensed vaccine, there is an unmet need for improved arenaviral intervention strategies. Understanding the distinct structural states of arenaviral glycoproteins is of critical importance for the design of immunogens that elicit antibodies capable of targeting the desired states of the arenaviral GPC, namely the mature neutral pH conformation displayed on mature virions, or drugs that target important intermediates.

5.8 Conclusion

Defining features of the multi-functional arenaviral GPC and its GP1 sub-component, that differentially adopt discrete conformations, is critical to understanding the mechanisms of arenaviral entry and infection. Moreover, an appreciation of the varied structural classes, the triggers for their formation, and their functional properties, is crucial for informing therapeutic intervention and vaccination strategies. Indeed, we propose that subunit-based vaccination strategies utilising isolated GP1 proteins from OW viruses will likely be ineffective at generating neutralising antibodies capable of targeting mature virions.

Chapter 6:

Future perspectives and concluding remarks

Chapter 6: Future perspectives and concluding remarks

6.1 Future perspectives

6.1.1 Structural studies of viral glycoprotein-receptor complexes

X-ray crystallography has, over many decades, provided a wealth of high-resolution information regarding the nature of viral attachment glycoproteins and their interactions with host-cell receptors (Jolly and Sattentau, 2013). Indeed, the crystal structures of the Machupo virus (MACV) and the prototypic henipavirus—Hendra (HeV) and Nipah (NiV)—attachment glycoproteins in complex with their cellular entry receptors (Bowden et al., 2008a, Abraham et al., 2010), provided crucial information that both confirmed earlier cellular and biochemical data, and prompted further studies of the determinants of attachment-mediated spillover into human populations and the basis of antibody-mediated neutralisation (Lee et al., 2015, Zeltina et al., 2017, Clark et al., 2018).

The crystal structure of the isolated receptor-binding domain of the Cedar virus (CedPV) G protein provides a rationale for the viruses' unique ephrin-B1 tropism (**Chapter 3**). Identification of a tyrosine ($Y525^{CedPV}$), unique amongst known henipaviruses, in the putative receptor binding region, highlights the utility of structure-based appraisal of receptor binding surfaces (**Fig 3.7.6ii**). Notably, the conformation of $Y525^{CedPV}$ appears uniquely modulated by the tertiary architecture of the protein fold and thus its conformation would be almost impossibly difficult to predict without experimental structural data. Indeed, homology models of CedPV G, based upon available henipaviral G proteins, generated to test this hypothesis, do not recapitulate the tertiary architecture of this region of the protein. To this end, the crystal structure of CedPV G in complex with ephrin-B1, and to a lesser extent ephrin-B2, will be critical in determining the importance of $Y525^{CedPV}$ and to further extend our knowledge of henipavirus G-receptor interactions.

The crystal structure of the GP1 from Whitewater Arroyo virus (WWAV) revealed substantial structural diversity amongst transferrin receptor-1 (TfR1) utilising arenaviruses (**Chapter 4**). Given the structural diversity of GP1 glycoproteins, determination of the interactions between such GPIs and their cognate rodent host and human receptors will be critical for understanding the various modes of TfR1 recognition.

The nature of the interaction between the Lassa virus (LASV) GPC and the intracellular receptor, lysosomal associate glycoprotein-1 (LAMP1), remains incompletely understood. Indeed, data presented herein and elsewhere fail to form a unified description of the LASV GPC-LAMP1 interaction (**Chapter 5**). The only structural data for the LASV GPC-LAMP1 interaction are low-resolution (14.8 Å), and whilst informative, are of limited application in confirming functional hypotheses (Li et al., 2016). As such, a high-resolution structure of the LASV GPC-LAMP1 complex would be of great value. Whilst the recent high-resolution structure of the trimeric LASV GPC ectodomain was solved by X-ray crystallography, generation of stable and trimeric protein required protein engineering and the presence of a trimer-stabilising antibody fragment that binds a quaternary protomer-bridging epitope (Hastie et al., 2017). Intriguingly, this neutralising antibody (37.7H), which was crucial to crystallogenesis, inhibits LAMP1 binding through an allosteric mechanism (Hastie et al., 2017). The mutually exclusive binding of LAMP1 and 37.7H therefore limits its utility as a crystallisation chaperone for LASV GPC-LAMP1 complexes (Koide, 2009).

Given that GPC is highly glycosylated, metastable, and conformationally labile, and that the LAMP1 interaction may be sub-stoichiometric (Li et al., 2016), crystallographic study is likely to be exceedingly challenging. Therefore, utilisation of cryo-electron microscopy (either single-particle or tomography with sub-tomogram averaging), may be most appropriate for generation of high-resolution structural information, particularly due to its ability to deal with conformationally and compositionally heterogeneous samples through *in silico* classification and data processing. Furthermore, such studies may yield structural information regarding the

nature of SSP and its interactions with the rest of the complex, that are critical to the fusion functionality of GP2.

Initial attempts to obtain cryo-electron microscopy-derived structures of full-length LASV GPC were performed. Tripartite LASV GPC, comprising the entirety of SSP-GP1-GP2, was kindly supplied Prof. Jack Nurnberg (University of Montana, USA), as described previously (Thomas et al., 2012). Despite initial observations evidencing trimeric GPC, similar to that observed on authentic fixed virus (**Fig 6.1.1**), data processing failed to yield sufficiently good quality 2D classes to permit further processing. Optimisation utilising LASV GPC-specific antibody Fab fragments, to aid both alignment and improve stability, and detergent substitutes (e.g. amphipols (Flotenmeyer et al., 2007)), are ongoing.

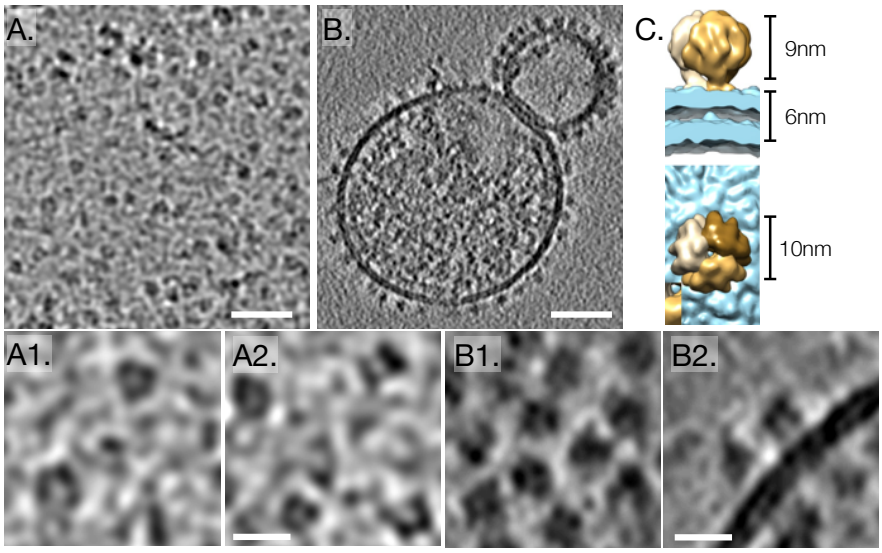


Figure 6.1.1: Towards high-resolution cryo-electron microscopy analysis of LASV GPC. Comparison of recombinant LASV GPC in the context of detergent micelles or virion resident lipids. (A.) Micrograph of DDM solubilised LASV GPC. A1 and A2 show enlarged views exhibiting potential top and side views, respectively. (B.) A slice through a 3D tomogram showing two LASV particles. B1 and B2 show enlarged top and side views, respectively. Scale bars represent 50nm in A and B and 10nm in A1, A2, B1 & B2. (C.) LASV GPC spike complex derived from sub-tomogram averaging with approximate dimensions displayed (Li et al., 2016).

6.1.2 Towards a method for high-throughput identification of antibody and receptor footprints

The generation of protein structures by X-ray crystallography is fundamentally limited by the ability to generate crystals with desirable diffraction properties (Price et al., 2009). Formation

of such high-quality protein crystals is a limiting factor in many structural projects, herein demonstrated by the failure to form suitable crystals for LUJV GP1, WWAV GP1-naTfR1, CedPV G-ephrin-B1, and CedPV G-ephrin-B2, despite the ability to produce milligram quantities of high-purity, monodisperse protein. Techniques that are able to rapidly generate structural information utilising un-labelled samples, optimised for, but apparently refectory to crystallisation, are of great value.

X-ray footprinting mass spectrometry (XF-MS) is an emerging technique that has utility in defining solution-state protein-ligand interactions and dynamic conformational changes (Gupta et al., 2016b). XF-MS utilises synchrotron-derived X-rays to stimulate the radiolysis of water molecules within an aqueous protein solution (Gupta et al., 2016a). Highly reactive hydroxyl radicals produced as a consequence of radiolysis rapidly and stably modify solvent-accessible amino acid side chains, therein generating adducts with an associated specific mass increase, that can be identified by mass spectrometry (MS). Comparative mapping of the solvent accessibility of both individual proteins and their complexes allows the determination of solvent excluded interfaces that are formed upon complexation, thus elucidating protein binding sites. XF-MS has the ability to elucidate protein-protein interaction interfaces with amino-acid level precision, utilising label free protein samples and in a relatively high-throughput manner (Kiselar and Chance, 2010). XF-MS is advantageous over similar methods, such as hydrogen-deuterium exchange mass spectrometry (HDX-MS), in that the observed modifications are covalent and temporally stable, and samples can be analysed in a number of common buffers that more faithfully reflect physiological conditions (Marcsisin and Engen, 2010).

In an attempt to experimentally validate the hypothesis that CedPV G recognises ephrin-B2 with an overall binding-mode that is conserved amongst the henipaviruses (**Chapter 3**), and to further assess the potential of the technique for future receptor and antibody footprint identification, XF-MS of CedPV G-ephrinB2 complexes was performed (**Chapter 2 section 2.7**). MS identified a number of modified amino acids, suggestive of successful sample irradiation

and suitable data analysis. Furthermore, modified amino acids mapped exclusively to solvent accessible regions of the protein consistent with expectation, thus additionally validating the data. The overall MS coverage was relatively low (<60% of the total protein), and the number of detected adducts even lower (<20%). These data do not provide convincing evidence of ephrin-B2-binding-mediated protection from oxidative modification (**Fig 6.1.3**).

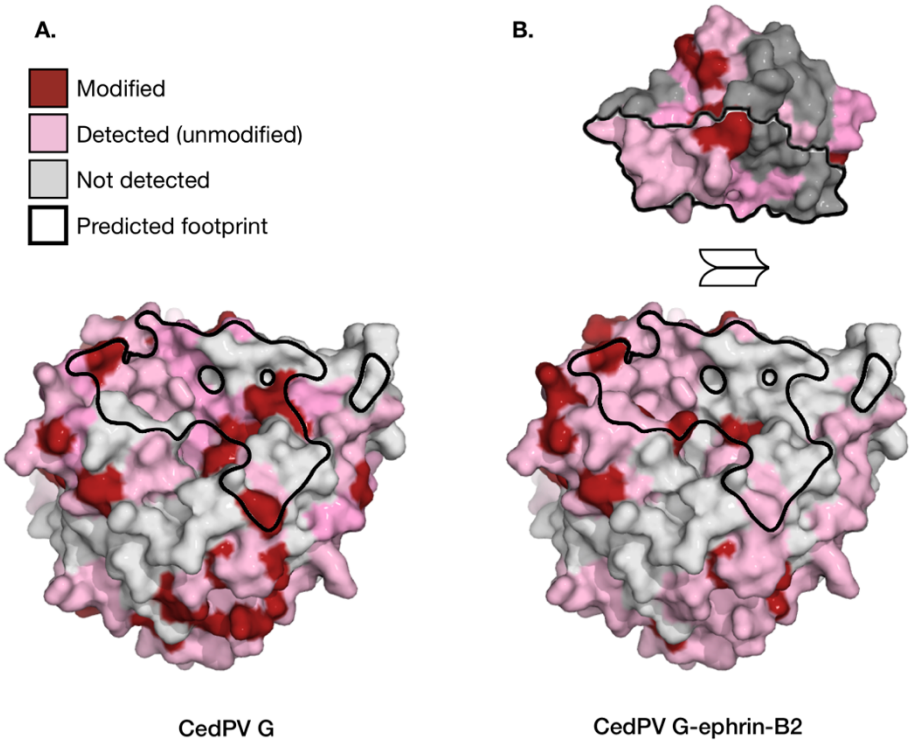


Figure 6.1.3: X-ray footprinting analysis of the CedPV G-ephrin-B2 interface. (A) Mapping oxidative modifications onto unliganded CedPV G. The crystal structure of CedPV G is shown in surface representation, coloured according to mass-spectrometry (MS) analysis: unmodified residues detected by MS are coloured pink; detected residues with oxidative modifications are coloured red; and undetected residues are grey. The predicted ephrin-B2 binding surface is outlined in black. (B) Mapping oxidative modifications to the CedPV G-ephrin-B2 complex. Modifications are shown as in panel A. Modifications of ephrin-B2 are mapped onto the crystal structure derived from PDB 2VSM.

Modifications within the putative receptor binding site of the ephrin-B2-bound complex can be rationalised by partial solvent accessibility within the complex. It is plausible that the data analysis strategy employed may require further optimisation, specifically, the current approach may be removing valuable information due to a failure to account for all potential oxidative

modifications. As such, the data analysis approach employed herein requires optimisation. Furthermore, peptide fragments corresponding to only one of seven N-linked glycosylation sequons were observed in the MS data, which may suggest hardware limitations that result in systematic underrepresentation of glycopeptides.

While the receptor footprints for both WWAV GP1 and CedPV G are unlikely to differ substantially from those predicted by comparison with homologous structures (Abraham et al., 2010, Bowden et al., 2008a), XF-MS analysis has the potential to provide solution-state experimental validation of this hypothesis. Furthermore, XF-MS may have even greater utility in defining novel receptor interactions or providing a high-throughput means of defining antibody epitopes without the need for crystallisation. Crystallographic studies of antibody-antigen complexes are prevalent (Abraham et al., 2010, Zeltina et al., 2017, Clark et al., 2018), though structural information regarding the majority of the bound antibody fragments and viral proteins, outside of the epitope-paratope interaction interface—where not the first instance of their publication—are often of limited interest. Thus, high-throughput mapping of protein epitopes by XF-MS, in cases where the structure of the viral protein is already known, has the potential to provide time- and cost-effective identification of protein interfaces. Nevertheless, for XF-MS to truly become a prominent technique for the structural characterisation of protein-protein interactions, future optimisation, both in terms of increased automation in sample irradiation procedures and accessibility of MS analyses for non-expert users, is essential.

6.1.3 Biochemical and cellular validation of the determinants of receptor specificity

Whilst structural data often provide unequivocal molecular-level detail of individual proteins and their ligand-/receptor-bound complexes, experimental interrogation of structure-guided hypotheses, *in vitro* and *in vivo,* is critically important. Indeed, the biological complexity of virus-host interactions extends beyond static view afforded by crystallography, and thus more complex systems are required to recapitulate more representative biological environments. To

this end, we identify a number of residues that are likely involved in determining the differential receptor tropism characteristics of both WWAV and CedPV, that warrant such further investigation.

Ongoing collaboration with the laboratory of Prof. Benhur Lee (Icahn School of Medicine at Mount Sinai, USA), that has developed a CedPV reverse genetics system, provides a platform to assess the effects of structure-guided mutants on the cellular receptor binding and entry of CedPV.

There is increasing evidence of the spillover of clade D arenaviruses, specifically WWAV, into human populations in North America (Enserink, 2000, Milazzo et al., 2011). To this end, it has been shown that only minor changes to hTfR1 are sufficient to permit its utilisation by clade D arenaviruses (Zong et al., 2014); Tamiami virus (TAMV) appears to rapidly adapt to hTfR1-mediated cellular entry (**Chapter 4**); and hTfR1-binding strains of WWAV are naturally occurring (Milazzo et al., 2008). Taken together, these observations suggest that clade D viruses represent a reservoir with the adaptive capacity to become human-infective. Owing to the complexity of protein tertiary structure formation, structure-based prediction of the determinants of acquired hTfR1 usage is difficult, as highlighted by the diversity within the TfR1 binding scaffold of arenaviral GPls (**Chapter 4**). Nonetheless, changes that confer hTfR1 usage are likely to be modest (Zong et al., 2014). Indeed, the rapid adaption of TAMV to hTfR1 usage elegantly demonstrates the power of virus- and cell-based methods in resolving such crucial determinants of hTfR1 tropism, that would be difficult to predict from structure alone. Intriguingly, variants observed within the TAMV quasi-species—that also exhibit shared features with the hTfR1-tropic WWAV strain—are located throughout the protein, and thus would unlikely be targeted based upon structure-guided mutations alone. Nevertheless, a combination of cellular and structural data will prove crucial in determining and rationalising the molecular facets of varied GP1-TfR1 interactions.

6.2 Concluding remarks

The interactions of a virus and its host are numerous and complex. Whilst no single method can shed light on the plethora of virus-host interactions, integration of reductive strategies has yielded profound insight into general viral biology and pathogenesis. Amongst these, X-ray crystallography has proved invaluable in elucidating many aspects of virology, particularly those concerning cellular attachment and entry (Jolly and Sattentau, 2013). Structural studies of viral attachment proteins has enabled: enhanced understanding of basic biology; the discovery and development of anti-viral drugs (Zhao et al., 2016); improvement of diagnostic tools (Banerjee and Mukhopadhyay, 2016); and enhanced structural-vaccinology efforts (Grimm and Ackerman, 2013).

Here, through crystallographic study of the attachment glycoproteins of two non-pathogenic viruses, each members of genera that contain important human pathogens with BSL-4 containment status, we provide a molecular level rationale for the different receptor tropism characteristics therein. Furthermore, by delineating the ability of these viruses to utilise a given repertoire of cellular entry receptors, we provide a structural template for understanding the interplay between receptor tropism, zoonosis, and pathogenesis. The idiosyncrasies observed amongst viral genera is reflective of their diversity and thus highlights the importance of basic science in understanding the expansive and varied virosphere. For example, while CedPV is able to utilise human receptors for functional cellular entry (Marsh et al., 2012, Laing et al., 2018), thereby underscoring potential for zoonotic spillover into human populations, receptor tropism alone is not able to account for its apparent lack of pathogenicity. Indeed, the ostensible lack of RNA editing and therefore non-structural proteins capable of antagonising the host immune response, likely contributes to CedPVs non-pathogenic phenotype (Marsh et al., 2012). Nevertheless, the structural principles determining receptor tropism, outlined herein, provide valuable insight that may be applied to as yet undiscovered henipaviruses with pathogenic potential. Conversely, the use of the human orthologue of the transferrin receptor-

1 (TfR1) by NW arenaviruses appears to be a correlate of pathogenicity (Choe et al., 2011). However, use of auxiliary receptors by NW viruses means that hTfR1 usage does not unequivocally determine zoonotic potential at the stage of host-cell entry (Flanagan et al., 2008). Ultimately, this work reveals that while conserved receptor-binding architectures can confer human cell tropism and enable zoonotic transfer, the ability for an emerging virus to specifically recognise a human host receptor is not always the sole pre-requisite for pathogenicity. Thus, the contrasting themes observed for each genus studied herein highlights the diversity amongst RNA viral pathogens and underscores the importance of continued study of such agents, which present a significant global health burden.

Modern technologies that can rapidly generate immense quantities of nucleotide sequence data have transformed the field of genomics (Margulies et al., 2005). Whilst the application of such high-throughput sequencing technologies to surveying viral diversity has greatly expanded appreciation of the virosphere, the resultant genome sequences have principally been applied to phylogenetic and epidemiological classification (Suttle, 2005, Li et al., 2015). Though fundamental to the understanding of viral evolution, such analyses give little consideration to the clinical importance of novel viral species. The paucity of genome-based assessments of clinical aspects of a given virus is principally a consequence of the complexity of virus-host interactions (Holmes, 2009). Indeed, even for well characterised systems such as the henipa- and arenaviruses discussed herein, many aspects of zoonosis and pathobiology are incompletely understood. Thus, the wealth of diverse viral genomic data already available, and the relative ease with which it can now be generated, represents a vast and underutilised resource that could provide crucial insight to predicting and preventing the emergence of as yet undefined human pathogens.

Targeted structural studies of the receptor tropism characteristics—like those undertaken herein—and determination of the pathobiological implications of such receptor utilisation, may provide a foundation for predictive tools that can assess zoonotic and

pathogenic potential of novel viruses, based on sequence alone. Continued study of novel attachment glycoprotein architectures, and the modes by which they mediate receptor recognition, is of critical importance to defining generalised principles for viral attachment within families.

Appendix A:

List of publications

Appendix A: List of publications

Publications arising from work undertaken during the course of this DPhil:

Li, S., Sun, Z., **Pryce, R**., Parsy, M. L., Fehling, S. K., Schlie, K., Siebert, C. A., Garten, W., Bowden, T. A., Strecker, T. & Huiskonen, J. T. 2016. Acidic pH-Induced Conformations and LAMP1 Binding of the Lassa Virus Glycoprotein Spike. *PLoS Pathog,* 12, e1005418.

Lu, J., Raghwani, J., **Pryce, R**., Bowden, T. A., Theze, J., Huang, S., Song, Y., Zou, L., Liang, L., Bai, R., Jing, Y., Zhou, P., Kang, M., Yi, L., Wu, J., Pybus, O. G. & Ke, C. 2018. Molecular Evolution, Diversity, and Adaptation of Influenza A (H7N9) Viruses in China. *Emerg Infect Dis, 24, 1795-1805.*

Pryce, R., Ng, W. M., Zeltina, A., Watanabe, Y., El Omari, K., Wagner, A., Bowden, T. A. 2018. Structure-Based Classification Defines the Discrete Conformational Classes Adopted by the Arenaviral GP1. 2018. *J Virol. 93 (1) e01048-18;* ***DOI****: 10.1128/JVI.01048-18*

Appendix B:

Bibliography

Appendix B: Bibliography

1997. Unified nomenclature for Eph family receptors and their ligands, the ephrins. Eph Nomenclature Committee. *Cell,* 90**,** 403-4.

ABRAHAM, J., CORBETT, K. D., FARZAN, M., CHOE, H. & HARRISON, S. C. 2010. Structural basis for receptor recognition by New World hemorrhagic fever arenaviruses. *Nat Struct Mol Biol,* 17**,** 438-44.

ABRAHAM, J., KWONG, J. A., ALBARINO, C. G., LU, J. G., RADOSHITZKY, S. R., SALAZAR-BRAVO, J., FARZAN, M., SPIROPOULOU, C. F. & CHOE, H. 2009. Host-species transferrin receptor 1 orthologs are cellular receptors for nonpathogenic new world clade B arenaviruses. *PLoS Pathog,* 5**,** e1000358.

ACCIANI, M., ALSTON, J. T., ZHAO, G., REYNOLDS, H., ALI, A. M., XU, B. & BRINDLEY, M. A. 2017. Mutational Analysis of Lassa Virus Glycoprotein Highlights Regions Required for Alpha-Dystroglycan Utilization. *J Virol,* 91.

ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L. W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr,* 66**,** 213-21.

ANG, B. S. P., LIM, T. C. C. & WANG, L. 2018. Nipah Virus Infection. *J Clin Microbiol,* 56.

ANON. 2017. *Hendra Virus Infection* [Online]. Online: Queensland Government. Available: http://conditions.health.qld.gov.au/HealthCondition/condition/14/217/363/hendra-virus-infection [Accessed 25/09/2018 2018].

APPEL, W. 1986. Chymotrypsin: molecular and catalytic properties. *Clin Biochem,* 19**,** 317-22.

ARCHER, A. M. & RICO-HESSE, R. 2002. High genetic divergence and recombination in Arenaviruses from the Americas. *Virology,* 304**,** 274-81.

ARICESCU, A. R., LU, W. & JONES, E. Y. 2006. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr D Biol Crystallogr,* 62**,** 1243-50.

AUDSLEY, M. D. & MOSELEY, G. W. 2013. Paramyxovirus evasion of innate immunity: Diverse strategies for common targets. *World J Virol,* 2**,** 57-70.

BAKER, N. A., SEPT, D., JOSEPH, S., HOLST, M. J. & MCCAMMON, J. A. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A,* 98**,** 10037-41.

BAMFORD, D. H., GRIMES, J. M. & STUART, D. I. 2005. What does structure tell us about virus evolution? *Current Opinion in Structural Biology,* 15**,** 655-663.

BANERJEE, N. & MUKHOPADHYAY, S. 2016. Viral glycoproteins: biological role and application in diagnosis. *Virusdisease,* 27**,** 1-11.

BARRERA ORO, J. G. & MCKEE, K. T., JR. 1991. Toward a vaccine against Argentine hemorrhagic fever. *Bull Pan Am Health Organ,* 25**,** 118-26.

BARTON, L. L. & HYNDMAN, N. J. 2000. Lymphocytic choriomeningitis virus: reemerging central nervous system pathogen. *Pediatrics,* 105**,** E35.

BEHRENS, A. J., VASILJEVIC, S., PRITCHARD, L. K., HARVEY, D. J., ANDEV, R. S., KRUMM, S. A., STRUWE, W. B., CUPO, A., KUMAR, A., ZITZMANN, N., SEABRIGHT, G. E., KRAMER, H. B., SPENCER, D. I., ROYLE, L., LEE, J. H., KLASSE, P. J., BURTON, D. R., WILSON, I. A., WARD, A. B., SANDERS, R. W., MOORE, J. P., DOORES, K. J. & CRISPIN, M. 2016. Composition and Antigenic Effects of Individual Glycan Sites of a Trimeric HIV-1 Envelope Glycoprotein. *Cell Rep,* 14**,** 2695-706.

BERGEMANN, A. D., ZHANG, L., CHIANG, M.-K., BRAMBILLA, R., KLEIN, R. & FLANAGAN, J. G. 1998. Ephrin-B3, a ligand for the receptor EphB3, expressed at the midline of the developing neural tube. *Oncogene,* 16**,** 471.

BIERING, S. B., HUANG, A., VU, A. T., ROBINSON, L. R., BRADEL-TRETHEWAY, B., CHOI, E., LEE, B. & AGUILAR, H. C. 2012. N-Glycans on the Nipah virus attachment glycoprotein modulate fusion and viral entry as they protect against antibody neutralization. *J Virol,* 86**,** 11991-2002.

BIRD, L. E., RADA, H., FLANAGAN, J., DIPROSE, J. M., GILBERT, R. J. & OWENS, R. J. 2014. Application of In-Fusion cloning for the parallel construction of E. coli expression vectors. *Methods Mol Biol,* 1116**,** 209-34.

BLASDELL, K. R., DUONG, V., ELOIT, M., CHRETIEN, F., LY, S., HUL, V., DEUBEL, V., MORAND, S. & BUCHY, P. 2016. Evidence of human infection by a new mammarenavirus endemic to Southeastern Asia. *Elife,* 5.

BONHOMME, C. J., CAPUL, A. A., LAURON, E. J., BEDERKA, L. H., KNOPP, K. A. & BUCHMEIER, M. J. 2011. Glycosylation modulates arenavirus glycoprotein expression and function. *Virology,* 409**,** 223-33.

BORROW, P. & OLDSTONE, M. B. 1994. Mechanism of lymphocytic choriomeningitis virus entry into cells. *Virology,* 198**,** 1-9.

BOSE, S., JARDETZKY, T. S. & LAMB, R. A. 2015. Timing is everything: Fine-tuned molecular machines orchestrate paramyxovirus entry. *Virology,* 479-480**,** 518-31.

BOSE, S., SONG, A. S., JARDETZKY, T. S. & LAMB, R. A. 2014. Fusion activation through attachment protein stalk domains indicates a conserved core mechanism of paramyxovirus entry into cells. *J Virol,* 88**,** 3925-41.

BOUSSIF, O., LEZOUALC'H, F., ZANTA, M. A., MERGNY, M. D., SCHERMAN, D., DEMENEIX, B. & BEHR, J. P. 1995. A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. *Proc Natl Acad Sci U S A,* 92**,** 7297-301.

BOWDEN, T. A., ARICESCU, A. R., GILBERT, R. J., GRIMES, J. M., JONES, E. Y. & STUART, D. I. 2008a. Structural basis of Nipah and Hendra virus attachment to their cell-surface receptor ephrin-B2. *Nat Struct Mol Biol,* 15**,** 567-72.

BOWDEN, T. A., ARICESCU, A. R., NETTLESHIP, J. E., SIEBOLD, C., RAHMAN-HUQ, N., OWENS, R. J., STUART, D. I. & JONES, E. Y. 2009a. Structural plasticity of eph receptor A4 facilitates cross-class ephrin signaling. *Structure,* 17**,** 1386-97.

BOWDEN, T. A., CRISPIN, M., GRAHAM, S. C., HARVEY, D. J., GRIMES, J. M., JONES, E. Y. & STUART, D. I. 2009b. Unusual molecular architecture of the machupo virus attachment glycoprotein. *J Virol,* 83**,** 8259-65.

BOWDEN, T. A., CRISPIN, M., HARVEY, D. J., ARICESCU, A. R., GRIMES, J. M., JONES, E. Y. & STUART, D. I. 2008b. Crystal structure and carbohydrate analysis of Nipah virus attachment glycoprotein: a template for antiviral and vaccine design. *J Virol,* 82**,** 11628-36.

BOWDEN, T. A., CRISPIN, M., HARVEY, D. J., JONES, E. Y. & STUART, D. I. 2010a. Dimeric architecture of the Hendra virus attachment glycoprotein: evidence for a conserved mode of assembly. *J Virol,* 84**,** 6208-17.

BOWDEN, T. A., CRISPIN, M., JONES, E. Y. & STUART, D. I. 2010b. Shared paramyxoviral glycoprotein architecture is adapted for diverse attachment strategies. *Biochem Soc Trans,* 38**,** 1349-55.

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. *Mol Phylogenet Evol,* 8**,** 301-16.

BRADEL-TRETHEWAY, B. G., LIU, Q., STONE, J. A., MCINALLY, S. & AGUILAR, H. C. 2015. Novel Functions of Hendra Virus G N-Glycans and Comparisons to Nipah Virus. *J Virol,* 89**,** 7235-47.

BRAGG, W. H. & BRAGG, W. L. 1913. The reflection of X-rays by crystals. *Proceedings of the Royal Society of London. Series A,* 88**,** 428.

BRANCO, L. M. & GARRY, R. F. 2009. Characterization of the Lassa virus GP1 ectodomain shedding: implications for improved diagnostic platforms. *Virology Journal,* 6**,** 147.

BRANCO, L. M., GROVE, J. N., MOSES, L. M., GOBA, A., FULLAH, M., MOMOH, M., SCHOEPP, R. J., BAUSCH, D. G. & GARRY, R. F. 2010. Shedding of soluble glycoprotein 1 detected during acute Lassa virus infection in human subjects. *Virology Journal,* 7**,** 306.

BREITBART, M. & ROHWER, F. 2005. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol,* 13**,** 278-84.

BRIESE, T., PAWESKA, J. T., MCMULLAN, L. K., HUTCHISON, S. K., STREET, C., PALACIOS, G., KHRISTOVA, M. L., WEYER, J., SWANEPOEL, R., EGHOLM, M., NICHOL, S. T. & LIPKIN, W. I. 2009. Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated arenavirus from southern Africa. *PLoS Pathog,* 5**,** e1000455.

BRÜNGER, A. T. 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature,* 355**,** 472.

BURRI, D. J., DA PALMA, J. R., KUNZ, S. & PASQUATO, A. 2012. Envelope glycoprotein of arenaviruses. *Viruses,* 4**,** 2162-81.

CALISHER, C. H., TZIANABOS, T., LORD, R. D. & COLEMAN, P. H. 1970. Tamiami virus, a new member of the TaCaribe group. *Am J Trop Med Hyg,* 19**,** 520-6.

CAO, W., HENRY, M. D., BORROW, P., YAMADA, H., ELDER, J. H., RAVKOV, E. V., NICHOL, S. T., COMPANS, R. W., CAMPBELL, K. P. & OLDSTONE, M. B. 1998. Identification of alpha-dystroglycan as a receptor for lymphocytic choriomeningitis virus and Lassa fever virus. *Science,* 282**,** 2079-81.

CAPUL, A. A., PEREZ, M., BURKE, E., KUNZ, S., BUCHMEIER, M. J. & DE LA TORRE, J. C. 2007. Arenavirus Z-glycoprotein association requires Z myristoylation but not functional RING or late domains. *J Virol,* 81**,** 9451-60.

CARETTE, J. E., RAABEN, M., WONG, A. C., HERBERT, A. S., OBERNOSTERER, G., MULHERKAR, N., KUEHNE, A. I., KRANZUSCH, P. J., GRIFFIN, A. M., RUTHEL, G., DAL CIN, P., DYE, J. M.,

WHELAN, S. P., CHANDRAN, K. & BRUMMELKAMP, T. R. 2011. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature,* 477, 340-3.

CARRILLO-BUSTAMANTE, P., NGUYEN, T. H. T., OESTEREICH, L., GUNTHER, S., GUEDJ, J. & GRAW, F. 2017. Determining Ribavirin's mechanism of action against Lassa virus infection. *Sci Rep,* 7, 11693.

CDC. 2018. *Remembering the 1918 Influenza Pandemic* [Online]. Online: CDC. Available: https://www.cdc.gov/features/1918-flu-pandemic/index.html [Accessed 2018 24/09/2018].

CHANG, A. & DUTCH, R. E. 2012. Paramyxovirus fusion and entry: multiple paths to a common end. *Viruses,* 4, 613-36.

CHANG, V. T., CRISPIN, M., ARICESCU, A. R., HARVEY, D. J., NETTLESHIP, J. E., FENNELLY, J. A., YU, C., BOLES, K. S., EVANS, E. J., STUART, D. I., DWEK, R. A., JONES, E. Y., OWENS, R. J. & DAVIS, S. J. 2007. Glycoprotein structural genomics: solving the glycosylation problem. *Structure,* 15, 267-73.

CHAYEN, N. E. & SARIDAKIS, E. 2008. Protein crystallization: from purified protein to diffraction-quality crystal. *Nat Methods,* 5, 147-53.

CHEN, V. B., ARENDALL, W. B., 3RD, HEADD, J. J., KEEDY, D. A., IMMORMINO, R. M., KAPRAL, G. J., MURRAY, L. W., RICHARDSON, J. S. & RICHARDSON, D. C. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr,* 66, 12-21.

CHOE, H., JEMIELITY, S., ABRAHAM, J., RADOSHITZKY, S. R. & FARZAN, M. 2011. Transferrin receptor 1 in the zoonosis and pathogenesis of New World hemorrhagic fever arenaviruses. *Curr Opin Microbiol,* 14, 476-82.

CHONG, H. T., KAMARULZAMAN, A., TAN, C. T., GOH, K. J., THAYAPARAN, T., KUNJAPAN, S. R., CHEW, N. K., CHUA, K. B. & LAM, S. K. 2001. Treatment of acute Nipah encephalitis with ribavirin. *Ann Neurol,* 49, 810-3.

CHRENCIK, J. E., BROOUN, A., KRAUS, M. L., RECHT, M. I., KOLATKAR, A. R., HAN, G. W., SEIFERT, J. M., WIDMER, H., AUER, M. & KUHN, P. 2006. Structural and biophysical characterization of the EphB4*ephrinB2 protein-protein interaction and receptor specificity. *J Biol Chem,* 281, 28185-92.

CHUA, K. B., BELLINI, W. J., ROTA, P. A., HARCOURT, B. H., TAMIN, A., LAM, S. K., KSIAZEK, T. G., ROLLIN, P. E., ZAKI, S. R., SHIEH, W., GOLDSMITH, C. S., GUBLER, D. J., ROEHRIG, J. T., EATON, B., GOULD, A. R., OLSON, J., FIELD, H., DANIELS, P., LING, A. E., PETERS, C. J., ANDERSON, L. J. & MAHY, B. W. 2000. Nipah virus: a recently emergent deadly paramyxovirus. *Science,* 288, 1432-5.

CLARK, L. E., MAHMUTOVIC, S., RAYMOND, D. D., DILANYAN, T., KOMA, T., MANNING, J. T., SHANKAR, S., LEVIS, S. C., BRIGGILER, A. M., ENRIA, D. A., WUCHERPFENNIG, K. W., PAESSLER, S. & ABRAHAM, J. 2018. Vaccine-elicited receptor-binding site antibodies neutralize two New World hemorrhagic fever arenaviruses. *Nat Commun,* 9, 1884.

COGSWELL-HAWKINSON, A., BOWEN, R., JAMES, S., GARDINER, D., CALISHER, C. H., ADAMS, R. & SCHOUNTZ, T. 2012. Tacaribe virus causes fatal infection of an ostensible reservoir host, the Jamaican fruit bat. *J Virol,* 86, 5791-9.

COHEN-DVASHI, H., COHEN, N., ISRAELI, H. & DISKIN, R. 2015. Molecular Mechanism for LAMP1 Recognition by Lassa Virus. *J Virol,* 89, 7584-92.

COHEN-DVASHI, H., ISRAELI, H., SHANI, O., KATZ, A. & DISKIN, R. 2016. Role of LAMP1 Binding and pH Sensing by the Spike Complex of Lassa Virus. *J Virol,* 90, 10329-10338.

CORPET, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res,* 16, 10881-90.

CRENNELL, S., TAKIMOTO, T., PORTNER, A. & TAYLOR, G. 2000. Crystal structure of the multifunctional paramyxovirus hemagglutinin-neuraminidase. *Nat Struct Biol,* 7, 1068-74.

DE ARAUJO, J., LO, M. K., TAMIN, A., OMETTO, T. L., THOMAZELLI, L. M., NARDI, M. S., HURTADO, R. F., NAVA, A., SPIROPOULOU, C. F., ROTA, P. A. & DURIGON, E. L. 2017. Antibodies Against Henipa-Like Viruses in Brazilian Bats. *Vector Borne Zoonotic Dis,* 17, 271-274.

DELGADO, S., ERICKSON, B. R., AGUDO, R., BLAIR, P. J., VALLEJO, E., ALBARINO, C. G., VARGAS, J., COMER, J. A., ROLLIN, P. E., KSIAZEK, T. G., OLSON, J. G. & NICHOL, S. T. 2008. Chapare virus, a newly discovered arenavirus isolated from a fatal hemorrhagic fever case in Bolivia. *PLoS Pathog,* 4, e1000047.

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, 455-65.

DOLNIK, O., VOLCHKOVA, V., GARTEN, W., CARBONNELLE, C., BECKER, S., KAHNT, J., STROHER, U., KLENK, H. D. & VOLCHKOV, V. 2004. Ectodomain shedding of the glycoprotein GP of Ebola virus. *EMBO J,* 23, 2175-84.

DONG, A., XU, X., EDWARDS, A. M., MIDWEST CENTER FOR STRUCTURAL, G., STRUCTURAL GENOMICS, C., CHANG, C., CHRUSZCZ, M., CUFF, M., CYMBOROWSKI, M., DI LEO, R., EGOROVA, O., EVDOKIMOVA, E., FILIPPOVA, E., GU, J., GUTHRIE, J., IGNATCHENKO, A., JOACHIMIAK, A., KLOSTERMANN, N., KIM, Y., KORNIYENKO, Y., MINOR, W., QUE, Q., SAVCHENKO, A., SKARINA, T., TAN, K., YAKUNIN, A., YEE, A., YIM, V., ZHANG, R., ZHENG, H., AKUTSU, M., ARROWSMITH, C., AVVAKUMOV, G. V., BOCHKAREV, A., DAHLGREN, L. G., DHE-PAGANON, S., DIMOV, S., DOMBROVSKI, L., FINERTY, P., JR., FLODIN, S., FLORES, A., GRASLUND, S., HAMMERSTROM, M., HERMAN, M. D., HONG, B. S., HUI, R., JOHANSSON, I., LIU, Y., NILSSON, M., NEDYALKOVA, L., NORDLUND, P., NYMAN, T., MIN, J., OUYANG, H., PARK, H. W., QI, C., RABEH, W., SHEN, L., SHEN, Y., SUKUMARD, D., TEMPEL, W., TONG, Y., TRESAGUES, L., VEDADI, M., WALKER, J. R., WEIGELT, J., WELIN, M., WU, H., XIAO, T., ZENG, H. & ZHU, H. 2007. In situ proteolysis for protein crystallization and structure determination. *Nat Methods,* 4, 1019-21.

DOWNS, W. G., ANDERSON, C. R., SPENCE, L., AITKEN, T. H. & GREENHALL, A. H. 1963. Tacaribe virus, a new agent isolated from Artibeus bats and mosquitoes in Trinidad, West Indies. *Am J Trop Med Hyg,* 12, 640-6.

DRAKE, J. W. & HOLLAND, J. J. 1999. Mutation rates among RNA viruses. *Proc Natl Acad Sci U S A,* 96, 13910-3.

DREXLER, J. F., CORMAN, V. M., MULLER, M. A., MAGANGA, G. D., VALLO, P., BINGER, T., GLOZA-RAUSCH, F., COTTONTAIL, V. M., RASCHE, A., YORDANOV, S., SEEBENS, A., KNORNSCHILD, M., OPPONG, S., ADU SARKODIE, Y., PONGOMBO, C., LUKASHEV, A. N., SCHMIDT-CHANASIT, J., STOCKER, A., CARNEIRO, A. J., ERBAR, S., MAISNER, A., FRONHOFFS, F., BUETTNER, R., KALKO, E. K., KRUPPA, T., FRANKE, C. R., KALLIES, R., YANDOKO, E. R., HERRLER, G., REUSKEN, C., HASSANIN, A., KRUGER, D. H., MATTHEE, S., ULRICH, R. G., LEROY, E. M. & DROSTEN, C. 2012. Bats host major mammalian paramyxoviruses. *Nat Commun,* 3, 796.

DUFFY, S., SHACKELTON, L. A. & HOLMES, E. C. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet,* 9, 267-76.

DUTCH, R. E. 2010. Entry and fusion of emerging paramyxoviruses. *PLoS Pathog,* 6, e1000881.

EATON, B. T., BRODER, C. C., MIDDLETON, D. & WANG, L. F. 2006. Hendra and Nipah viruses: different and dangerous. *Nat Rev Microbiol,* 4, 23-35.

EATON, B. T., WRIGHT, P. J., WANG, L. F., SERGEYEV, O., MICHALSKI, W. P., BOSSART, K. N. & BRODER, C. C. 2004. Henipaviruses: recent observations on regulation of transcription and the nature of the cell receptor. *Arch Virol Suppl*, 122-31.

ELBEIN, A. D., TROPEA, J. E., MITCHELL, M. & KAUSHAL, G. P. 1990. Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. *J Biol Chem,* 265, 15599-605.

EMONET, S. F., DE LA TORRE, J. C., DOMINGO, E. & SEVILLA, N. 2009. Arenavirus genetic diversity and its biological implications. *Infection, Genetics and Evolution,* 9, 417-429.

EMSLEY, P. & COWTAN, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr,* 60, 2126-32.

ENSERINK, M. 2000. New Arenavirus Blamed for Recent Deaths in California. *Science,* 289, 842-843.

FELSENSTEIN, J. 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics,* 5, 164-166.

FLANAGAN, M. L., OLDENBURG, J., REIGNIER, T., HOLT, N., HAMILTON, G. A., MARTIN, V. K. & CANNON, P. M. 2008. New world clade B arenaviruses can use transferrin receptor 1 (TfR1)-dependent and -independent entry pathways, and glycoproteins from human pathogenic strains are associated with the use of TfR1. *J Virol,* 82, 938-48.

FLOTENMEYER, M., WEISS, H., TRIBET, C., POPOT, J. L. & LEONARD, K. 2007. The use of amphipathic polymers for cryo electron microscopy of NADH:ubiquinone oxidoreductase (complex I). *J Microsc,* 227, 229-35.

FLOYD, D. L., RAGAINS, J. R., SKEHEL, J. J., HARRISON, S. C. & VAN OIJEN, A. M. 2008. Single-particle kinetics of influenza virus membrane fusion. *Proc Natl Acad Sci U S A,* 105, 15382-7.

FRIEDEN, T. R. & DAMON, I. K. 2015. Ebola in West Africa--CDC's Role in Epidemic Detection, Control, and Prevention. *Emerg Infect Dis,* 21, 1897-905.

FULHORST, C. F., BOWEN, M. D., KSIAZEK, T. G., ROLLIN, P. E., NICHOL, S. T., KOSOY, M. Y. & PETERS, C. J. 1996. Isolation and characterization of Whitewater Arroyo virus, a novel North American arenavirus. *Virology,* 224, 114-20.

GAGO, S., ELENA, S. F., FLORES, R. & SANJUAN, R. 2009. Extremely high mutation rate of a hammerhead viroid. *Science,* 323, 1308.

GATTER, K. C., BROWN, G., TROWBRIDGE, I. S., WOOLSTON, R. E. & MASON, D. Y. 1983. Transferrin receptors in human tissues: their distribution and possible clinical relevance. *J Clin Pathol,* 36, 539-45.

GONCALVES, A. R., MORAZ, M. L., PASQUATO, A., HELENIUS, A., LOZACH, P. Y. & KUNZ, S. 2013. Role of DC-SIGN in Lassa virus entry into human dendritic cells. *J Virol,* 87, 11504-15.

GONZALEZ, J. P., EMONET, S., DE LAMBALLERIE, X. & CHARREL, R. 2007. Arenaviruses. *Curr Top Microbiol Immunol,* 315, 253-88.

GOUET, P., ROBERT, X. & COURCELLE, E. 2003. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res,* 31, 3320-3.

GRIMM, S. K. & ACKERMAN, M. E. 2013. Vaccine design: emerging concepts and renewed optimism. *Curr Opin Biotechnol,* 24, 1078-88.

GROVE, J. & MARSH, M. 2011. The cell biology of receptor-mediated virus entry. *J Cell Biol,* 195, 1071-82.

GRUENINGER-LEITCH, F., D'ARCY, A., D'ARCY, B. & CHENE, C. 1996. Deglycosylation of proteins for crystallization using recombinant fusion protein glycosidases. *Protein Sci,* 5, 2617-22.

GUNTHER, S. & LENZ, O. 2004. Lassa virus. *Crit Rev Clin Lab Sci,* 41, 339-90.

GUPTA, S., CELESTRE, R., FENG, J. & RALSTON, C. 2016a. Advancements and Application of Microsecond Synchrotron X-ray Footprinting at the Advanced Light Source. *Synchrotron Radiation News,* 29, 39-44.

GUPTA, S., FENG, J., CHANCE, M. & RALSTON, C. 2016b. Recent Advances and Applications in Synchrotron X-Ray Protein Footprinting for Protein Structure and Dynamics Elucidation. *Protein Pept Lett,* 23, 309-22.

GURLEY, E. S., MONTGOMERY, J. M., HOSSAIN, M. J., BELL, M., AZAD, A. K., ISLAM, M. R., MOLLA, M. A., CARROLL, D. S., KSIAZEK, T. G., ROTA, P. A., LOWE, L., COMER, J. A., ROLLIN, P., CZUB, M., GROLLA, A., FELDMANN, H., LUBY, S. P., WOODWARD, J. L. & BREIMAN, R. F. 2007. Person-to-person transmission of Nipah virus in a Bangladeshi community. *Emerg Infect Dis,* 13, 1031-7.

HALLDORSSON, S., LI, S., LI, M., HARLOS, K., BOWDEN, T. A. & HUISKONEN, J. T. 2018. Shielding and activation of a viral membrane fusion protein. *Nat Commun,* 9, 349.

HALPIN, K., BANKAMP, B., HARCOURT, B. H., BELLINI, W. J. & ROTA, P. A. 2004. Nipah virus conforms to the rule of six in a minigenome replication assay. *J Gen Virol,* 85, 701-7.

HALPIN, K., HYATT, A. D., FOGARTY, R., MIDDLETON, D., BINGHAM, J., EPSTEIN, J. H., RAHMAN, S. A., HUGHES, T., SMITH, C., FIELD, H. E., DASZAK, P. & HENIPAVIRUS ECOLOGY RESEARCH, G. 2011. Pteropid bats are confirmed as the reservoir hosts of henipaviruses: a comprehensive experimental study of virus transmission. *Am J Trop Med Hyg,* 85, 946-51.

HANGARTNER, L., ZINKERNAGEL, R. M. & HENGARTNER, H. 2006. Antiviral antibody responses: the two extremes of a wide spectrum. *Nat Rev Immunol,* 6, 231-43.

HARRISON, L. H., HALSEY, N. A., MCKEE, K. T., JR., PETERS, C. J., BARRERA ORO, J. G., BRIGGILER, A. M., FEUILLADE, M. R. & MAIZTEGUI, J. I. 1999. Clinical case definitions for Argentine hemorrhagic fever. *Clin Infect Dis,* 28, 1091-4.

HARRISON, S. C. 2008. Viral membrane fusion. *Nat Struct Mol Biol,* 15, 690-8.

HARRISON, S. C. 2015. Viral membrane fusion. *Virology,* 479-480, 498-507.

HARVEY, R., MARTIN, A. C., ZAMBON, M. & BARCLAY, W. S. 2004. Restrictions to the adaptation of influenza a virus h5 hemagglutinin to the human host. *J Virol,* 78, 502-7.

HASTIE, K. M., IGONET, S., SULLIVAN, B. M., LEGRAND, P., ZANDONATTI, M. A., ROBINSON, J. E., GARRY, R. F., REY, F. A., OLDSTONE, M. B. & SAPHIRE, E. O. 2016. Crystal structure of the prefusion surface glycoprotein of the prototypic arenavirus LCMV. *Nat Struct Mol Biol,* 23, 513-521.

HASTIE, K. M., KIMBERLIN, C. R., ZANDONATTI, M. A., MACRAE, I. J. & SAPHIRE, E. O. 2011. Structure of the Lassa virus nucleoprotein reveals a dsRNA-specific 3' to 5' exonuclease activity essential for immune suppression. *Proc Natl Acad Sci U S A,* 108**,** 2396-401.

HASTIE, K. M., ZANDONATTI, M. A., KLEINFELTER, L. M., HEINRICH, M. L., ROWLAND, M. M., CHANDRAN, K., BRANCO, L. M., ROBINSON, J. E., GARRY, R. F. & SAPHIRE, E. O. 2017. Structural basis for antibody-mediated neutralization of Lassa virus. *Science,* 356**,** 923-928.

HIMANEN, J. P., RAJASHANKAR, K. R., LACKMANN, M., COWAN, C. A., HENKEMEYER, M. & NIKOLOV, D. B. 2001. Crystal structure of an Eph receptor-ephrin complex. *Nature,* 414**,** 933-8.

HIMANEN, J. P., SAHA, N. & NIKOLOV, D. B. 2007. Cell-cell signaling via Eph receptors and ephrins. *Curr Opin Cell Biol,* 19**,** 534-42.

HOCHULI, E., BANNWARTH, W., DÖBELI, H., GENTZ, R. & STÜBER, D. 1988. Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Bio/Technology,* 6**,** 1321.

HOLMES, E. C. 2009. RNA virus genomics: a world of possibilities. *J Clin Invest,* 119**,** 2488-95.

HOOPER, P., ZAKI, S., DANIELS, P. & MIDDLETON, D. 2001. Comparative pathology of the diseases caused by Hendra and Nipah viruses. *Microbes Infect,* 3**,** 315-22.

HOULIHAN, C. & BEHRENS, R. 2017. Lassa fever. *BMJ,* 358**,** j2986.

HSU, V. P., HOSSAIN, M. J., PARASHAR, U. D., ALI, M. M., KSIAZEK, T. G., KUZMIN, I., NIEZGODA, M., RUPPRECHT, C., BRESEE, J. & BREIMAN, R. F. 2004. Nipah virus encephalitis reemergence, Bangladesh. *Emerg Infect Dis,* 10**,** 2082-7.

HUSSAIN, R., JAVORFI, T. & SILIGARDI, G. 2012. Circular dichroism beamline B23 at the Diamond Light Source. *J Synchrotron Radiat,* 19**,** 132-5.

HYATT, A. D., ZAKI, S. R., GOLDSMITH, C. S., WISE, T. G. & HENGSTBERGER, S. G. 2001. Ultrastructure of Hendra virus and Nipah virus within cultured cells and host animals. *Microbes Infect,* 3**,** 297-306.

IGONET, S., VANEY, M. C., VONRHEIN, C., BRICOGNE, G., STURA, E. A., HENGARTNER, H., ESCHLI, B. & REY, F. A. 2011. X-ray structure of the arenavirus glycoprotein GP2 in its postfusion hairpin conformation. *Proc Natl Acad Sci U S A,* 108**,** 19967-72.

ISRAELI, H., COHEN-DVASHI, H., SHULMAN, A., SHIMON, A. & DISKIN, R. 2017. Mapping of the Lassa virus LAMP1 binding site reveals unique determinants not shared by other old world arenaviruses. *PLoS Pathog,* 13**,** e1006337.

JAE, L. T., RAABEN, M., HERBERT, A. S., KUEHNE, A. I., WIRCHNIANSKI, A. S., SOH, T. K., STUBBS, S. H., JANSSEN, H., DAMME, M., SAFTIG, P., WHELAN, S. P., DYE, J. M. & BRUMMELKAMP, T. R. 2014. Virus entry. Lassa virus entry requires a trigger-induced receptor switch. *Science,* 344**,** 1506-10.

JÁVORFI, T., HUSSAIN, R., MYATT, D. & SILIGARDI, G. 2010. Measuring circular dichroism in a capillary cell using the b23 synchrotron radiation CD beamline at diamond light source. *Chirality,* 22**,** E149-E153.

JOLLY, C. L. & SATTENTAU, Q. J. 2013. Attachment factors. *Adv Exp Med Biol,* 790**,** 1-23.

KABSCH, W. & SANDER, C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* 22**,** 2577-637.

KARPLUS, P. A. & DIEDERICHS, K. 2012. Linking crystallographic model and data quality. *Science,* 336**,** 1030-3.

KELLY, S. M., JESS, T. J. & PRICE, N. C. 2005. How to study proteins by circular dichroism. *Biochim Biophys Acta,* 1751**,** 119-39.

KILPATRICK, A. M. & RANDOLPH, S. E. 2012. Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *Lancet,* 380**,** 1946-55.

KISELAR, J. G. & CHANCE, M. R. 2010. Future directions of structural mass spectrometry using hydroxyl radical footprinting. *J Mass Spectrom,* 45**,** 1373-82.

KOIDE, S. 2009. Engineering of recombinant crystallization chaperones. *Curr Opin Struct Biol,* 19**,** 449-57.

KOONIN, E. V., SENKEVICH, T. G. & DOLJA, V. V. 2006. The ancient Virus World and evolution of cells. *Biol Direct,* 1**,** 29.

KRISSINEL, E. & HENRICK, K. 2007. Inference of macromolecular assemblies from crystalline state. *J Mol Biol,* 372**,** 774-97.

KUNZ, S., ROJEK, J. M., KANAGAWA, M., SPIROPOULOU, C. F., BARRESI, R., CAMPBELL, K. P. & OLDSTONE, M. B. 2005. Posttranslational modification of alpha-dystroglycan, the cellular receptor for arenaviruses, by the glycosyltransferase LARGE is critical for virus binding. *J Virol,* 79**,** 14282-96.

LAING, E. D., AMAYA, M., NAVARATNARAJAH, C. K., FENG, Y. R., CATTANEO, R., WANG, L. F. & BRODER, C. C. 2018. Rescue and characterization of recombinant cedar virus, a non-pathogenic Henipavirus species. *Virol J,* 15**,** 56.

LAWRENCE, C. M., RAY, S., BABYONYSHEV, M., GALLUSER, R., BORHANI, D. W. & HARRISON, S. C. 1999. Crystal structure of the ectodomain of human transferrin receptor. *Science,* 286**,** 779-82.

LEE, B., PERNET, O., AHMED, A. A., ZELTINA, A., BEATY, S. M. & BOWDEN, T. A. 2015. Molecular recognition of human ephrinB2 cell surface receptor by an emergent African henipavirus. *Proc Natl Acad Sci U S A,* 112**,** E2156-65.

LEE, K. J., NOVELLA, I. S., TENG, M. N., OLDSTONE, M. B. & DE LA TORRE, J. C. 2000. NP and L proteins of lymphocytic choriomeningitis virus (LCMV) are sufficient for efficient transcription and replication of LCMV genomic RNA analogs. *J Virol,* 74**,** 3470-7.

LEFKOWITZ, E. J., DEMPSEY, D. M., HENDRICKSON, R. C., ORTON, R. J., SIDDELL, S. G. & SMITH, D. B. 2018. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res,* 46**,** D708-D717.

LI, C. X., SHI, M., TIAN, J. H., LIN, X. D., KANG, Y. J., CHEN, L. J., QIN, X. C., XU, J., HOLMES, E. C. & ZHANG, Y. Z. 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *Elife,* 4.

LI, S., SUN, Z., PRYCE, R., PARSY, M. L., FEHLING, S. K., SCHLIE, K., SIEBERT, C. A., GARTEN, W., BOWDEN, T. A., STRECKER, T. & HUISKONEN, J. T. 2016. Acidic pH-Induced Conformations and LAMP1 Binding of the Lassa Virus Glycoprotein Spike. *PLoS Pathog,* 12**,** e1005418.

LISABETH, E. M., FALIVELLI, G. & PASQUALE, E. B. 2013. Eph receptor signaling and ephrins. *Cold Spring Harb Perspect Biol,* 5.

LISIEUX, T., COIMBRA, M., NASSAR, E. S., BURATTINI, M. N., DE SOUZA, L. T., FERREIRA, I., ROCCO, I. M., DA ROSA, A. P., VASCONCELOS, P. F., PINHEIRO, F. P. & ET AL. 1994. New arenavirus isolated in Brazil. *Lancet,* 343**,** 391-2.

LOUREIRO, M. E., D'ANTUONO, A., LEVINGSTON MACLEOD, J. M. & LOPEZ, N. 2012. Uncovering viral protein-protein interactions and their role in arenavirus life cycle. *Viruses,* 4**,** 1651-67.

LUKASHEVICH, I. S., CLEGG, J. C. & SIDIBE, K. 1993. Lassa virus activity in Guinea: distribution of human antiviral antibody defined using enzyme-linked immunosorbent assay with recombinant antigen. *J Med Virol,* 40**,** 210-7.

MAHMUTOVIC, S., CLARK, L., LEVIS, S. C., BRIGGILER, A. M., ENRIA, D. A., HARRISON, S. C. & ABRAHAM, J. 2015. Molecular Basis for Antibody-Mediated Neutralization of New World Hemorrhagic Fever Mammarenaviruses. *Cell Host Microbe,* 18**,** 705-13.

MAIZTEGUI, J. I., MCKEE, K. T., JR., BARRERA ORO, J. G., HARRISON, L. H., GIBBS, P. H., FEUILLADE, M. R., ENRIA, D. A., BRIGGILER, A. M., LEVIS, S. C., AMBROSIO, A. M., HALSEY, N. A. & PETERS, C. J. 1998. Protective efficacy of a live attenuated vaccine against Argentine hemorrhagic fever. AHF Study Group. *J Infect Dis,* 177**,** 277-83.

MALEY, F., TRIMBLE, R. B., TARENTINO, A. L. & PLUMMER, T. H. 1989. Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Analytical Biochemistry,* 180**,** 195-204.

MANN, J. K. & NDUNG'U, T. 2015. HIV-1 vaccine immunogen design strategies. *Virol J,* 12**,** 3.

MARCSISIN, S. R. & ENGEN, J. R. 2010. Hydrogen exchange mass spectrometry: what is it and what can it tell us? *Anal Bioanal Chem,* 397**,** 967-72.

MARGULIES, M., EGHOLM, M., ALTMAN, W. E., ATTIYA, S., BADER, J. S., BEMBEN, L. A., BERKA, J., BRAVERMAN, M. S., CHEN, Y. J., CHEN, Z., DEWELL, S. B., DU, L., FIERRO, J. M., GOMES, X. V., GODWIN, B. C., HE, W., HELGESEN, S., HO, C. H., IRZYK, G. P., JANDO, S. C., ALENQUER, M. L., JARVIE, T. P., JIRAGE, K. B., KIM, J. B., KNIGHT, J. R., LANZA, J. R., LEAMON, J. H., LEFKOWITZ, S. M., LEI, M., LI, J., LOHMAN, K. L., LU, H., MAKHIJANI, V. B., MCDADE, K. E., MCKENNA, M. P., MYERS, E. W., NICKERSON, E., NOBILE, J. R., PLANT, R., PUC, B. P., RONAN, M. T., ROTH, G. T., SARKIS, G. J., SIMONS, J. F., SIMPSON, J. W., SRINIVASAN, M., TARTARO, K. R., TOMASZ, A., VOGT, K. A., VOLKMER, G. A., WANG, S. H., WANG, Y., WEINER, M. P., YU, P., BEGLEY, R. F. & ROTHBERG, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-80.

MARSH, G. A., DE JONG, C., BARR, J. A., TACHEDJIAN, M., SMITH, C., MIDDLETON, D., YU, M., TODD, S., FOORD, A. J., HARING, V., PAYNE, J., ROBINSON, R., BROZ, I., CRAMERI, G., FIELD, H. E. & WANG, L. F. 2012. Cedar virus: a novel Henipavirus isolated from Australian bats. *PLoS Pathog,* 8**,** e1002836.

MARSH, G. A. & WANG, L.-F. 2012. Hendra and Nipah viruses: why are they so deadly? *Current Opinion in Virology,* 2**,** 242-247.

MAVRIDIS, L. & JANES, R. W. 2017. PDB2CD: a web-based application for the generation of circular dichroism spectra from protein atomic coordinates. *Bioinformatics,* 33**,** 56-63.

MCCORMICK, J. B., KING, I. J., WEBB, P. A., SCRIBNER, C. L., CRAVEN, R. B., JOHNSON, K. M., ELLIOTT, L. H. & BELMONT-WILLIAMS, R. 1986. Lassa fever. Effective therapy with ribavirin. *N Engl J Med,* 314**,** 20-6.

MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. 2007. Phaser crystallographic software. *Journal of Applied Crystallography,* 40**,** 658-674.

MCLAY, L., LIANG, Y. & LY, H. 2014. Comparative analysis of disease pathogenesis and molecular mechanisms of New World and Old World arenavirus infections. *J Gen Virol,* 95, 1-15.

MCLELLAN, J. S., CHEN, M., JOYCE, M. G., SASTRY, M., STEWART-JONES, G. B., YANG, Y., ZHANG, B., CHEN, L., SRIVATSAN, S., ZHENG, A., ZHOU, T., GRAEPEL, K. W., KUMAR, A., MOIN, S., BOYINGTON, J. C., CHUANG, G. Y., SOTO, C., BAXA, U., BAKKER, A. Q., SPITS, H., BEAUMONT, T., ZHENG, Z., XIA, N., KO, S. Y., TODD, J. P., RAO, S., GRAHAM, B. S. & KWONG, P. D. 2013. Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. *Science,* 342, 592-8.

MILAZZO, M. L., CAJIMAT, M. N., HAYNIE, M. L., ABBOTT, K. D., BRADLEY, R. D. & FULHORST, C. F. 2008. Diversity among tacaribe serocomplex viruses (family Arenaviridae) naturally associated with the white-throated woodrat (Neotoma albigula) in the southwestern United States. *Vector Borne Zoonotic Dis,* 8, 523-40.

MILAZZO, M. L., CAMPBELL, G. L. & FULHORST, C. F. 2011. Novel arenavirus infection in humans, United States. *Emerg Infect Dis,* 17, 1417-20.

MILLS, J. N., ELLIS, B. A., MCKEE, K. T., JR., KSIAZEK, T. G., ORO, J. G., MAIZTEGUI, J. I., CALDERON, G. E., PETERS, C. J. & CHILDS, J. E. 1991. Junin virus activity in rodents from endemic and nonendemic loci in central Argentina. *Am J Trop Med Hyg,* 44, 589-97.

MORAZ, M. L. & KUNZ, S. 2011. Pathogenesis of arenavirus hemorrhagic fevers. *Expert Rev Anti Infect Ther,* 9, 49-59.

MORENO, H., GALLEGO, I., SEVILLA, N., DE LA TORRE, J. C., DOMINGO, E. & MARTIN, V. 2011. Ribavirin can be mutagenic for arenaviruses. *J Virol,* 85, 7246-55.

MORSE, S. S., MAZET, J. A., WOOLHOUSE, M., PARRISH, C. R., CARROLL, D., KARESH, W. B., ZAMBRANA-TORRELIO, C., LIPKIN, W. I. & DASZAK, P. 2012. Prediction and prevention of the next pandemic zoonosis. *Lancet,* 380, 1956-65.

NEEDLEMAN, S. B. & WUNSCH, C. D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol,* 48, 443-53.

NEGRETE, O. A., LEVRONEY, E. L., AGUILAR, H. C., BERTOLOTTI-CIARLET, A., NAZARIAN, R., TAJYAR, S. & LEE, B. 2005. EphrinB2 is the entry receptor for Nipah virus, an emergent deadly paramyxovirus. *Nature,* 436, 401-5.

NEGRETE, O. A., WOLF, M. C., AGUILAR, H. C., ENTERLEIN, S., WANG, W., MUHLBERGER, E., SU, S. V., BERTOLOTTI-CIARLET, A., FLICK, R. & LEE, B. 2006. Two key residues in ephrinB3 are critical for its use as an alternative receptor for Nipah virus. *PLoS Pathog,* 2, e7.

NICHOL, S. T., ARIKAWA, J. & KAWAOKA, Y. 2000. Emerging viral diseases. *Proc Natl Acad Sci U S A,* 97, 12411-2.

NIKOLOV, D. B., LI, C., BARTON, W. A. & HIMANEN, J.-P. 2005. Crystal Structure of the Ephrin-B1 Ectodomain: Implications for Receptor Recognition and Signaling. *Biochemistry,* 44, 10947-10953.

NUNBERG, J. H. & YORK, J. 2012. The curious case of arenavirus entry, and its inhibition. *Viruses,* 4, 83-101.

PAGER, C. T., CRAFT, W. W., JR., PATCH, J. & DUTCH, R. E. 2006. A mature and fusogenic form of the Nipah virus fusion protein requires proteolytic processing by cathepsin L. *Virology,* 346, 251-7.

PAGER, C. T. & DUTCH, R. E. 2005. Cathepsin L is involved in proteolytic processing of the Hendra virus fusion protein. *J Virol,* 79, 12714-20.

PANTOPHLET, R., WILSON, I. A. & BURTON, D. R. 2003. Hyperglycosylated mutants of human immunodeficiency virus (HIV) type 1 monomeric gp120 as novel antigens for HIV vaccine design. *J Virol,* 77, 5889-901.

PARK, M. S., SHAW, M. L., MUNOZ-JORDAN, J., CROS, J. F., NAKAYA, T., BOUVIER, N., PALESE, P., GARCIA-SASTRE, A. & BASLER, C. F. 2003. Newcastle disease virus (NDV)-based assay demonstrates interferon-antagonist activity for the NDV V protein and the Nipah virus V, W, and C proteins. *J Virol,* 77, 1501-11.

PARSY, M. L., HARLOS, K., HUISKONEN, J. T. & BOWDEN, T. A. 2013. Crystal structure of Venezuelan hemorrhagic fever virus fusion glycoprotein reveals a class 1 postfusion architecture with extensive glycosylation. *J Virol,* 87, 13070-5.

PASQUALE, E. B. 2008. Eph-ephrin bidirectional signaling in physiology and disease. *Cell,* 133, 38-52.

PATCH, J. R., CRAMERI, G., WANG, L. F., EATON, B. T. & BRODER, C. C. 2007. Quantitative analysis of Nipah virus proteins released as virus-like particles reveals central role for the matrix protein. *Virol J,* 4, 1.

PAWESKA, J. T., SEWLALL, N. H., KSIAZEK, T. G., BLUMBERG, L. H., HALE, M. J., LIPKIN, W. I., WEYER, J., NICHOL, S. T., ROLLIN, P. E., MCMULLAN, L. K., PADDOCK, C. D., BRIESE, T., MNYALUZA, J., DINH, T. H., MUKONKA, V., CHING, P., DUSE, A., RICHARDS, G., DE JONG, G., COHEN, C., IKALAFENG, B., MUGERO, C., ASOMUGHA, C., MALOTLE, M. M., NTEO, D. M., MISIANI, E., SWANEPOEL, R., ZAKI, S. R., OUTBREAK, C. & INVESTIGATION, T. 2009. Nosocomial outbreak of novel arenavirus infection, southern Africa. *Emerg Infect Dis,* 15, 1598-602.

PEREZ, M., CRAVEN, R. C. & DE LA TORRE, J. C. 2003. The small RING finger protein Z drives arenavirus budding: implications for antiviral strategies. *Proc Natl Acad Sci U S A,* 100, 12978-83.

PERNET, O., SCHNEIDER, B. S., BEATY, S. M., LEBRETON, M., YUN, T. E., PARK, A., ZACHARIAH, T. T., BOWDEN, T. A., HITCHENS, P., RAMIREZ, C. M., DASZAK, P., MAZET, J., FREIBERG, A. N., WOLFE, N. D. & LEE, B. 2014. Evidence for henipavirus spillover into human populations in Africa. *Nat Commun,* 5, 5342.

PERNET, O., WANG, Y. E. & LEE, B. 2012. Henipavirus receptor usage and tropism. *Curr Top Microbiol Immunol,* 359, 59-78.

PETERS, C. J. 2006. Lymphocytic choriomeningitis virus--an old enemy up to new tricks. *N Engl J Med,* 354, 2208-11.

PINSCHEWER, D. D., PEREZ, M. & DE LA TORRE, J. C. 2005. Dual role of the lymphocytic choriomeningitis virus intergenic region in transcription termination and virus propagation. *J Virol,* 79, 4519-26.

POLIAKOV, A., COTRINA, M. & WILKINSON, D. G. 2004. Diverse roles of eph receptors and ephrins in the regulation of cell migration and tissue assembly. *Dev Cell,* 7, 465-80.

PRICE, W. N., 2ND, CHEN, Y., HANDELMAN, S. K., NEELY, H., MANOR, P., KARLIN, R., NAIR, R., LIU, J., BARAN, M., EVERETT, J., TONG, S. N., FOROUHAR, F., SWAMINATHAN, S. S., ACTON, T., XIAO, R., LUFT, J. R., LAURICELLA, A., DETITTA, G. T., ROST, B., MONTELIONE, G. T. & HUNT, J. F. 2009. Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. *Nat Biotechnol,* 27, 51-7.

PROVENCHER, S. W. & GLOECKNER, J. 1981. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry,* 20**,** 33-37.

PRYCE, R., NG, W. M., ZELTINA, A., WATANABE, Y., EL OMARI, K., WAGNER, A. & BOWDEN, T. A. 2018. Structure-based classification defines the discrete conformational classes adopted by the arenaviral GP1. *J Virol*.

RAABEN, M., JAE, L. T., HERBERT, A. S., KUEHNE, A. I., STUBBS, S. H., CHOU, Y. Y., BLOMEN, V. A., KIRCHHAUSEN, T., DYE, J. M., BRUMMELKAMP, T. R. & WHELAN, S. P. 2017. NRP2 and CD63 Are Host Factors for Lujo Virus Cell Entry. *Cell Host Microbe,* 22**,** 688-696 e5.

RADOSHITZKY, S. R., ABRAHAM, J., SPIROPOULOU, C. F., KUHN, J. H., NGUYEN, D., LI, W., NAGEL, J., SCHMIDT, P. J., NUNBERG, J. H., ANDREWS, N. C., FARZAN, M. & CHOE, H. 2007. Transferrin receptor 1 is a cellular receptor for New World haemorrhagic fever arenaviruses. *Nature,* 446**,** 92.

RADOSHITZKY, S. R., KUHN, J. H., SPIROPOULOU, C. F., ALBARINO, C. G., NGUYEN, D. P., SALAZAR-BRAVO, J., DORFMAN, T., LEE, A. S., WANG, E., ROSS, S. R., CHOE, H. & FARZAN, M. 2008. Receptor determinants of zoonotic transmission of New World hemorrhagic fever arenaviruses. *Proc Natl Acad Sci U S A,* 105**,** 2664-9.

RAHMAN, M. & CHAKRABORTY, A. 2012. Nipah virus outbreaks in Bangladesh: a deadly infectious disease. *WHO South East Asia J Public Health,* 1**,** 208-212.

RAMACHANDRAN, G. N., RAMAKRISHNAN, C. & SASISEKHARAN, V. 1963. Stereochemistry of polypeptide chain configurations. *J Mol Biol,* 7**,** 95-9.

REIGNIER, T., OLDENBURG, J., FLANAGAN, M. L., HAMILTON, G. A., MARTIN, V. K. & CANNON, P. M. 2008. Receptor use by the Whitewater Arroyo virus glycoprotein. *Virology,* 371**,** 439-46.

RHODES, G. 2006. *Crystallography made crystal clear : a guide for users of macromolecular models,* Amsterdam ; Oxford, Academic.

RISSANEN, I., AHMED, A. A., AZARM, K., BEATY, S., HONG, P., NAMBULLI, S., DUPREX, W. P., LEE, B. & BOWDEN, T. A. 2017. Idiosyncratic Mojiang virus attachment glycoprotein directs a host-cell entry pathway distinct from genetically related henipaviruses. *Nat Commun,* 8**,** 16060.

ROBINSON, J. E., HASTIE, K. M., CROSS, R. W., YENNI, R. E., ELLIOTT, D. H., ROUELLE, J. A., KANNADKA, C. B., SMIRA, A. A., GARRY, C. E., BRADLEY, B. T., YU, H., SHAFFER, J. G., BOISEN, M. L., HARTNETT, J. N., ZANDONATTI, M. A., ROWLAND, M. M., HEINRICH, M. L., MARTINEZ-SOBRIDO, L., CHENG, B., DE LA TORRE, J. C., ANDERSEN, K. G., GOBA, A., MOMOH, M., FULLAH, M., GBAKIE, M., KANNEH, L., KOROMA, V. J., FONNIE, R., JALLOH, S. C., KARGBO, B., VANDI, M. A., GBETUWA, M., IKPONMWOSA, O., ASOGUN, D. A., OKOKHERE, P. O., FOLLARIN, O. A., SCHIEFFELIN, J. S., PITTS, K. R., GEISBERT, J. B., KULAKOSKI, P. C., WILSON, R. B., HAPPI, C. T., SABETI, P. C., GEVAO, S. M., KHAN, S. H., GRANT, D. S., GEISBERT, T. W., SAPHIRE, E. O., BRANCO, L. M. & GARRY, R. F. 2016. Most neutralizing human monoclonal antibodies target novel epitopes requiring both Lassa virus glycoprotein subunits. *Nat Commun,* 7**,** 11544.

ROHWER, F., PRANGISHVILI, D. & LINDELL, D. 2009. Roles of viruses in the environment. *Environ Microbiol,* 11**,** 2771-4.

ROSENBERG, R. 2015. Detecting the emergence of novel, zoonotic viruses pathogenic to humans. *Cell Mol Life Sci,* 72**,** 1115-25.

RUGGIERO, H. A., PEREZ ISQUIERDO, F., MILANI, H. A., BARRI, A., VAL, A., MAGLIO, F., ASTARLOA, L., GONZALEZ CAMBACERES, C., MILANI, H. L. & TALLONE, J. C. 1986.

[Treatment of Argentine hemorrhagic fever with convalescent's plasma. 4433 cases]. *Presse Med,* 15**,** 2239-42.

RUSNAK, J. M., BYRNE, W. R., CHUNG, K. N., GIBBS, P. H., KIM, T. T., BOUDREAU, E. F., COSGRIFF, T., PITTMAN, P., KIM, K. Y., ERLICHMAN, M. S., REZVANI, D. F. & HUGGINS, J. W. 2009. Experience with intravenous ribavirin in the treatment of hemorrhagic fever with renal syndrome in Korea. *Antiviral Res,* 81**,** 68-76.

SATTERFIELD, B. A., CROSS, R. W., FENTON, K. A., AGANS, K. N., BASLER, C. F., GEISBERT, T. W. & MIRE, C. E. 2015. The immunomodulating V and W proteins of Nipah virus determine disease course. *Nat Commun,* 6**,** 7483.

SCHLIE, K., MAISA, A., FREIBERG, F., GROSETH, A., STRECKER, T. & GARTEN, W. 2010. Viral protein determinants of Lassa virus entry and release from polarized epithelial cells. *J Virol,* 84**,** 3178-88.

SCHNEIDER, T. R. & SHELDRICK, G. M. 2002. Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr,* 58**,** 1772-9.

SELVEY, L. A., WELLS, R. M., MCCORMACK, J. G., ANSFORD, A. J., MURRAY, K., ROGERS, R. J., LAVERCOMBE, P. S., SELLECK, P. & SHERIDAN, J. W. 1995. Infection of humans and horses by a newly described morbillivirus. *Med J Aust,* 162**,** 642-5.

SHAO, J., LIANG, Y. & LY, H. 2015. Human hemorrhagic Fever causing arenaviruses: molecular mechanisms contributing to virus virulence and disease pathogenesis. *Pathogens,* 4**,** 283-306.

SHAO, J., LIU, X., LY, H. & LIANG, Y. 2016. Characterization of the Glycoprotein Stable Signal Peptide in Mediating Pichinde Virus Replication and Virulence. *J Virol,* 90**,** 10390-10397.

SHIMON, A., SHANI, O. & DISKIN, R. 2017. Structural Basis for Receptor Selectivity by the Whitewater Arroyo Mammarenavirus. *J Mol Biol,* 429**,** 2825-2839.

SHTYRYA, Y. A., MOCHALOVA, L. V. & BOVIN, N. V. 2009. Influenza virus neuraminidase: structure and function. *Acta Naturae,* 1**,** 26-32.

SOMMERSTEIN, R., FLATZ, L., REMY, M. M., MALINGE, P., MAGISTRELLI, G., FISCHER, N., SAHIN, M., BERGTHALER, A., IGONET, S., TER MEULEN, J., RIGO, D., MEDA, P., RABAH, N., COUTARD, B., BOWDEN, T. A., LAMBERT, P. H., SIEGRIST, C. A. & PINSCHEWER, D. D. 2015. Arenavirus Glycan Shield Promotes Neutralizing Antibody Evasion and Protracted Infection. *PLoS Pathog,* 11**,** e1005276.

SPIROPOULOU, C. F., KUNZ, S., ROLLIN, P. E., CAMPBELL, K. P. & OLDSTONE, M. B. 2002. New World arenavirus clade C, but not clade A and B viruses, utilizes alpha-dystroglycan as its major receptor. *J Virol,* 76**,** 5140-6.

STEFFEN, D. L., XU, K., NIKOLOV, D. B. & BRODER, C. C. 2012. Henipavirus mediated membrane fusion, virus entry and targeted therapeutics. *Viruses,* 4**,** 280-308.

STEGMANN, T. 2000. Membrane fusion mechanisms: the influenza hemagglutinin paradigm and its implications for intracellular fusion. *Traffic,* 1**,** 598-604.

STUART, D. I., LEVINE, M., MUIRHEAD, H. & STAMMERS, D. K. 1979. Crystal structure of cat muscle pyruvate kinase at a resolution of 2.6 A. *J Mol Biol,* 134**,** 109-42.

SUTTLE, C. A. 2005. Viruses in the sea. *Nature,* 437**,** 356-61.

TERASAWA, K., TOMABECHI, Y., IKEDA, M., EHARA, H., KUKIMOTO-NIINO, M., WAKIYAMA, M., PODYMA-INOUE, K. A., RAJAPAKSHE, A. R., WATABE, T., SHIROUZU, M. & HARA-

YOKOYAMA, M. 2016. Lysosome-associated membrane proteins-1 and -2 (LAMP-1 and LAMP-2) assemble via distinct modes. *Biochem Biophys Res Commun,* 479, 489-495.

THOMAS, C. J., SHANKAR, S., CASQUILHO-GRAY, H. E., YORK, J., SPRANG, S. R. & NUNBERG, J. H. 2012. Biochemical reconstitution of hemorrhagic-fever arenavirus envelope glycoprotein-mediated membrane fusion. *PLoS One,* 7, e51114.

THOMAS, S. M., LAMB, R. A. & PATERSON, R. G. 1988. Two mRNAs that differ by two nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. *Cell,* 54, 891-902.

VIGERUST, D. J. & SHEPHERD, V. L. 2007. Virus glycosylation: role in virulence and immune interactions. *Trends Microbiol,* 15, 211-8.

VUILLARD, L., RABILLOUD, T., LEBERMAN, R., BERTHET-COLOMINAS, C. & CUSACK, S. 1994. A new additive for protein crystallization. *FEBS Lett,* 353, 294-6.

WAGNER, A., DUMAN, R., HENDERSON, K. & MYKHAYLYK, V. 2016. In-vacuum long-wavelength macromolecular crystallography. *Acta Crystallogr D Struct Biol,* 72, 430-9.

WALTER, T. S., DIPROSE, J. M., MAYO, C. J., SIEBOLD, C., PICKFORD, M. G., CARTER, L., SUTTON, G. C., BERROW, N. S., BROWN, J., BERRY, I. M., STEWART-JONES, G. B. E., GRIMES, J. M., STAMMERS, D. K., ESNOUF, R. M., JONES, E. Y., OWENS, R. J., STUART, D. I. & HARLOS, K. 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,* 61, 651-657.

WATANABE, Y., RAGHWANI, J., ALLEN, J. D., SEABRIGHT, G. E., LI, S., MOSER, F., HUISKONEN, J. T., STRECKER, T., BOWDEN, T. A. & CRISPIN, M. 2018. Structure of the Lassa virus glycan shield provides a model for immunological resistance. *Proc Natl Acad Sci U S A,* 115, 7320-7325.

WEISS, S., NOWAK, K., FAHR, J., WIBBELT, G., MOMBOULI, J. V., PARRA, H. J., WOLFE, N. D., SCHNEIDER, B. S. & LEENDERTZ, F. H. 2012. Henipavirus-related sequences in fruit bat bushmeat, Republic of Congo. *Emerg Infect Dis,* 18, 1536-7.

WHITE, J. M., DELOS, S. E., BRECHER, M. & SCHORNBERG, K. 2008. Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Crit Rev Biochem Mol Biol,* 43, 189-219.

WHO. 2018. *Disease outbreak news* [Online]. WHO. Available: http://www.who.int/csr/don/archive/year/2018/en/ [Accessed 24/09/2018 2018].

WILEN, C. B., TILTON, J. C. & DOMS, R. W. 2012. HIV: cell binding and entry. *Cold Spring Harb Perspect Med,* 2.

WINTER, G. 2010. xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography,* 43, 186-190.

WOLFE, N. D., DUNAVAN, C. P. & DIAMOND, J. 2007. Origins of major human infectious diseases. *Nature,* 447, 279-83.

WU, Z., YANG, L., YANG, F., REN, X., JIANG, J., DONG, J., SUN, L., ZHU, Y., ZHOU, H. & JIN, Q. 2014. Novel Henipa-like virus, Mojiang Paramyxovirus, in rats, China, 2012. *Emerg Infect Dis,* 20, 1064-6.

XU, K., CHAN, Y. P., BRADEL-TRETHEWAY, B., AKYOL-ATAMAN, Z., ZHU, Y., DUTTA, S., YAN, L., FENG, Y., WANG, L. F., SKINIOTIS, G., LEE, B., ZHOU, Z. H., BRODER, C. C., AGUILAR, H. C.

& NIKOLOV, D. B. 2015. Crystal Structure of the Pre-fusion Nipah Virus Fusion Glycoprotein Reveals a Novel Hexamer-of-Trimers Assembly. *PLoS Pathog,* 11, e1005322.

XU, K., RAJASHANKAR, K. R., CHAN, Y. P., HIMANEN, J. P., BRODER, C. C. & NIKOLOV, D. B. 2008. Host cell recognition by the henipaviruses: crystal structures of the Nipah G attachment glycoprotein and its complex with ephrin-B3. *Proc Natl Acad Sci U S A,* 105, 9953-8.

XU, R. & WILSON, I. A. 2011. Structural characterization of an early fusion intermediate of influenza virus hemagglutinin. *J Virol,* 85, 5172-82.

YAMAUCHI, Y. & HELENIUS, A. 2013. Virus entry at a glance. *J Cell Sci,* 126, 1289-95.

YANG, Z. R., THOMSON, R., MCNEIL, P. & ESNOUF, R. M. 2005. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics,* 21, 3369-76.

YORK, J. & NUNBERG, J. H. 2016. Myristoylation of the Arenavirus Envelope Glycoprotein Stable Signal Peptide Is Critical for Membrane Fusion but Dispensable for Virion Morphogenesis. *J Virol,* 90, 8341-50.

YUAN, P., SWANSON, K. A., LESER, G. P., PATERSON, R. G., LAMB, R. A. & JARDETZKY, T. S. 2011. Structure of the Newcastle disease virus hemagglutinin-neuraminidase (HN) ectodomain reveals a four-helix bundle stalk. *Proc Natl Acad Sci U S A,* 108, 14920-5.

YUAN, P., THOMPSON, T. B., WURZBURG, B. A., PATERSON, R. G., LAMB, R. A. & JARDETZKY, T. S. 2005. Structural studies of the parainfluenza virus 5 hemagglutinin-neuraminidase tetramer in complex with its receptor, sialyllactose. *Structure,* 13, 803-15.

ZELTINA, A., KRUMM, S. A., SAHIN, M., STRUWE, W. B., HARLOS, K., NUNBERG, J. H., CRISPIN, M., PINSCHEWER, D. D., DOORES, K. J. & BOWDEN, T. A. 2017. Convergent immunological solutions to Argentine hemorrhagic fever virus neutralization. *Proc Natl Acad Sci U S A,* 114, 7031-7036.

ZHAO, Y., REN, J., HARLOS, K., JONES, D. M., ZELTINA, A., BOWDEN, T. A., PADILLA-PARRA, S., FRY, E. E. & STUART, D. I. 2016. Toremifene interacts with and destabilizes the Ebola virus glycoprotein. *Nature,* 535, 169-172.

ZONG, M., FOFANA, I. & CHOE, H. 2014. Human and host species transferrin receptor 1 use by North American arenaviruses. *J Virol,* 88, 9418-28.

Appendix C:

List of Abbreviations

Appendix C: List of Abbreviations

Å: Ångströms = 10^{-10} metres

α-DG: Alpha dystroglycan

CAPV: Chapare virus

CDC: Center for disease control

CedPV: Cedar paramyxovirus

Cryo-EM: Cryogenic electron-microscopy

Da: Daltons

DMEM: Dulbecco's Modified Eagle's Medium

DNA: Deoxyribonucleic Acid

***E. coli*:** *Escherichia coli*

EBOV: Ebola virus

EDTA: Ethylenediaminetetraacetic acid

EndoF1: Endoglycosidase-F1

F: Fusion protein (Paramyxovirus)

Fab: 'Fragment antigen-binding': the antigen binding region of an antibody

FCS: Fetal Calf Serum

G: Attachment Glycoprotein (Paramyxovirus)

g: Gram

GhV: Ghanaian henipavirus (aka Kumasi virus)

GlcNAc: N-acetyl glucosamine

GP1: Glycoprotein 1 (Arenavirus attachment glycoprotein)

GP2: Glycoprotein 2 (Arenavirus fusion glycoprotein)

GPC: Glycoprotein Complex (Arenavirus glycoprotein spike complex)

GTOV: Guanarito virus

H: Hemagglutinin

HEK293T: Human Embryonic Kidney 293T cells

HeV: Hendra virus

HIV: Human immunodeficiency virus

HN: Hemagglutinin Neuraminidase

JUNV: Junín virus

K: Kelvin

k: Kilo (10^3)

Kb: Kilobase

L: Litre

LB: Luria Broth

LCMV: Lymphocitic choriomeningitis virus

LORV: Loei river virus

LUJV: Lujo virus

µ: Micro (10^{-6})

m: Milli (10^{-3})

Man: Mannose

MERS-CoV: Middle East respiratory syndrome virus

Min: Minutes

MOJV: Mójiāng virus

MORV: Morogoro virus

MS: Mass spectrometry

MWCO: Molecular Weight Cut-Off

n.c.s: Non-crystallographic symmetry

NaCl: Sodium Chloride

NEAA: Non-Essential Amino Acids

NiV: Nipah virus

NW: New World

OW: Old World

PCR: Polymerase Chain Reaction

PDB: Protein Data Bank

PEI: Polyethylenimine

RMSD: Root mean square deviation

RNA: Ribonucleic acid

SABV: Sabia virus

SDS-PAGE: Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

Sec: Seconds

SEC: Size Exclusion Chromatography

SOC: Super Optimal Broth

SSP: Stable Signal Peptide (Arenavirus)

TAE: Tris Acetic acid EDTA

TAMV: Tamiami virus

TfR1: Transferrin receptor-1

T_m: Melting temperature (of DNA duplexes)

Tris: Tris(hydroxymethyl)aminomethane

UV: Ultra Violet (light)

WENV: Wēnzhōu virus

WWAV: Whitewater Arroyo virus

x g: Multiples of the standard gravity of earth, 9.8 ms^{-1}

XF-MS: X-ray footprinting mass spectrometry

Standard one-letter amino acid abbreviations are used throughout.