Structural and Cellular Biology of the Nairovirus Nucleocapsid Protein

Rebecca Anna Surtees

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

The University of Leeds

The Astbury Centre for Structural Molecular Biology

June 2014

ACKNOWLEDGEMENTS

Chapter 3: Stephen Carter picked crystals and collected data for the RNA-bound HAZV N structure presented in this chapter. Dr Chi Trinh picked crystals and helped collect data for the Apo-HAZV N structure presented in this chapter. Dr Antonio Ariza refined both datasets, and helped solve the structure of Apo-HAZV N by molecular replacement.

Chapter 4: Mass spectrometry analysis of the SILAC immunoprecipitation data was performed by Dundee Cell Products. Mass spectrometry analysis of HAZV virions was carried out by Dr Stuart Armstrong at the University of Liverpool. All experiments involving actively replicating CCHFV were carried out by Dr Stuart Dowall, of Public Health England, at Porton Down.

I would primarily like to thank John and Ed for all their invaluable help and guidance. Thank you for always supporting me, and encouraging me to go that little bit further – I couldn't have done it without you! I would also like to thank everybody in Virology at Leeds, especially in Garstang 8.61, for all of their help, support and advice. Thank you to Chi, Ed, and Tony for teaching me about crystallography, and to Gareth Howall for all his advice about confocal microscopy. Thank you to Roger Hewson and Stuart Dowall for all of their enthusiasm, support and help, and for supplying me with reagents, experiments and ideas. I am also grateful to my funding bodies, the BBSRC and Public Health England for agreeing to fund this project. Thank you to my friends and family for their unwavering support, and to Leo, who always believed in me and has spent an inordinate amount of time over the past 4 years waiting for me, and driving me to and from the lab at every hour of the day and night.

ABSTRACT

Crimean Congo haemorrhagic fever virus (CCHFV) is a tick borne nairovirus responsible for a highly fatal human disease with increased recent incidence within the human population of southern Europe. Hazara virus (HAZV), another tick borne nairovirus of the same serogroup as CCHFV, is not pathogenic in humans, but infection with HAZV results in a similar disease progression in interferon receptor knockout mice to CCHFV. In order to investigate further similarities between HAZV and CCHFV, and to support the use of HAZV as a model for CCHFV infection, various structural and cell biological aspects of the CCHFV and HAZV replication cycle were investigated. We present the X-ray crystal structure of the nucleocapsid protein (N) of HAZV at a resolution of 2.7 Å, and the crystal structure of RNA-bound HAZV N at a resolution of 3 Å. HAZV N had a highly similar overall structure to CCHFV N, consisting of a globular domain containing amino acid residues from both the N and the C-termini, and an extended arm domain. A crevice within the globular domain of HAZV N had 3 nucleotides of RNA bound, in the RNA-bound HAZV N structure. RNA binding is a crucial function of HAZV N and CCHFV N, and the identification of an RNA binding site in HAZV N provides a model for the structure guided design of anti-virals to disrupt this interaction.

Cellular heat shock protein 70 (HSP70) was found to be an interacting partner of the N protein of CCHFV and HAZV, both in cells and in secreted HAZV particles. The functional role of this interacting partner within the HAZV replication cycle was investigated using siRNA knock down and small molecule inhibitors of endogenous HSP70. HAZV induced apoptosis in SW13 cells, and HAZV N was cleaved during a 72 hr time course of infection. The data presented here also reveals new insights into the replication cycles of HAZV and CCHFV.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ABBREVIATIONS

AU asymmetric unit

BSA bovine serum albumin

BUNV bunyamwera virus

CCHF Crimean congo haemorrhagic fever

CCHFV Crimean congo haemorrhagic fever virus

cRNA viral RNA complimentary to genomic RNA

DAPI 4',6-diamidino-2-phenylindole

DMEM Dulbecco’s modified eagles medium

DMSO dimethyl sulphoxide

dsRNA double stranded RNA

EGFP enhanced green fluorescent protein

ER endoplasmic reticulum

FBS foetal bovine serum

G_C virus glycoprotein C terminal fragment

GFP green fluorescent protein

G_N virus glycoprotein N terminal fragment

GST glutathione S-transferase

HAZV Hazara virus

HCV hepatitis C virus

HRV human rhinovirus

HSC70 heat shock cognate 70

HSP70 heat shock protein 70

HSP90 heat shock protein 90

HSV herpes simplex virus

IAV influenza A virus

ICTV International Committee on Taxonomy of Viruses

IF immunofluorescence

IFN interferon

IFN-α/βR$^{-/-}$ interferon α/β (type 1) receptor knockout

IL interleukin

IP immunoprecipitation

IPTG isopropyl β-D-1-thiogalactopyranoside

IRF-3 interferon regulatory transcription factor 3

ISG15 interferon stimulated gene 15

L segment large segment

LACV La Crosse virus

LASV Lassa virus

M segment medium segment

MAP mix after precipitation

MOI multiplicity of infection

MR molecular replacement

MS mass spectrometry

MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide)

MW molecular weight

N nucleocapsid protein

nsNSV non-segmented negative-strand RNA virus

NTR non-translated region

ORF open reading frame

OTU ovarian tumor protease-like domain

PAMP pathogen associated molecular pattern

PARP poly(ADP) ribose polymerase

PDB Protein Data Bank

PEG polyethylene glycol

pfu plaque forming units

PIF pifithrin-µ

RdRp (L protein) RNA dependent RNA polymerase

RMSD Cα root mean squared deviation

RNP ribonucleoprotein

RSV respiratory syncytial virus

RT-PCR real time polymerase chain reaction

RV rabies virus

RVFV Rift Valley fever virus

S segment small segment

SDS PAGE sodium dodecyl sulphate polyacrylamide gel electrophoresis

SFM serum- and antibiotic-free media

SFTSV severe fever with thrombocytopenia syndrome virus

SILAC stable isotope labelling of amino acids in cell culture

siRNA small interfering RNA

sNSV segmented negative-strand RNA virus

ssRNA single stranded RNA

STAT1 signal transducer and activator of transcription 1 protein

TNF tumor necrosis factor

TUNEL terminal dUTP nick-end labelling

UPR unfolded protein response

VER VER155008

WHO world health organisation

WT wild type

CHAPTER 1: INTRODUCTION

1.1 GENERAL INTRODUCTION

1.1.1 Discovery of Crimean Congo haemorrhagic fever virus and Hazara virus

Crimean Congo haemorrhagic fever (CCHF) was first described as a new tick borne disease during a large outbreak of severe haemorrhagic fever in Crimea in 1945 when approximately 200 patients were infected (M. P. Chumakov 1945). The disease was named Crimean haemorrhagic fever (CHF), and a viral aetiology was suggested after infecting psychiatric patients with a filterable agent isolated from the blood of CCHF patients (Whitehouse 2004; Bente, Forrester et al. 2013). The virus was first isolated through inoculation of new born white mice with samples from a CCHF patient in Astrakhan in 1967 and the resulting virus strain (Drozdov) enabled researchers to develop the necessary reagents (e.g antigens and antibodies) for subsequent serological surveys and investigations of virus isolates from several different geographic regions (Chumakov, Butenko et al. 1968; Whitehouse 2004). In 1969 it was shown that the virus isolated in Crimea in 1945 was antigenically indistinguishable from a virus isolated from human patients from the Congo and Uganda in 1956, and the virus was renamed Crimean Congo haemorrhagic fever virus (CCHFV) (Hoogstraal 1979; Simpson, Knight et al. 1967; Woodall, Williams et al. 1967; Casals 1969).

The outbreak of CCHF in Crimea in 1945 was the first full medical description of this disease that brought it to the attention of modern medical science. However references have been found in several scholarly texts which date from as far back as the 12th century to a human disease circulating in the region that is present day Tajikistan. This disease presented with haemorrhagic symptoms (the presence of blood in the urine, vomit, gums, rectum, and sputum) and was associated with ticks (Semiatkovskaia and Sidtdykova 1950; Shapiro and Barkagan 1960; Hoogstraal 1979, Whitehouse 2004). In southern Uzbekistan a disease that is likely to be CCHFV has been known for centuries under several different names including *khungribta* (blood taking) and *khunymuny* (nose bleeding) (Chumakov 1974; Whitehouse 2004). It is therefore probable that CCHFV has been circulating in human populations in Asia for centuries. However, the recent expansion of the geographic range of this virus, and the increase in the number of countries in which CCHFV is now considered to be endemic, is a cause for concern. As a result CCHFV is now described as an emerging virus.

Hazara virus (HAZV) was first isolated from *Ixodes redikorzevi* ticks collected from a mountain vole (*Alticola roylei*) in the Kaghan Valley, Hazara District, Western Pakistan in 1964 (Begum, Wisseman et al. 1970; Begum, Wisseman et al. 1970). This strain of HAZV (JC280) was isolated in suckling mice, and antigen and immune serum from these mice were shown to cross react strongly in neutralisation assays with CCHFV, but not in the haemagglutination inhibition or complement fixation assays (Begum, Wisseman et al. 1970), indicating that HAZV was serologically related to, but a distinct virus from CCHFV. Similar to CCHFV, HAZV was lethal when suckling mice were inoculated intracerebrally, with the highest viral loads detected in the brain and liver (Smirnova, Shestopalova et al. 1977). In contrast to CCHFV, HAZV has not been documented to infect or cause disease in humans and current knowledge regarding the global distribution of HAZV is limited.

1.1.2 Classification

CCHFV and HAZV are both members of the *Bunyaviridae* family (ICTV), which together with members of the *Arenaviridae* and *Orthomyxoviridae* family are collectively known as the segmented negative-strand RNA viruses (sNSVs) by virtue of their multiple genome segments. The *Bunyaviridae* family is one of the largest virus families, containing over 350 named species, which are classified into five genera, the *Orthobunyaviruses, Tospoviruses, Phleboviruses, Hantaviruses* and *Nairoviruses*. CCHFV and HAZV both belong in the *Nairovirus* genus. Defining features of members of the *Bunyaviridae* family include the possession of a segmented genome containing three RNA segments of negative sense or ambi-sense polarity, a cytoplasmic replication strategy and a spherical virion morphology of 90-120 nm in diameter (Walter and Barr 2011). The three genomic segments are termed small (S), medium (M) and large (L), based on their relative sizes, and bunyaviruses share fundamental mechanisms of RNA synthesis (Walter and Barr 2011).

Viruses within the *Bunyaviridae* family are arthropod and rodent-borne, infecting animals, plants, humans, and insects. Viruses within the *Tosposvirus* genus infect plants and are predominantly transmitted by thrips, and viruses within the *Hantavirus* genus are transmitted by rodents; humans become infected primarily through inhalation of aerosolised virus particles present in rodent excrement and saliva (Song, Baek et al. 2007). Viruses in the *Orthobunyavirus, Phlebovirus* and *Nairovirus* genus are arthropod-borne and are transmitted by ticks, mosquitos, midges and sandflies (Walter and Barr 2011).

Noteworthy human pathogens of the *Orthobunyavirus* genus include Oropouche virus, which causes Oropouche fever, a febrile disease often associated with headache, dizziness, weakness, myalgias (muscle pain) and arthralgias (joint pain) (Pinheiro, Travassos da Rosa et al. 1981; Aguilar, Barrett et al. 2011). Oropouche virus (OROV) is transmitted by mosquitos and midges and is widespread in South America, with >30 epidemics reported between 1960-2009, causing significant morbidity with over half a million people estimated to have been infected (Vasconcelos, Nunes et al. 2011). Bunyamwera virus (BUNV), the prototypic member of the *Orthobunyaviruses* also causes febrile illness, the symptoms of which include headache, arthalgias, rash, and infrequent nervous system involvement (Aguilar, Barrett et al. 2011). La Crosse virus is the leading cause of paediatric encephalitis in North America, and Tahyna virus causes an influenza-like syndrome in central Europe (McJunkin, de los Reyes et al. 2001; Gould, Higgs et al. 2006).

In the *Phlebovirus* genus, Rift Valley fever virus (RVFV) is a human pathogen endemic to sub-Saharan Africa, generally causing an influenza-like illness; however in ~1% of cases patients develop haemorrhagic manifestations, which have a 50% case fatality rate (Boshra, Lorenzo et al. 2011; Raymond, Piper et al. 2012). Various member of the *Hantavirus* genus are known to be pathogenic to humans (Hantaan virus, Seoul virus, Puumala virus, Sin nombre virus) causing two distinct disease syndromes: haemorrhagic fever with renal syndrome (HFRS) and hantavirus cardio-pulmonary syndrome (HCPS) (Manigold and Vial 2014, Walter and Barr 2011). Members of the *Tospovirus* genus (e.g Tomato spotted wilt virus, the most important plant virus in economic terms) cause significant losses in yield and quality of produce of legume, vegetable and ornamental crops from many different countries in Africa, Asia, North America, Europe and Australasia (Pappu, Jones et al. 2009).

Within the *Nairovirus* genus, the most significant human pathogen is CCHFV. The *Nairovirus* genus is separated from other bunyavirus genera by their large L segments and includes seven different serogroups of tick transmitted viruses, which contain 43 virus strains; the CCHF serogroup contains only CCHFV and HAZV (Bente, Forrester et al. 2013, Whitehouse 2004, ICTV). The other *Nairovirus* serogroup containing human pathogens is the Nairobi sheep disease group, which includes Nairobi sheep disease virus; this is primarily a pathogen of sheep and goats, causing severe haemorrhagic gastroenteritis (but has been known to infect humans), and Dugbe virus, which causes a mild febrile illness and thrombocytopenia in humans (Burt, Spencer et al. 1996, Marczinke and Nichol 2002).

Individual viruses are classified into the 5 genera within the *Bunyaviridae* family based on similar segment lengths and protein sizes, the presence of conserved nucleotide sequences within each segment termini and the segment coding strategy used to express viral proteins. For example members of the *Phlebovirus* genus encode a non-structural protein (NSs) generated by an ambisense coding strategy of the S segment, whereas members of the *Nairovirus* genus do not encode an NSs protein (Walter and Barr 2011).

1.1.3 The CCHFV Tick vector and mammalian host life cycle

CCHFV is a zoonotic virus, in nature CCHFV cycles between tick and vertebrate hosts; humans become infected when they are bitten by an infected tick, or come into contact with the blood of an infected animal or human (Figure 1.2). CCHFV is maintained within several species of *Ixodes* (hard) ticks (*Hyalomma, Rhipicephalus* and *Dermacentor* ticks), although ticks of the *Hyalomma* genus (Figure 1.1) are thought to be the principle source of human infection; indeed the global distribution of *Hyalomma* ticks closely mirrors that of CCHF incidence (Bente, Forrester et al. 2013; Yesilbag, Aydin et al. 2013).

Figure 1.1 Female Hyalomma dromedarii tick

Ticks of the Hyalomma species are thought to be the principle source of human CCHFV infection. Photograph courtesy of Alan Walker.

Ticks become infected with CCHFV when they take a blood meal to acquire the nutrients they need for metamorphosis from larvae to nymph to adult; ticks need to acquire these nutrients once during every developmental stage. In addition, adult female ticks also need a blood meal to acquire the nutrients required to produce eggs (Bente, Forrester et al. 2013). Once a tick has become infected with CCHFV it remains infected throughout its lifetime; CCHFV is maintained throughout metamorphosis from larvae to nymph to adult and is transmitted to other ticks through trans-ovarial (passed from adult females into their eggs), and venereal routes (transmitted through copulation) (Dohm, Logan et al. 1996). In addition, CCHFV can also spread directly to other ticks through the mechanism of co-feeding, which involves the transference of CCHFV from an infected tick feeding on a vertebrate host directly through their saliva to another tick feeding nearby on the same infected host (Jones, Davies et al. 1987; Nuttall and Labuda 2003; Nuttall and Labuda 2004). This is also known as 'non-viraemic' transmission, and enhances the efficiency of virus spread since it does not require CCHFV to reach the certain titre in infected vertebrates required to initiate infection in ticks (a factor that varies among different ticks hosts), and overcomes the transmission barrier due to the fact that not all the mammalian hosts that ticks feed on are susceptible to CCHFV infection (Bente, Forrester et al. 2013).

Several members of the *Hyalomma* genus are two-host ticks; they feed on small animal hosts (generally hares, ground-feeding birds, and rodents) as larvae and nymphs; adult ticks feed on large animal hosts (both domestic animals such as cattle and sheep, and wild animals such as deer) and mate while attached to the host (Figure 1.2). When mammalian hosts become infected with CCHFV they display a transient, asymptomatic viraemia which is resolved in 1-2 weeks (Bente, Forrester et al. 2013). CCHFV can replicate in a wide variety of wild and domestic animals (horses, donkeys, sheep, cattle, scrub hares, mice, rats and ostriches) in the absence of any clear signs of illness (Smirnova 1979, Bente, Forrester et al. 2013). CCHFV is capable of being transmitted from ticks to mammals; naïve ticks can then become infected when they feed on the blood of a viraemic animal (Figure 1.2). Humans however, are not a source of infection for ticks and are therefore thought to be only accidental or 'dead-end' hosts for the virus (Bente, Forrester et al. 2013). Ticks are considered to be the true natural reservoirs for CCHFV as ticks remain infected with CCHFV throughout their several-year long life spans.

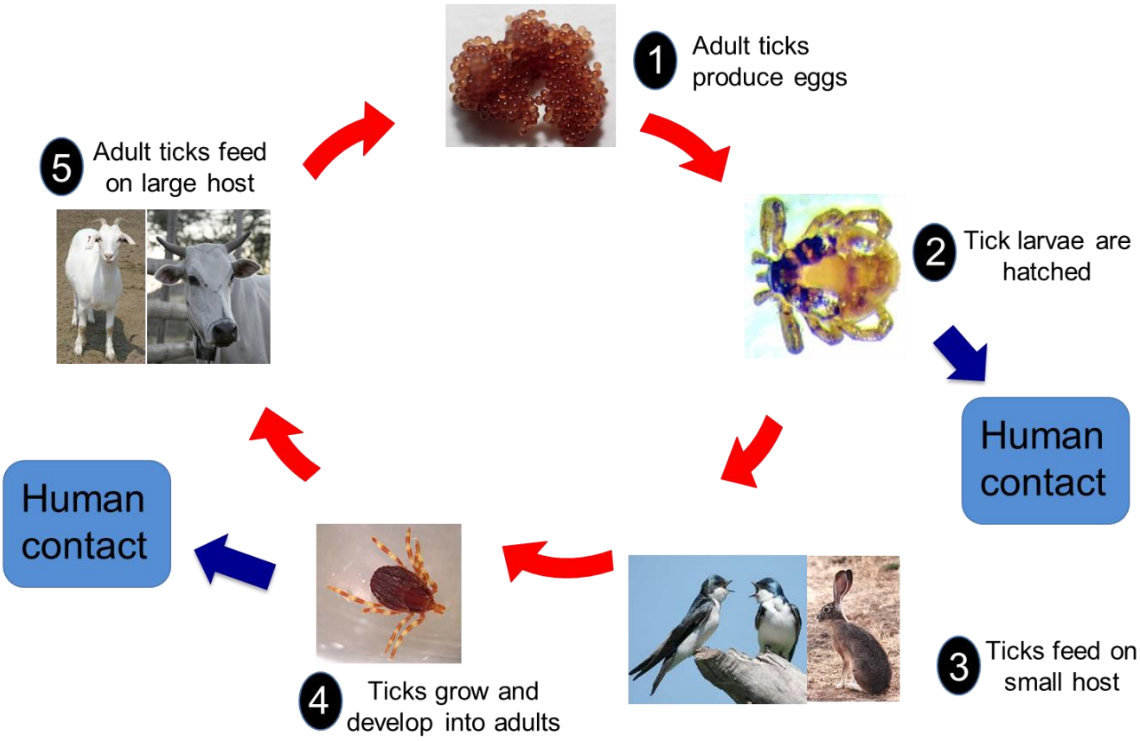


Figure 1.2 The natural life cycle of Hyalomma ticks

The life cycle of Hyalomma spp. requires both a small animal host (3) and a large animal host (5). A CCHFV infected tick maintains CCHFV replication for the duration of its life cycle, and females transmit CCHFV to their offspring transovarially. Humans become infected with CCHFV when they are bitten by an infected tick at any stage of the tick lifecycle, however questing adult Hyalomma ticks are considered to be the main source of human infection.

CCHFV has also been detected in other arthropod vectors such as biting midges, however this probably reflects the recent feeding of the biting midge on a CCHFV infected mammalian host, as there is no evidence that CCHFV is maintained throughout the lifecycle of any other arthropod host, or that any other arthropod apart from an *Ixodid* tick is a competent vector (Causey, Kemp et al. 1970).

1.1.4 HAZV mammalian hosts

Although the natural mammalian hosts of HAZV are largely unknown, antibodies against HAZV have been detected in wild rodent sera (Darwish, Hoogstraal et al. 1983). In a similar manner to CCHFV, experimentally HAZV has been shown to replicate in a variety of mammals including various species of mice and rats, guinea pigs, rabbits, and donkeys, amongst others (Smirnova 1979). In the case of both CCHFV and HAZV, the only animals that display clinical symptoms and die are suckling mice (Smirnova, Shestopalova et al. 1977; Bente, Forrester et al. 2013). Excluding the *Ixodes redikorzevi* ticks from which HAZV was first isolated, there have been no further reports of other tick species maintaining HAZV replication.

1.1.5 Epidemiology and Geographic range of CCHFV

CCHF is the most geographically widespread tick borne human viral disease and CCHFV is the second most widespread of all medically important arboviruses (insect-borne viruses) after Dengue virus (Appannanavar and Mishra 2011). The known geographic distribution of CCHFV now ranges from western China, throughout Asia and the Middle East, across South Eastern Europe and throughout most of Africa (Mild, Simon et al. 2010). After CCHFV was first described in Crimea in 1945, the majority of reported cases initially originated from the former Soviet Union and Bulgaria (Appannanavar and Mishra 2011). Outbreaks of CCHF were then reported from several African countries such as the Democratic Republic of the Congo, Uganda and Mauritania. Middle Eastern countries such as Iraq, the United Arab Emirates and Saudi Arabia also reported a significant number of cases, and more recently, a large number of CCHF cases have been reported from Pakistan, Iran, Bulgaria and Turkey (Appannanavar and Mishra 2011).

Currently 52 countries are recognised as endemic or potentially endemic regions; most endemic countries report a number of human CCHF cases every year; some only report evidence of CCHFV circulation through the recovery of CCHFV RNA, or live virus, from wild or domestic animals, or from ticks (Appannanavar and Mishra 2011; Bente, Forrester et al. 2013). In some countries evidence of CCHFV circulation is based on detection of CCHFV specific antibodies in serosurveys of humans and animals. Figure 1.3 is taken from the World Health Organisation (WHO) website and shows the known global distribution of CCHFV in 2008 as well as the underlying evidence for the circulation of CCHFV in these regions.

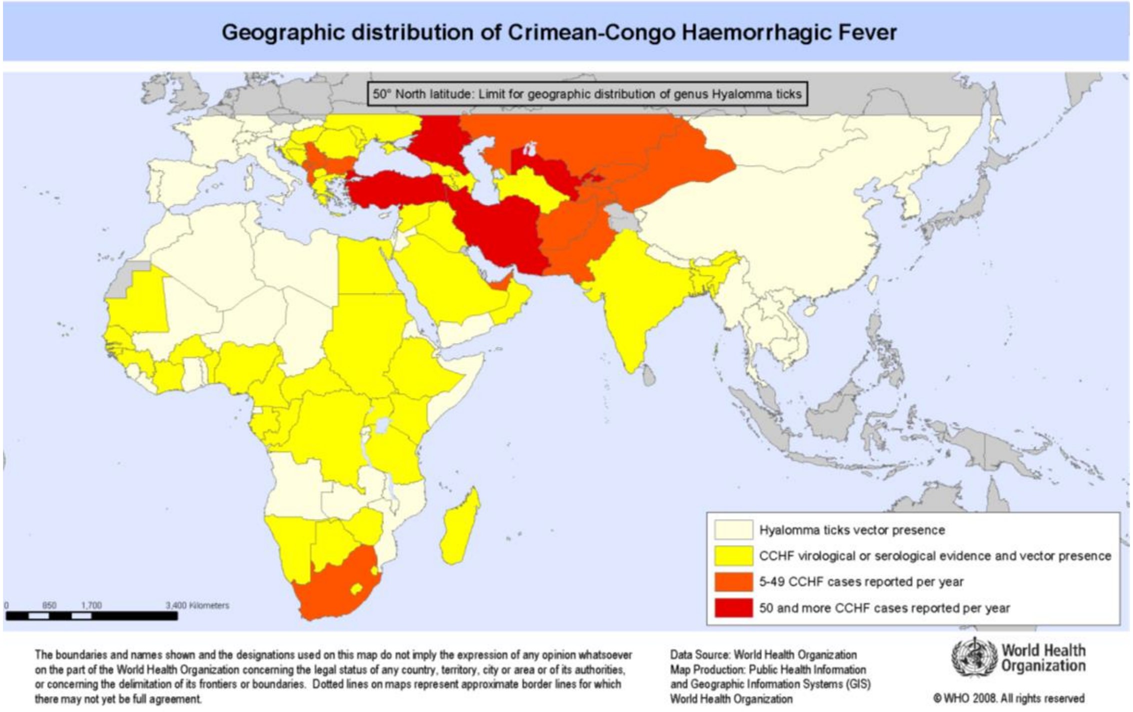


Figure 1.3 Geographic distribution of CCHFV

More than 50 human cases of CCHF are reported annually to the WHO by countries coloured red, whereas fewer than 50 annual cases are reported by countries coloured orange. Human cases of CCHFV have not been reported by countries coloured yellow, but CCHFV has been isolated from ticks or animals in these countries, or the presence of CCHFV has been inferred from serological studies, and a competent CCHFV tick vector has also been detected. The northernmost limit of Hyalomma marginatum and Hyalomma asiaticum tick species is indicated. The figure is taken from the World Health Organisation (WHO) website (http://www.who.int/csr/disease/crimean_ congoHF).

CCHF is considered to be an emerging disease due to a recent increase in both the number of cases of CCHF in countries where CCHF is endemic, and an increase in the number of countries where CCHFV is now considered to be endemic. The first reported case of CCHF in Turkey was documented in 2002 and since then there have been over 6000 reported cases, representing a special case of CCHF epidemiology (Karti, Odabasi et al. 2004, Bente, Forrester et al. 2013). The first CCHF case was reported in Greece in 2008; a person who was living close to the Bulgarian border was bitten by a tick and subsequently developed a haemorrhagic fever (Papa, Dalla et al. 2010) and more recently CCHFV RNA was isolated from ticks in Spain in 2010

(Estrada-Pena, Palomar et al. 2012). The CCHFV tick vector (adult *Hyalomma marginatum*) has been identified on livestock in Hungary and Romania, demonstrating these ticks can survive and replicate in more northern regions than had previously been documented, and in 2013 ticks of the *Hyalomma* species were identified in Cornwall in the UK (Jameson and Medlock 2011; Ceianu, Panculescu-Gatej et al. 2012; Hornok and Horvath 2012; Bente, Forrester et al. 2013).

The expanding geographic distribution of the CCHFV tick vector increases the possibility that the geographic range of the disease might also expand as the global distribution of CCHFV closely mirrors that of its primary vector. People most at risk of developing CCHFV include those with an increased risk of exposure to tick bites, such as agricultural workers and foresters, as well as abattoir workers, veterinarians and hospital staff, who come into contact with the blood of potentially infected animals or humans.

The majority of reported cases of CCHFV occur as sporadic, isolated incidents, however occasionally larger outbreaks occur, such as the infection of 200 Soviet soldiers in 1944 (M. P. Chumakov 1945). Nosocomial outbreaks of infection in hospital workers who do not employ appropriate barrier nursing methods are also common in countries where CCHFV is endemic. This can be illustrated by the central government hospital in Pakistan, where a nosocomial outbreak of CCHF occurred in 1976 (Hoogstraal 1979; Burney, Ghafoor et al. 1980). A shepherd was admitted displaying typical CCHF symptoms and was operated on the day of his admission, but died later that night. His father, who was caring for him at home, died of CCHF two days later despite intensive care and repeated blood transfusions. Of the 12 hospital staff who came into contact with, or treated the shepherd, 10 became ill with CCHF and were hospitalised, two of these died (a nurse who assisted during the operation died three days later, and the surgeon, who cut his finger during the operation, died two weeks later) and 8 eventually recovered after becoming severely ill (including the doctor who admitted the shepherd after he vomited blood on her hands and face, and an assistant surgeon who pricked his finger whilst operating) (Whitehouse 2004). Incidences such as this highlight the threat posed to medical personnel by CCHFV and illustrate why CCHF is considered to be such a public health problem in countries where the virus is endemic.

Excluding the *Ixodes redikorzevi* ticks found in the Kaghan valley in Western Pakistan from which HAZV was originally isolated, and the anti-HAZV antibodies detected in the sera of a wild jird (a gerbil-like mammal) found in the lowland, semi-arid environment of Pakistan, there have been no further reports of HAZV being isolated from any other country. Knowledge regarding the known geographic distribution of HAZV therefore remains limited.

1.1.6 Genetic diversity and evolution

Isolates of CCHFV show great genetic diversity, ranging from 20 % nucleotide sequence differences between viral S segments, to 31 % nucleotide sequence differences between viral M segments (Bente, Forrester et al. 2013). Several processes are thought to contribute to the present genetic diversity of CCHFV including genetic drift, reassortment, recombination, and long distance virus transport. Genetic drift is proposed to be caused by an error prone viral RNA dependent RNA polymerase (RdRp) which has helped the adaptation of CCHFV to many different vertebrate and tick hosts by introducing random sequence changes to the genome, thereby providing the fundamental diversity needed for the evolution of CCHFV (Bente, Forrester et al. 2013). This has contributed to the emergence of 6-7 discrete CCHFV lineages or clades, which have several labels assigned to them by various investigators, including the designation by Atkinson et al who refer to them as Africa 1, Africa 2, Africa 3, Asia 1, Asia 2, Europe 1 and Europe 2 (Atkinson, Chamberlain et al. 2012; Atkinson, Latham et al. 2012). These clades are thought to have arisen due to the fact that CCHFV and its tick vectors are endemic in several different geographic regions. Tick species have evolved separately in these regions as they have different climates and vegetation. It is thought the co-evolution of CCHFV with region specific ticks has resulted in the generation of many different local virus isolates, which gave rise to the different clades (Bente, Forrester et al. 2013).

There is evidence that genetic reassortment occurs between different CCHFV strains; the genome of CCHFV is tripartite and simultaneous infection of the same host with multiple strains of CCHFV can result in the generation of viruses with novel combinations of S, M and L segments (Hewson, Gmyl et al. 2004; Burt, Paweska et al. 2009). Reassortment is thought to occur within the tick host as ticks remain infected with CCHFV throughout their life spans, whereas replication of CCHFV within mammals is transient and mammals are only viraemic for 1-2 weeks. It is therefore more likely for a tick to become simultaneously infected with two different strains of CCHFV to allow reassortment to occur (Davies, Jones et al. 1987; Jones, Davies et al. 1987). M segment reassortment is detected most frequently and it is thought that viruses which contain an RdRp and N protein that have co-evolved together replicate more efficiently than those viruses whose RdRp and N protein have not co-evolved with each other (Chamberlain, Cook et al. 2005; Goedhals, Bester et al. 2014).

Genetic reassortment has the potential to create CCHFV strains with novel phenotypes, such as an altered ability to be transmitted between vertebrate and tick hosts, or an increased pathogenicity in humans; a phenomenon that has been detected in other bunyaviruses. A recently detected member of the *Orthobunyavirus* genus has been associated with outbreaks of haemorrhagic fever in Somalia and Kenya and was named Ngari virus (Bowen, Trappier et al. 2001; Gerrard, Li et al. 2004; Briese, Bird et al. 2006). Genetic analysis revealed Ngari virus was composed of the M segment of Batai virus, and the S and L segments of Bunyamwera virus. Iquitos virus (specifically the IQT9924 strain) was associated with cases of febrile illness in Peru and was shown to be a novel reassortment of the S and L segments of Oropouche virus, genotype II, and the M segment of a still unidentified Simbu-serogroup virus (Aguilar, Barrett et al. 2011).

Long distance virus transport is thought to contribute to the genetic diversity of CCHFV through the introduction of CCHFV strains to regions that were previously naïve to them. This increases the possibility that a vertebrate or tick host may become simultaneously infected with two different strains of CCHFV and enables the possible emergence of new virus strains through reassortment and recombination (Mild, Simon et al. 2010). Evidence of long distance virus transport can be seen through the analysis of CCHFV phylogenetic trees; these reveal that several clades contain virus isolates with similar sequences that originate from very distant geographic regions. Long distance movement of CCHFV strains is thought to occur when migratory birds transport CCHFV infected ticks to different geographic regions, and through international livestock trading between countries (Carroll, Bird et al. 2010; Mild, Simon et al. 2010).

There is also evidence that recombination has taken place between CCHFV strains; sequence alignment of genome segments reveals the interruption of the nucleotide sequence of a particular CCHFV strain, and its replacement with the sequence of a different strain, which suggests that strand exchange took place through recombination between different CCHFV isolates (Hewson, Chamberlain et al. 2004; Lukashev 2005; Bente, Forrester et al. 2013).

1.1.7 CCHF symptoms and pathogenesis in humans

Humans become infected with CCHFV through several different routes, including through the bite of an infected tick, by crushing an engorged infected tick or through contact with the infected body fluids of a viraemic animal or human (Whitehouse 2004; Bente, Forrester et al. 2013). Following an incubation period typically lasting 1-6 days (however there have been reports of incubation periods lasting > 13 days (Kaya, Engin et al. 2011)), the onset of symptoms of CCHF is rapid, and are typified by a sudden onset of high fever (39-41 °C), headache, myalgia and arthralgia, fatigue, and dizziness. The onset of symptoms may also include nausea, vomiting and diarrhoea. This is known as the prehaemorrhagic phase and typically lasts 3-7 days. These initial symptoms are sometimes described as ‘general influenza-like symptoms’ (Ergonul 2008) and are not specific to CCHF, therefore early misdiagnosis is common; it has been estimated that ~2/3 patients with CCHF are initially misdiagnosed (Tasdelen Fisgin, Doganci et al. 2010).

The haemorrhagic phase of infection follows in severe cases of CCHF, frequently beginning with a petechial rash which develops into ecchymosis. Uncontrolled bleeding from the gums and nose is common (gingival haemorrhage and epistaxis) and is sometimes followed by haemorrhagic involvement of the airways, gastrointestinal tract, brain and urinary tract. This haemorrhagic phase lasts 2-3 days, and it is during this stage that patients die, often due to shock, haemorrhage, severe liver necrosis and multi-organ failure (Burt, Swanepoel et al. 1997). For those who survive, the final stage of CCHF disease is convalescence, which typically begins 10-20 days after symptom onset and is lengthy, often involving clinical symptoms such as breathing impairment, poor vision, loss of hearing, loss of memory, polyneuritis and tachycardia. Fortunately these sequelae are scarcely ever permanent (Ergonul 2008; Bente, Forrester et al. 2013; Zivcec, Safronetz et al. 2013).

Hallmarks of CCHFV infection include leukopenia, thrombocytopenia, increased serum levels of proinflammatory cytokines and chemokines, and elevated serum levels of liver enzymes (transaminase alanine aminotransferase or ALT) (Cevik, Erbay et al. 2007). These, as well as high viral loads, increased clotting times and low antibody titres correlate with disease severity, and can be used to predict a fatal outcome in up to 90% cases (Ergonul, Tuncbilek et al. 2006; Papa, Bino et al. 2006; Duh, Saksida et al. 2007; Cevik, Erbay et al. 2008; Weber and Mirazimi 2008; Onguru, Dagdas et al. 2010; Saksida, Duh et al. 2010). A viral load of > 10^9 copies/mL is a strong factor ($p < 0.001$) for differentiating CCHF patients who die from those who survive (Cevik, Erbay et al. 2007). It has been suggested that the ability of CCHFV to replicate to such high copy numbers in the blood may be a contributing factor to the propensity of CCHFV to be transmitted nosocomially; other arboviruses such as Dengue virus, which are not commonly transmitted nosocomially do not replicate to such high levels (Chen and Wilson 2004). CCHF has been reported to be fatal in 5-70% cases. This large range of case fatality rates has been attributed to several different factors, including the mode of transmission, as well as improved diagnostics and raised awareness of CCHF; this increases detection of mild cases of CCHF, which may not present with haemorrhagic manifestations (Atkinson, Chamberlain et al. 2013).

Patients who die from CCHF generally have an absent or weak antibody response and high levels of pro-inflammatory cytokines (Bente, Forrester et al. 2013). Cytokines are important factors influencing the pathogenesis of viral haemorrhagic fevers and cytokine levels often correlate with disease severity. During the course of CCHFV infection patient serum levels of TNFα (tumor necrosis factor alpha), IL-6 (interleukin 6) and IL-10 are elevated, however significantly higher levels of TNFα and IL-6 are seen in patients with fatal outcome, than in non-fatal cases (Ergonul, Tuncbilek et al. 2006; Papa, Bino et al. 2006). Other cytokines thought to play a role in CCHF pathogenesis include IFN-γ, IL-1β and CCL-2; IL-1β is a major mediator of innate immune reactions and CCL-2 is a marker of severe disease in murine and primate models of Ebola virus (Dinarello 1996; Geisbert, Hensley et al. 2003). It was hypothesised that as CCHFV infects primarily macrophages and dendritic cells the high levels of IFN-1β and CCL-2 expression enhances virus infection by recruiting these cells to sites of inflammation at the beginning of the infection, enabling increased CCHFV replication and dissemination (Bente, Alimonti et al. 2010; Peyrefitte, Perret et al. 2010).

The pathogenesis resulting from CCHFV infection is still not completely understood and a lot of what is known is derived from human case studies and some *in vitro* experiments, as well as being extrapolated from what is known about other haemorrhagic fever viruses. Hepatocytes, Kupffer cells, endothelial cells, macrophages and dendritic cells are all targets for CCHFV infection (Burt, Swanepoel et al. 1997; Peyrefitte, Perret et al. 2010). During severe CCHFV infection erythrocytes and plasma leak from the vascular system into surrounding tissues (Ergonul 2006; Karlberg, Tan et al. 2011). In the case of CCHF this vasculature dysfunction is linked to endothelial cell destruction, which can lead to capillary fragility, and can cause deregulated stimulation of platelet aggregation and degranulation, which can impair the bloods ability to clot, and lead to haemorrhages through activation of the intrinsic coagulation cascade (Schnittler and Feldmann 2003; Karlberg, Tan et al. 2011). It is not known whether the destruction of endothelial cells seen during CCHF is a direct result of CCHFV infection of these cells, or an indirect consequence of virus induced host factors, such as high levels of pro-inflammatory cytokines. A correlation has been observed between disease severity, vascular leakage and the strength of the proinflammatory response during infection with other haemorrhagic fever viruses (Bray and Mahanty 2003; Schnittler and Feldmann 2003; Karlberg, Tan et al. 2011). In addition, varying degrees of hepatic damage have been observed histopathologically in human CCHF patients, ranging from mild to more severe necrosis (Joubert, King et al. 1985; Burt, Swanepoel et al. 1997).

1.1.8 The CCHFV mouse model

In 2010 two different transgenic mouse models, an interferon α/β (type 1) receptor knockout mouse (IFN-α/βR$^{-/-)}$, and a STAT1 (signal transducer and activator of transcription 1 protein) knockout mouse, were evaluated for their ability to recapitulate human CCHF disease (Bente, Alimonti et al. 2010; Bereczky, Lindegren et al. 2010). The pathogenesis reported from the STAT-1 and IFN-α/βR$^{-/-}$ mouse models shows a similar disease progression to human CCHFV infection, beginning with elevated body temperature, and the detection of leukopenia and thrombocytopenia on day 1 post infection. Increased serum levels of liver enzymes, such as liver transaminase alanine aminotransferase (ALT) are also detected in CCHFV infected STAT-1 and IFN-α/βR$^{-/-}$ mice, as well as a significant increase in serum levels of various proinflammatory cytokines and chemokines.

Similar to fatal human cases of CCHF, high levels of viraemia were detected in both mouse models by quantitative RT-PCR, reaching 7.32 x 10^9 GEQ/mL (genome equivalents per mL) at day 3 post infection in the STAT 1 knock out mouse model (Bente, Alimonti et al. 2010). In both mouse models immunohistochemical analysis revealed the presence of viral antigen in the liver and spleen with hepatocytes, Kupffer cells, occasionally the endothelial lining of sinusoids, and macrophages staining positive (Zivcec, Safronetz et al. 2013). The intensity and the frequency of the CCHFV antigen-positive cells in the liver increased until death, when severe liver damage could be detected.

Evidence of haemorrhages were not detected in either mouse model, in contrast to human disease, where haemorrhaging, especially intestinal haemorrhaging is common. This may be linked to the rapid disease progression in mice or differences in the murine and human coagulation cascades (Bente, Alimonti et al. 2010). However, both mouse models faithfully reproduce major hallmarks of human disease and currently represent the most appropriate animal models of CCHF, despite the limitations of the model due to the impaired IFN and immune response phenotypes (Bente, Alimonti et al. 2010; Bereczky, Lindegren et al. 2010).

1.1.9 The HAZV mouse model

The IFN-α/βR$^{-/-}$ mouse was also susceptible to HAZV infection, where a dose-dependent, lethal disease similar to CCHF was observed (Dowall, Findlay-Wilson et al. 2012). Similar to CCHFV infection, all mice inoculated with the highest virus dose ($4x10^4$ pfu) died, whereas wild type mice inoculated with this virus dose survived with no clinical signs of disease. Consistent with CCHFV infected mice, the target organs of HAZV were also the liver and spleen; the highest viral titre was detected in these organs day 3 post infection, and histopathological changes were observed in the liver and spleen that were similar to those resulting from CCHFV infection (Dowall, Findlay-Wilson et al. 2012). As HAZV infection of IFN-α/βR$^{-/-}$ mice follows a similar disease course to CCHFV, HAZV could be used as a model for CCHFV infection with the aim of rapidly assessing potential therapeutics and vaccines against CCHFV, without having to work under containment level 4 conditions (Dowall, Findlay-Wilson et al. 2012).

1.1.10 CCHF diagnosis, treatment and vaccine

CCHF is diagnosed through the detection of viral RNA in patient serum samples with specific RT-PCR and quantitative RT-PCR assays. Enzyme linked immunosorbant assays (ELISA) are also used to detect virus specific IgM or IgG antibodies (Bente, Forrester et al. 2013). The most reliable method of diagnosis is considered to be detection of viral RNA by RT-PCR however, as severely ill patients often fail to develop an antibody response and therefore will not test positive in an ELISA (Shepherd, Swanepoel et al. 1989; Burt, Leman et al. 1994). In addition, quantitative RT-PCR provides a measurement of viral load and therefore also enables a prediction of disease severity and likelihood of death (Saksida, Duh et al. 2010). A recently developed RT-PCR assay targets a highly conserved portion of the 5’ non-translated region of the viral S segment, that is required for the ‘panhandle’ formation and genome replication (Atkinson, Chamberlain et al. 2012). This RT-PCR assay is able to detect virus strains in all 7 clades and allays concerns that a sequence specific diagnostic assay may fail to detect all strains of CCHFV, based on the high genetic diversity of the virus (Duh, Saksida et al. 2006).

The majority of people infected with CCHFV are either asymptomatic or develop a nonspecific febrile illness that does not require hospitalisation; for those few who fall severely ill treatment is largely supportive. Ribavirin has been used in the treatment of CCHFV patients for decades and has been reported to be beneficial in most cases, as long as it is administered early on during the course of infection (Ergonul, Celikbas et al. 2004). However a recent review of human cases where ribavirin was used to treat CCHF patients concluded that the quality of the evidence was low, and a randomised clinical trial was unable to show that ribavirin treatment had any significant effect on CCHF disease progression (Koksal, Yilmaz et al. 2010; Soares-Weiser, Thomas et al. 2010). When administered to CCHFV infected IFNα/βR$^{-/-}$ mice, ribavirin prolonged the time until death, but did not prevent it (Oestereich, Rieger et al. 2014). The efficacy of ribavirin treatment therefore remains unclear.

CCHFV is considered to be interferon sensitive, and human IFN was administered to 6 patients during the Tygerberg outbreak in South Arica however no benefit was observed (Joubert, King et al. 1985). There is therefore an urgent need to develop novel therapeutics for CCHF. The IFNα/βR$^{-/-}$ mouse was recently used to test the efficacy of the drug T705 (favipiravir; currently in late stage clinical trials as an antiviral against influenza virus infection) during CCHFV infection; T705 protected 100 % mice against lethal challenge with CCHFV when administered 1 hr, or up to 2 days post exposure (Oestereich, Rieger et al. 2014).

There is no vaccine available for CCHFV that has been licensed by the European Medicines Agency, or the US Food and Drug Administration. A formalin inactivated mouse-brain derived vaccine has however, been in use in Bulgaria since the 1970’s/1980’s. The Bulgarian V42/81 vaccine strain is administered to soldiers, medical personnel and other high risk groups in endemic regions (Papa, Christova et al. 2004; Papa, Papadimitriou et al. 2011). This vaccine effectively protected inoculated individuals (CCHFV has never been reported in any vaccinated individuals) however, controlled studies on the protective efficacy of this vaccine have not been performed, and it is unlikely to gain widespread international approval due to its crude preparation.

Recent efforts to develop a more effective vaccine have focused on using a recombinant vaccine based on Modified Vaccinia virus Ankara (MVA), which is a poxviral vector and can tolerate the insertion of the cDNA corresponding to the CCHFV M segment (Buttigieg, Dowall et al. 2014). The CCHFV M segment encodes a large polyprotein precursor, which is proteolytically cleaved to generate multiple different protein products, including the mature glycoproteins G_N and G_C. The recombinant MVA vaccine (MVA-GP) was tested in the recently described IFNα/βR$^{-/-}$ mouse model, and was shown to elicit both a cellular and humoral response against CCHFV challenge. The MVA-GP vaccine candidate protected all of the IFNα/βR$^{-/-}$ mice against lethal challenge with CCHFV, and it is hoped that the MVA-GP vaccine could provide the basis for a novel CCHFV vaccine that can meet modern international regulatory standards (Buttigieg, Dowall et al. 2014).

1.2 THE VIRUS

1.2.1 Virion structure

The virions of CCHFV and HAZV are enveloped, spherical and approximately 80-100 nm in diameter (Martin, Lindsey-Regnery et al. 1985). The glycoproteins G_N and G_C protrude though the envelope creating ‘surface projections’ 5-10 nm in length, are responsible for virions binding to cellular receptors, and are the targets for neutralising antibodies (Bergeron, Vincent et al. 2007). The interior of the virions contain the viral ribonucleoprotein particles (RNPs); as previously stated, the genome of CCHFV and HAZV is encoded on 3 separate segments of negative sense RNA, termed the Large (L), Medium (M), and Small (S) segments, named according to their relative lengths. The RNA segments are encapsidated along their entire length by the nucleocapsid protein (N) to form functional RNPs. Viral RNPs are thought to have at least one copy of the viral RNA dependent RNA polymerase (RdRp, or L protein) associated with them, to initiate transcription and genome replication in the host cell. The cytoplasmic tails of the glycoproteins extend into the interior of the virion, and are thought to interact directly with the RNPs; this interaction is thought to be essential for virus assembly.

It is assumed that each infectious particle must contain at least one copy of each of the RNPs. The RNPs within the virion are thought to be relatively unorganised (in contrast to influenza for example, where the 8 RNPs generally maintain the same array) which suggests that virion morphology is more likely to be dependent on lateral interactions between G_N and G_C heterodimers (Walter and Barr 2011). A schematic representation of CCHFV and HAZV virions is depicted in Figure 1.4 (A). An electron micrograph of HAZV (Figure 1.4 (B)) reveals the spherical morphology of these virions, and their relative sizes (Clerx, Casals et al. 1981; Ellis, Southee et al. 1981; Martin, Lindsey-Regnery et al. 1985).

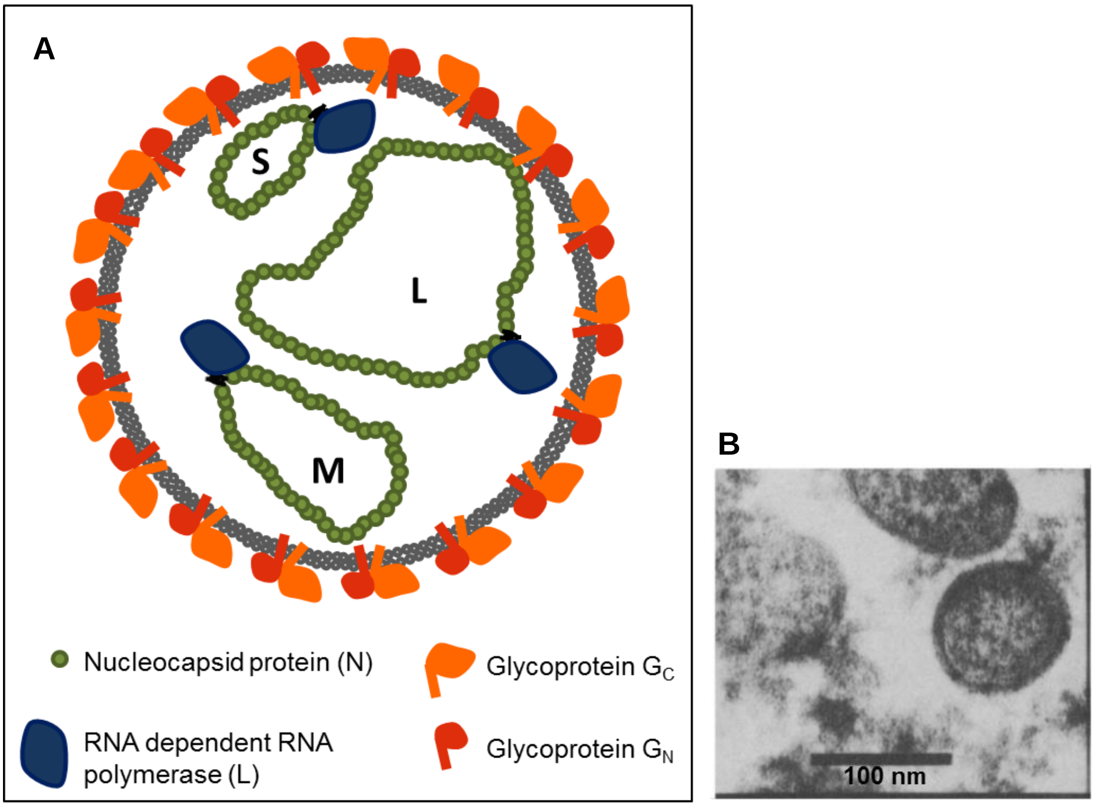


Figure 1.4 Schematic representation of CCHFV and HAZV virions and electron micrograph of HAZV.

(A) Schematic representation of CCHFV and HAZV virions depicting the glycoprotein spikes of G_N *and* G_C *in the virus envelope, surrounding the 3 viral RNPs, with attached L proteins. (B) Electron micrograph of HAZV virions, taken by (Clerx, Casals et al. 1981).*

1.2.2 CCHFV and HAZV genome structure

One of the interesting features of segmented negative sense RNA viruses (sNSVs) is the fact that their genomes are never found as naked RNA, but instead are always encapsidated by the viral nucleocapsid protein (N) (Green, Cox et al. 2014). The three genomic RNA segments possessed by CCHFV and HAZV are organised in a similar fashion across all 5 genera of the *Bunyaviridae* family; each segment termini contains 3’ and 5’ non-translated regions (NTRs) that surround an open reading frame (ORF) (Figure 1.5) (Walter and Barr 2011). In members of the same genus the length of each RNA segment is similar, and the nucleotide sequence of the extreme 3’ and 5’ termini of each segment is highly conserved and invariant (Table 1). All members of the *Nairovirus* genus have the following nucleotides at the extreme 3’ and 5’ termini of each of the 3 genome segments: 3' –AGAGUUUCU and 5'-UCUCAAAGA (Walter and Barr 2011).

There is extensive nucleotide complementarity between the 3' and 5' NTRs of each segment, which interact with each other through canonical Watson-Crick base pairing, to form an open circular panhandle structure. Bunyavirus RNA segments exist as open circular molecules both within virus particles and infected cells (Pettersson and von Bonsdorff 1975; Obijeski, Bishop et al. 1976; Raju and Kolakofsky 1989). Adjacent to the complimentary, conserved 11 nucleotides at the extreme 3' and 5' termini, there are additional complementary nucleotide motifs, which are shorter, and conserved only on a segment-specific basis. The interaction of the segment NTRs is essential in order to initialise RNA synthesis, as the promoters for RNA synthesis are thought to contain elements from both the 5' and 3' NTRs (Kohl, Dunn et al. 2004).

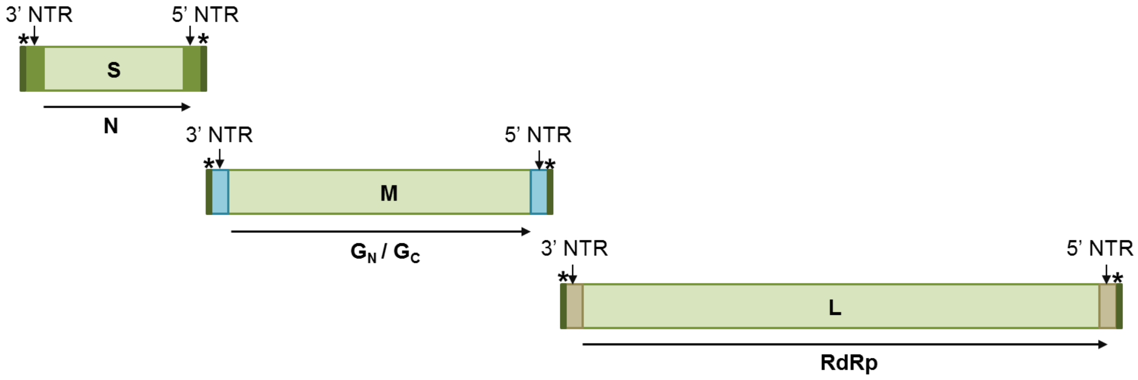


Figure 1.5 Schematic representation of Nairovirus genomic segments.
The genomic segments of CCHFV and HAZV (S, M and L) encode the N protein, viral glycoproteins G_N *and* G_C *and the RdRp, respectively. Each segment contains a single ORF, represented by horizontal black arrows, flanked by 3' and 5' NTRs. The NTRs possess both highly conserved 3' and 5' extreme termini nucleotide sequences (*), as well as non-conserved nucleotide sequences within the NTR regions (vertical arrows).*

In contrast to the extreme termini of the 3' and 5' NTRs, which show high sequence conservation, the internal nucleotides of the NTRs do not. There is much variation in the nucleotides sequences of these regions both among viruses in the same genera, and between genomic segments of the same virus (Walter and Barr 2011). Reverse genetic analysis has shown that the NTRs contain *cis* acting elements that are involved in regulating RNA synthesis and segment packaging for both BUNV and

RVFV (Barr, Elliott et al. 2003; Gauliard, Billecocq et al. 2006). In BUNV it was shown that 22 nucleotides at the genomic 3’ end and 113 nucleotides at the 5’ terminus within the unconserved regions of the segment NTRs were essential for infectious virus replication (Lowen and Elliott 2005). This implies that although many of these nucleotides are dispensable for virus replication, some sequences within segment specific NTRs are important for virus replication and affect virus fitness. During BUNV replication, the segment specific 3’ sequences were shown to mainly affect the balance of replication and transcription, whereas segment specific sequences at the 5’ regions predominantly affected overall virus growth (Lowen and Elliott 2005; Walter and Barr 2011).

Table 1. Features of Bunyavirus genome segments used for virus classification

Genus (prototype)	**Consensus terminal sequences**	**Segment length (nts)**		
		S	**M**	**L**
Nairovirus (Dugbe virus)	3’-AGAGUUUCU… 5’-UCUCAAAGA…	1712	4888	12255
Orthobunyavirus (Bunyamwera virus)	3’-UCAUACCAUG… 5’-AGUAGUGUGC…	961	4458	6875
Phlebovirus (Rift Valley fever virus)	3’-UGUGUUUC… 5’-ACACAAAG…	1690	3885	6404
Hantavirus (Hantaan virus)	3’-AUCAUCAUCUG… 5’-UAGUAGUAUGC…	1696	3616	6533
Tospovirus (Tomato spotted wilt virus)	3’-UCUCGUUA… 5’-AGAGCAAU…	2916	4821	8897

Rearrangement of the NTRs and ORFs of different BUNV genomic segments results in virus that cannot be rescued into infectious virus particles using the mini-genome replicon system; the only tolerated combination is when the L segment ORF is flanked by the M segment NTRs (MLM) (Lowen, Boyd et al. 2005). This suggests there is some kind of interplay between the NTRs and the ORF; in the MLM recombination, rescued virus had 100-fold decreased titres, and a greatly increased particle to plaque forming unit ratio, indicating there may be a reduced incorporation of MLM segments into infectious virus particles (Lowen, Boyd et al. 2005). The sequences at the NTR/ORF boundaries may therefore play a role in segment packaging, similar to Influenza packaging (Fujii, Goto et al. 2003; Fujii, Fujii et al. 2005; Walter and Barr

2011). Although not much is known about the specific NTR sequence requirements of CCHFV and HAZV, they are likely to play a similar role to those of BUNV during the virus replication cycle as organisation of the genomic segments is similar across the *Bunyaviridae* family.

1.3 THE REPLICATION CYCLE OF CCHFV

The replication cycle of CCHFV and HAZV takes place in the cytoplasm of infected cells. An overview of the CCHFV replication cycle is presented diagrammatically in Figure 1.6, and is adapted from (Bente, Forrester et al. 2013). The following sections describe the replication cycle of CCHFV; the replication cycle of HAZV has not been well characterised but is thought to be fairly analogous.

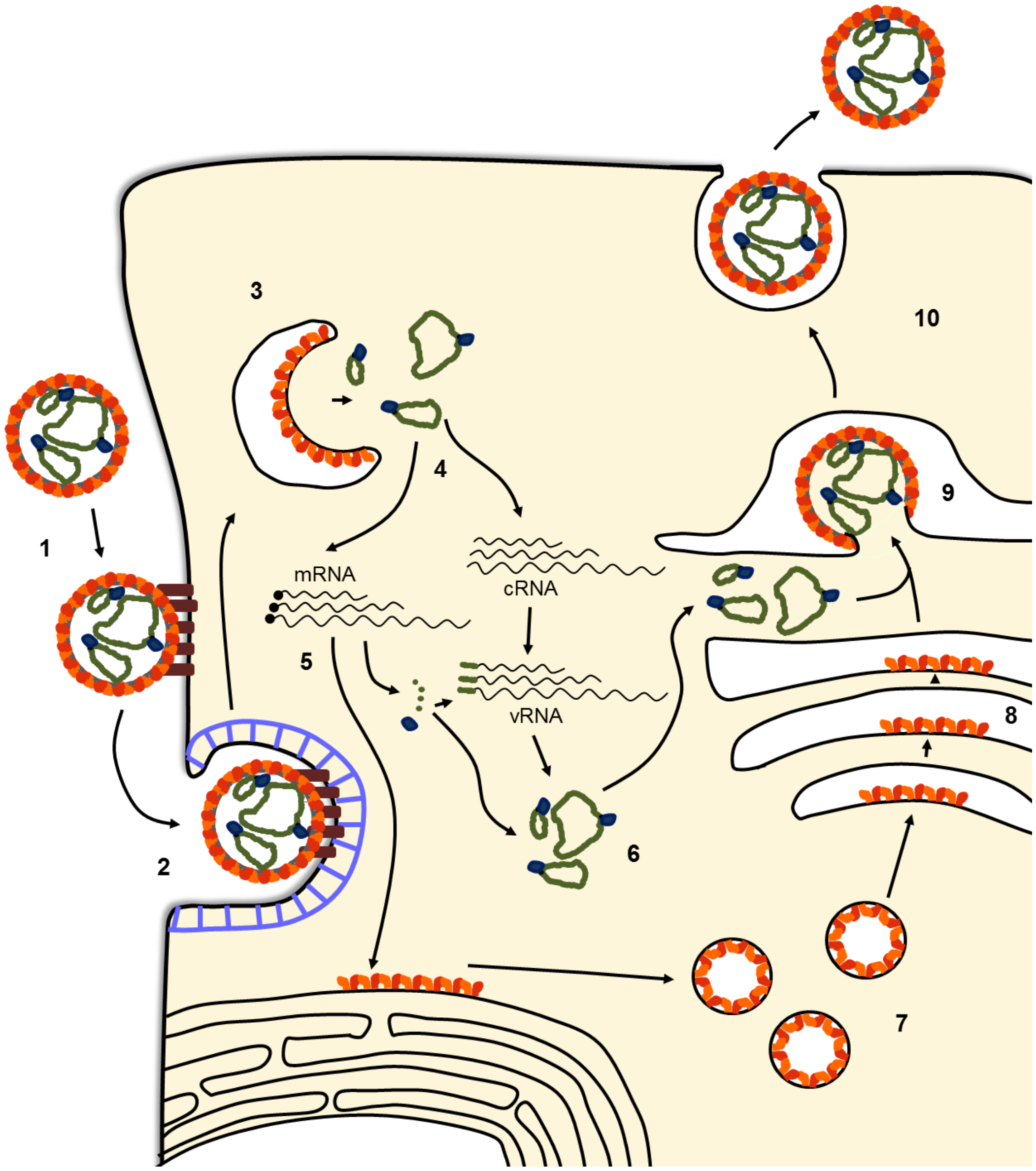


Figure 1.6. Overview of the replication cycle of CCHFV

Following binding to cellular receptors (1), CCHFV entry is via clathrin mediated endocytosis (2). CCHFV envelope fuses with early endosomal membranes to release viral RNPs (3), which traffic to perinuclear regions where viral factories are thought to form. Here, the genomic segments (vRNA) are transcribed to generate mRNA (4), which is translated to synthesise viral proteins (5). Genomic segments are also replicated to form complementary copies of the genomic segments (cRNA), which act as a template for subsequent vRNA production (4). vRNAs are encapsidated by newly synthesised N proteins to form RNPs (6). Once translated, viral glycoproteins G_N *and* G_C *traffic through the ER to the Golgi membrane (7), where they undergo extensive post-translational processing (8). Assembly occurs when newly synthesised viral RNPs interact with* G_N *and* G_C*, causing virions to bud into the Golgi cisternae (9). Virions are released from the cell after trafficking through the secretory pathway (10).*

1.3.1 Virus entry and uncoating

CCHFV entry is mediated predominantly by G_C, which binds to cell surface nucleolin, which has been shown to be essential for virus entry (Xiao, Feng et al. 2011). Internalisation of CCHFV occurs through clathrin-dependent endocytosis; siRNA knock down of clathrin, or one of its essential adaptor proteins, AP2, results in decreased CCHFV replication (Simon, Johansson et al. 2009). CCHFV entry is also dependent on low pH and cholesterol; the use of bafilomycin A1 or NH_4Cl to inhibit endosomal acidification, or methyl-β-cyclodextrin to reduce levels of cholesterol in cells, results in a loss of CCHFV replication (Simon, Johansson et al. 2009; Garrison, Radoshitzky et al. 2013).

Following cell entry, CCHFV virions fuse with early endosomal vesicle membranes to release viral RNPs into the cytoplasm. It is thought that acidification of early endocytic vesicles, which occurs as they bud from the plasma membrane and mature into late endocytic vesicles, changes the conformation of CCHFV G_N and G_C, allowing them to fuse with the endosomal membrane (Garrison, Radoshitzky et al. 2013). Efficient CCHFV replication relies on a functional Rab5 protein, rather than Rab7, indicating that CCHFV fusion occurs in the early endosomes, not the late endosomes, and the early endosomes rather than the late endosomes are important for CCHFV trafficking (Garrison, Radoshitzky et al. 2013).

CCHFV internalisation is also dependent on intact, dynamic microtubules. It was shown that in the first hour of CCHFV infection intact microtubules were required for efficient early RNA synthesis; it is thought they are required not only for virus internalisation, but also for subsequent transport of CCHFV virions or virion components to sites of virus replication (Simon, Johansson et al. 2009). Upon infection the N protein is trafficked to perinuclear regions in infected cells in an actin dependent manner; these perinuclear regions are thought to be the sites of CCHFV replication and virus factory formation (Andersson, Simon et al. 2004).

The mechanism of HAZV binding and internalisation is not known, however it is likely to be fairly similar to CCHFV, given the similarities in both the virion morphology and cell types targeted by these viruses.

1.3.2 RNA transcription and replication

Each encapsidated RNA segment within the virion must act as a template for two different RNA synthesis activities; transcription to generate mRNA, and replication to synthesise complementary copies of the full length genomic segments (cRNA), that subsequently act as templates for synthesis of viral genomic segments (vRNA)

(Figure 1.7). These RNA products have a fundamentally different structure; the mRNA is extended at the 5' end relative to the genome by the addition of a capped oligoribonucleotide stolen from host cell mRNAs, and the 3' end is truncated relative to the genome template (Walter and Barr 2011).

Although there is little specific information concerning the mechanism of RNA transcription and replication of CCHFV and HAZV, work with Dugbe virus, another Nairovirus, reveals that the mechanism of mRNA transcription initiation is similar in Nairoviruses to other members of the *Bunyaviridae* family (Jin and Elliott 1993). Therefore although the following section summarises what we know about transcription and replication of other members of the *Bunyaviridae* family, we can infer that the mechanisms employed by CCHFV and HAZV are similar.

Upon entry to an infected cell and transport to perinuclear regions, virus RNA transcription and replication takes place in the cytoplasm. In the currently accepted model, the first RNA synthesis event is transcription ('primary transcription'), carried out by the template associated RdRp on the input RNA strand to generate 5' capped mRNA (Walter and Barr 2011). The viral polymerase (L-protein) can only initiate RNA synthesis once per segment, at the segment termini; it is thought this might be due to the fact that sequences from both the 3' and 5' segment termini contribute to the formation of a functional promoter (Kohl, Dunn et al. 2004). The L protein therefore cannot re-initiate synthesis if RNA synthesis is prematurely terminated, and each bunyavirus genomic RNA strand must encode a maximum of one transcription unit (Walter and Barr 2011). This is in contrast to non-segmented negative sense RNA viruses (nsNSVs, e.g respiratory syncytial virus, RSV), where following transcription of each gene within the genomic RNA strand, the RNA polymerase terminates transcription at the gene end, then re-initiates transcription again at the next gene start site.

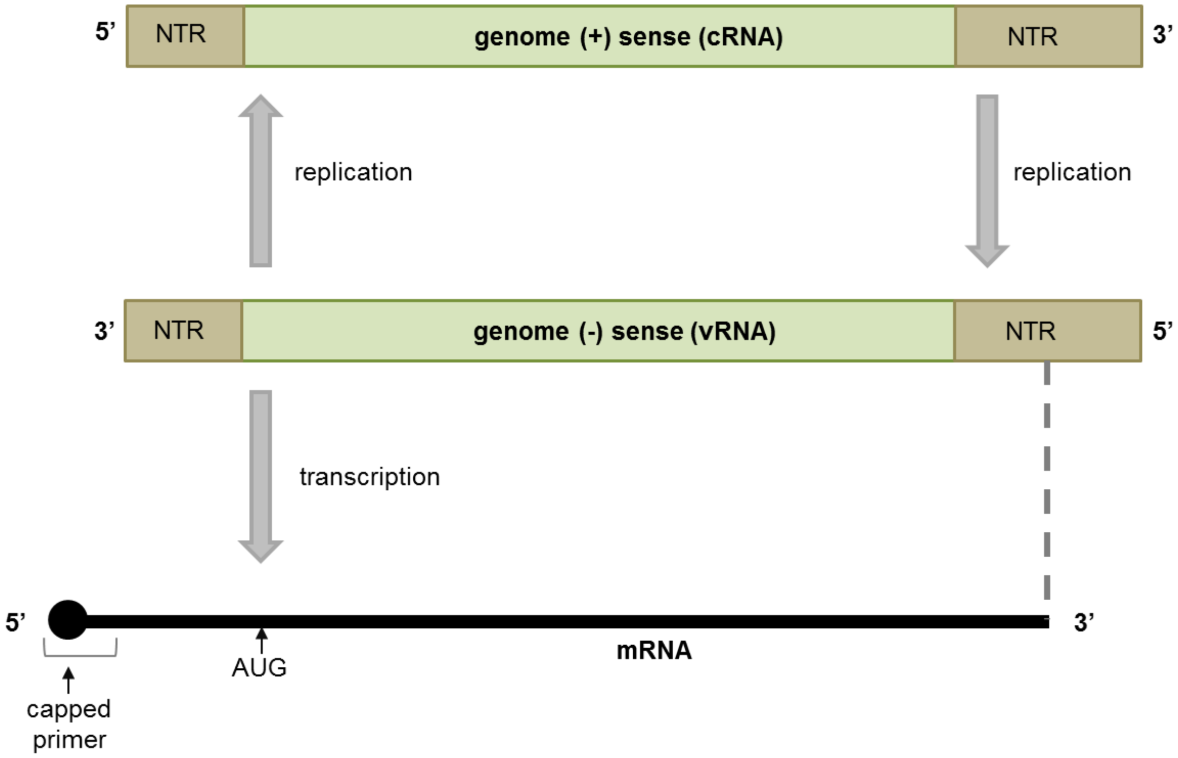


Figure 1.7. Negative sense gene expression.

CCHFV and HAZV vRNA must act as a template for two distinct RNA synthesis activities; transcription and replication.

Transcription initiation relies on primers that are generated from capped host cell mRNAs, and is distinct from replication in this regard. The evidence of cap-snatching for Dugbe virus (in addition to other members of the *Bunyaviridae* family) lies in the discovery of heterogeneous non-viral RNA sequences at the 5' ends of the Dugbe virus S segment mRNA (Jin and Elliott 1993). These are an average of 10 nucleotides in length, and are derived from host cell mRNA when the L protein cleaves cellular mRNA to steal the cap required for translation of viral mRNAs (Reguera, Weber et al. 2010). The mechanism by which capped mRNAs are selected is unknown, however it has been noted for Dugbe virus, as well as other bunyaviruses, that the 3' end of the primer resembles the 5' end of the viral RNA (Jin and Elliott 1993; van Knippenberg, Lamine et al. 2005). It could be that the L protein selects a specific subset of cellular mRNAs to generate primers, specifically those that allow at least partial base pairing with the 3' end of the vRNA template. It is thought the 'prime and realign' mechanism of transcription initiation then follows, which involves annealing the capped primer to internal regions of the 3' NTR within a tri-nucleotide repeat sequence (3'-AUCAUCAUC-5'). Following extension by the L protein by a few nucleotides the short primer slips backwards and re-aligns with the 3’ proximal triplet repeat on the template RNA (Garcin, Lezzi et al. 1995; Walter and Barr 2011). After a few cycles of prime and realign, a longer complementary primer is generated. This could form a more stable transcription complex that is more suitable for elongation, however the reason behind employing the prime and realign mechanism is still unclear.

The 3’ termini of bunyavirus mRNAs are consistently truncated with respect to the genomic RNA template, indicating there is some sort of transcription termination signal in the template RNA. In the case of BUNV S segment this transcription termination signal was shown to be around 33 nucleotides in length, and was located in the 5’ NTR (Barr, Rodgers et al. 2006). A related sequence was identified in the L segments 5’ NTR that also mapped closely to the 3’ end of the L mRNA; interestingly the M segment 5’ NTR did not possess a similar signal, raising the possibility that mRNA synthesis is terminated via a different mechanism. Bunyaviral mRNA is not polyadenylated at the 3’ termini; this could potentially affect protein synthesis, as the poly(A) tail is important for stability of mRNA transcripts, as well as recruitment of the ribosomes for efficient translation. There is some evidence that in the case of BUNV at least, nucleotide sequences at the 3’ terminus of the mRNA transcripts functionally replace the poly(A) tail through the formation of secondary structural elements, such as a strong stem loop structure (Blakqori, van Knippenberg et al. 2009; Walter and Barr 2011). The sequence and the structure of the BUNV 3’ mRNA terminus is strikingly similar to that of the 3’ end of cellular histone mRNA, which is also poly(A) deficient, possibly indicating that BUNV may have hijacked a pre-existing cellular mechanism for poly(A) independent mRNA translation (Marzluff 1992). Little is known about the RNA synthesis activities of RNPs of CCHFV and HAZV; however the overall mechanism could be analogous.

Viral mRNA is not recognised and bound by the N protein. This is in contrast to the cRNA, which is a full length copy of the viral genome, bound to multiple copies of N along its length, and in some members of the bunyavirus genera (Phleboviruses and Tospoviruses) can also act as a template to generate mRNA that encode non-structural (NS) proteins. In some genera (for example the Phlebovirus RVFV) cRNA segments are also packaged into virions to enable expression of NSs proteins immediately upon entry to the host cell (Ikegami, Won et al. 2005). Viral mRNA on the other hand, is not packaged into virions.

The promoters for transcription and replication of BUNV segments have been mapped to distinct sets of nucleotide sequences at both the 3’ and 5’ termini of the segments. These nucleotide sequences are arranged so that the genomic strand NTR contains signals for both transcription and replication of the segment, whereas the cRNA strand only contains signals for replication of the segment (Barr, Elliott et al.

2003; Flick, Katz et al. 2004; Kohl, Dunn et al. 2004; Gauliard, Billecocq et al. 2006). After primary transcription has occurred, the viral polymerase switches to segment RNA replication. The signal that initiates the switch to segment replication from mRNA synthesis is poorly defined, and may be as a result of a change in the RdRp function either through modification of the polymerase or its interaction with host-cell factors (Walter and Barr 2011). It is assumed, however, that cytoplasmic N protein concentration is also important for the switch from transcription to replication, as sufficient N protein must be translated to encapsidate the newly synthesised cRNA and vRNA segments (Reguera, Cusack et al. 2014). In contrast to mRNA synthesis, synthesis of cRNA is initiated using 5' nucleoside triphosphates. The relative abundance of each genomic segment varies within infected cells; nucleotide sequences within the segment specific NTRs contain elements that influence the relative abundance of each segment that is replicated. During BUNV infection, the relative abundance of genome replication of the 3 segments is M>L>S (Barr, Elliott et al. 2003).

The subcellular localisation of CCHFV cRNA and vRNA during infection has been determined (Andersson, Henriksson et al. 2012). Simultaneous staining of the N protein revealed that vRNA subcellular distribution is fairly diffuse in the cytoplasm, only occasionally co-localising in areas where N assembled or aggregated. cRNA on the other hand was greatly enriched in the CCHFV-N containing areas of the cytoplasm, and co-localised extensively with the perinuclear N aggregates, rather than having a more diffuse cytoplasmic distribution. The reason for the varying cytoplasmic distribution of cRNA and vRNA is unknown, but may be related to their different roles within the CCHFV replication cycle (Andersson, Henriksson et al. 2012).

1.3.3 CCHFV and HAZV assembly and egress

CCHFV and HAZV are thought to assemble and bud from perinuclear viral factories that are built around the Golgi complex, in a similar manner to other members of the *Bunyaviridae* family (Kuismanen, Hedman et al. 1982; Matsuoka, Chen et al. 1991; Novoa, Calderita et al. 2005). This may be a result of the accumulation of the structural glycoprotein heterodimer G_N and G_C in the membrane of the Golgi complex, where they are trafficked for post-translational processing following synthesis in the ER. Virion formation is thought to occur when viral RNPs traffic to the Golgi membrane and interact with the cytoplasmic tails of G_N and G_C, which protrude into the cytoplasm.

It is not known how CCHFV RNPs are transported to the G_N and G_C heterodimers in the Golgi complex, however several components of the cytoskeletal system, including actin filaments and microtubules, are thought to be involved, especially in N protein transportation (Andersson, Simon et al. 2004; Ramanathan, Chung et al. 2007; Simon, Johansson et al. 2009). Depolymerisation of microtubules affects CCHFV assembly, and results in a redistribution of both N and G_N, as well as significantly reducing progeny virus release. Stabilisation of microtubules also results in the accumulation of progeny virions in the cytoplasm, indicating a role for both intact and dynamic microtubules in CCHFV egress (Simon, Johansson et al. 2009).

Once viral RNPs have trafficked to the G_N and G_C heterodimer in the Golgi membrane they are thought to directly interact with the cytoplasmic tails of G_N and G_C, as CCHFV does not encode any non-structural or matrix proteins that can mediate this interaction. Recent structural studies of the cytoplasmic tail of CCHFV G_N revealed it contained a dual CCHC-type zinc finger domain (Estrada and De Guzman 2011). The cytoplasmic tail of G_N was able to bind RNA, specifically RNA resembling the hairpin-like M segment panhandle, raising the intriguing possibility that the cytoplasmic tail of G_N may interact with the RNA component of the RNPs, rather than the N protein (Estrada and De Guzman 2011). The concentration of structural components of CCHFV in the vicinity of the Golgi membrane within a ‘budding compartment’ is thought to drive the formation and budding of new CCHFV virions into the cisternae of the Golgi complex. Virions are subsequently released along the secretory pathway.

1.3.4 Proteins encoded by the CCHFV M segment

The CCHFV M segment possesses a single ORF that encodes a polyprotein precursor of 1,689 amino acids. This polyprotein is co-translationally inserted into the membrane of the rough endoplasmic reticulum (RER), where it undergoes a complex series of cleavage events and post-translational modifications, both in the RER, and as it migrates through the Golgi network. This results in the synthesis of multiple protein products, which are summarised in Figure 1.8.

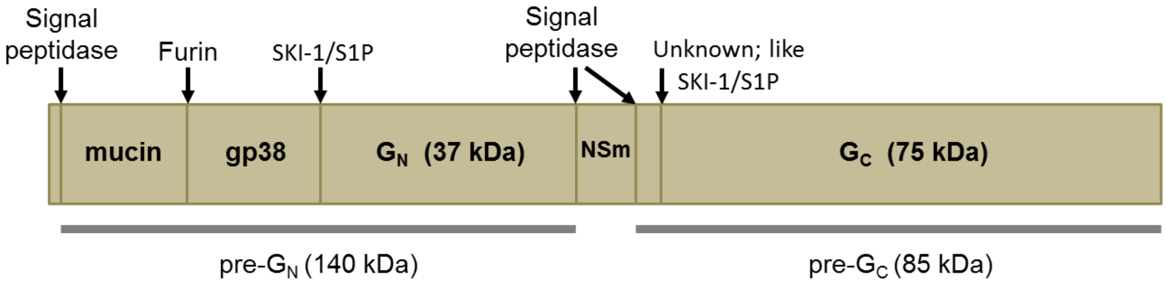


Figure 1.8 Diagrammatic representation of CCHFV glycoprotein polypeptide

The CCHFV M ORF is initially translated as a large polyprotein that undergoes complex processing by cellular proteases (listed above the polyprotein) to generate the structural glycoproteins G_N and G_C amongst other protein products (labelled on the polyprotein).

The polyprotein precursor contains two distinct regions; the C terminal 1,441 residues, which are highly conserved among CCHFV strains (91.6 % sequence identity) and the N terminal 243-248 amino acids residues, which are highly variable (up to 56.4 % sequence variation, labelled ‘mucin’ in Figure 1.8). This N terminal domain has mucin-like features, including a high serine, threonine and proline content (up to 47.3 %) and is predicted to be heavily O-glycosylated (Sanchez, Vincent et al. 2002). The possession of a mucin-like domain is a unique feature among the *Bunyaviridae* family, and notably, this mucin like domain resembles that present within the Ebola virus glycoprotein, which is predicted to play an important role in immune evasion (Simmons, Wool-Lewis et al. 2002). Interestingly, analysis of the efficacy of a novel CCHFV vaccine which uses the M segment ORF revealed that a large part of the cellular immune response detected in protected mice mapped to epitope regions in the non-structural regions of the M segment polyprotein, including the mucin-like domain (Buttigieg, Dowall et al. 2014).

The main structural proteins derived from the M segment are the membrane bound glycoproteins G_N and G_C, which are found inserted in the viral envelope. G_N is 37 kDa, and G_C is 75 kDa, and both are initially produced by cleavage of the polyprotein to generate a 140 kDa pre-G_N precursor, and an 85 kDa pre-G_C precursor. Cleavage of pre-G_N at a conserved tetrapeptide $RRLL_{519}$ motif by the protease SKI-1/S1P releases G_N from pre-G_N; this cleavage is essential for infectious virus release; cells deficient in SKI-1/S1P are susceptible to infection with CCHFV, but infectious particles are not released by these cells (Bergeron, Vincent et al. 2007, Vincent, Sanchez et al. 2003). G_C is cleaved from pre-G_C at the conserved $RKPL_{1040}$ motif, which closely resembles the motif recognised by SKI-1/S1P, however the protease involved in this cleavage remains unknown (Bergeron, Vincent et al. 2007, Vincent, Sanchez et al. 2003). During infection both G_N and G_C localise to the Golgi complex, however only G_N contains the Golgi localisation signal; heterodimerisation of G_N and G_C is required for the Golgi localisation of G_C (Bertolotti-Ciarlet, Smith et al. 2005). Both G_N and G_C undergo N-linked glycosylation, G_N is N-linked glycosylated at one site on its ectodomain (557N) and G_C is glycosylated at two sites on its ectodomain (1054N and 1563N) (Erickson, Deyde et al. 2007). Antibodies targeting G_C were able to neutralise CCHFV prior to infection, but antibodies directed against G_N could not, indicating that G_C is involved in recognising and binding to cellular receptors (Bertolotti-Ciarlet, Smith et al. 2005).

Other viral proteins generated by M ORF polyprotein cleavage include GP38 and a non-structural protein designated NSm (Sanchez, Vincent et al. 2006, Altamura, Bertolotti-Ciarlet et al. 2007). The function of GP38, NSm and the mucin-like domain during the virus replication cycle are not known; these proteins could not be detected in partially purified virions, indicating they are not structural proteins required for virion formation.

1.3.5 Proteins encoded by the HAZV M segment

Fewer studies have investigated the proteins encoded by the HAZV M segment ORF. Similar to CCHFV, HAZV also encodes at least two glycoproteins of ~75 kDa and ~30 kDa, which are embedded in the virus envelope (Clerx, Casals et al. 1981; Foulke, Rosato et al. 1981). There is some evidence to suggest that HAZV may also encode a third envelope glycoprotein, of ~45 kDa (Foulke, Rosato et al. 1981); however Clerx et al do not report the detection of this third envelope glycoprotein, therefore further work needs to be carried out to resolve this point.

1.3.6 Proteins encoded by the L segment

The L segment of CCHFV and HAZV encodes the viral RNA dependent RNA polymerase (RdRp), or L protein, whose function is to both transcribe and replicate genomic RNA segments. The nairovirus L protein is considerably larger than the L protein of other bunyaviruses; the full length protein has an approximate molecular weight of 448 kDa. The full length L protein possesses the core catalytic domains common to all viral RdRps; pre-A, A, B, C and D, in addition to a motif specific for all sNSVs, termed region E, that together make up the polymerase module (Muller, Poch et al. 1994; Walter and Barr 2011). The L proteins of nairoviruses also encode additional sequences, predominantly at their N termini, that perform as yet uncharacterised functions. In the case of CCHFV, some of these additional regions show some homology with known transcription factor, gyrase, helicase, and topoisomerase domains (Honig, Osborne et al. 2004; Kinsella, Martin et al. 2004).

Additionally, some of these N-terminal nairovirus specific sequences show strong similarity to cysteine proteases within the ovarian tumor (OTU) super-family and the L protein of CCHFV and Dugbe virus possesses OTU-like protease activity (Frias-Staheli, Giannakopoulos et al. 2007). The viral OTU (vOTU) protease domain of CCHFV L directly deconjugates ubiquitin and ISG15 from their target proteins *in vitro* and in cell culture, resulting in an overall decrease in the amount of ubiquitinated or ISG15 conjugated proteins in cells (Frias-Staheli, Giannakopoulos et al. 2007). Furthermore, this deconjugation of ubiquitin and ISG15 results in antagonism of the innate antiviral response, both through directly abrogating the antiviral effects of ISG15, which are mediated through its ability to conjugate to target proteins, and though interference with any immune pathways that are regulated by ubiquitination. The NF-$_{K}$B dependent antiviral signalling pathway is regulated by ubiquitination for example, and upon L protein expression, nuclear translocation of endogenous p65 is inhibited, thus expression of antiviral genes under the control of the NF-$_{K}$B promoter does not occur (Frias-Staheli, Giannakopoulos et al. 2007).

Interestingly cellular OTU-domain containing proteases only deconjugate ubiquitin, not ISG15. The molecular basis for the ability of vOTU proteases to cleave both substrates was recently revealed though structural analysis of CCHFV OTU-domain with either ubiquitin or ISG15 covalently bound in the active site (Akutsu, Ye et al. 2011; James, Frias-Staheli et al. 2011) (Figure 1.9).

It has been proposed that the OTU-like protease domain may cleave the CCHFV L protein autoproteolytically to release both the polymerase as well as various other proteins that are required to carry out transcription and replication of the viral genome. Such a mechanism has been described for the equine arteritis virus (EAV) nsp2 protein, which is an OTU-like protease with dual purpose; nsp2 cleaves the viral polyprotein to release individual proteins, and also deconjugates ubiquitin and ISG15 from cellular proteins to interfere with innate antiviral signalling pathways, and suppress the innate immune response (Frias-Staheli, Giannakopoulos et al. 2007; van Kasteren, Bailey-Elkin et al. 2013). However, the CCHFV vOTU domain is not necessary for efficient transcription and replication of the CCHFV genome, as deletion of this domain did not affect the activity of the L-protein in a CCHFV minireplicon assay (Bergeron, Albarino et al. 2010). This implies that the CCHFV vOTU is not required for polyprotein processing of the L protein to generate an efficient RNA polymerase.

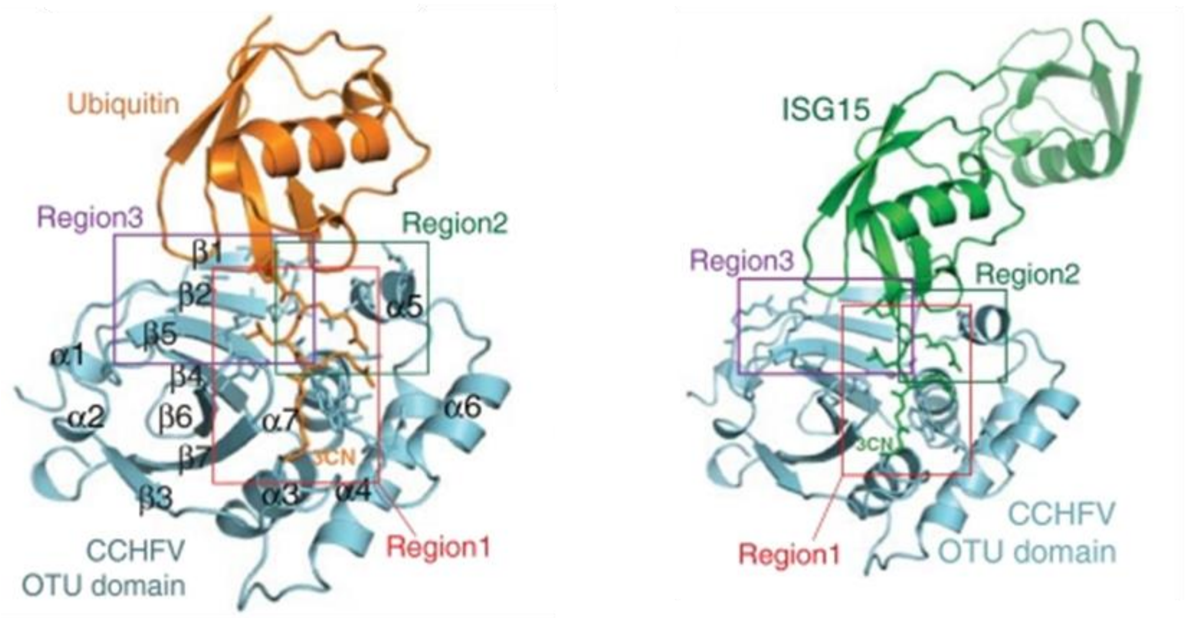


Figure 1.9 Crystal structure of vOTU domain of CCHFV L protein in complex with ubiquitin or ISG15

Taken from (James, Frias-Staheli et al. 2011). Crystal structure of CCHFV vOTU (cyan) bound to ubiquitin (orange) or ISG15 (green). In both cases ribbon diagrams of the complex highlight the 3 main regions of interaction between ubiquitin or ISG15, and the vOTU domain of CCHFV L protein. The C-terminal ubiquitin like domain of ISG15 binds the substrate binding groove of the protease in an analogous manner to ubiquitin.

It was recently shown that the N terminal domain of the LACV L protein encodes a cap snatching endonuclease with very close structural and biochemical features to that of the N terminal domain of the influenza virus polymerase PA subunit (Reguera, Weber et al. 2010). Mutational analysis revealed that this endonuclease activity was essential for efficient transcription of LACV mRNA, and the endonuclease activity was inhibited by DPBA, a drug that targets the influenza endonuclease. Sequence alignment with other members of the *Bunyaviridae* family indicated the endonuclease active site motif (typically composed of 5 potentially catalytic amino acid residues; H34, D52, D79, D92, and K94) is conserved among members of the *Orthobunyavirus, Phlebovirus, Hantavirus* and *Tospovirus* genus (Reguera, Weber et al. 2010). Although the nairoviruses were not included in this original alignment, they were found to possess a putative endonuclease motif of the Phlebo/Hanta/Orthomyxo type in the residue range 630–710, in the case of CCHFV. However, endonuclease activity of the CCHFV L protein needs to be confirmed with structural and functional data, and it should be noted that although the cap snatching endonuclease activity of CCHFV could possibly be attributed to this domain in the L protein, there is no evidence as yet that the L protein also contains the cap binding site.

1.4 PROTEINS ENCODED BY THE S SEGMENT: CCHFV N AND HAZV N

The S segment of CCHFV and HAZV encodes the nucleocapsid protein (N), often the most highly abundant structural protein present in infectious virions. CCHFV N and HAZV N are significantly larger than the N proteins of most other *Bunyaviridae* family members, apart from members of the *Hantavirus* genus (Table 1). The best characterised role of N is to bind to viral RNA to form functional RNPs, which are essential for both transcription and replication of the viral genome, by analogy with other bunyaviruses. One of the unique features of NSVs is the fact that RNPs, not naked genomic RNA segments, are used as a template for viral RNA synthesis (Green, Cox et al. 2014). Multiple copies of CCHFV N are proposed to interact with viral RNA along the length of the genome or anti-genome segments, possibly via homotypic interactions between adjacent monomers in a ‘head-tail’ manner (Wang, Dutta et al. 2012).

By encapsidating the viral genome or anti genome, but not the viral mRNA or cellular RNA, it is thought N protects the viral genome from degradation, helps avoid the formation of dsRNA structures between viral RNA strands of opposite polarity, and compacts the viral RNA into RNPs (Reguera, Cusack et al. 2014). It is thought genome encapsidation and RNP assembly occurs concurrently with genome synthesis; this property is thought to be one way of ensuring that only genomic or anti-genomic RNA is packaged into virions during assembly, rather than viral or cellular mRNA (Reguera, Cusack et al. 2014). RNP assembly of N with nascent genome or anti-genome strands is thought to be initiated at the RNA 5’ termini, as there are no conserved sequences, other than those in the NTRs, that could serve as initiation sites for N-assembly (Osborne and Elliott 2000; Reguera, Cusack et al. 2014).

Extensive analysis of N proteins and RNPs from multiple members of both the segmented NSVs (sNSVs) and non-segmented NSVs (nsNSVs), reveals the N proteins and RNPs of these two groups of viruses have fundamentally different properties (Reguera, Cusack et al. 2014). The RNPs of sNSVs are highly flexible and are able to adopt a number of different conformations (Young and Howard 1983). The flexibility of the sNSV RNPs has been proposed to be required for multiple stages of the virus replication cycle; the inherent flexibility of the N:RNA chains may be essential in enabling the complementary terminal sequences of each genomic RNA strand to find each other and anneal during genome replication, for example (Reguera, Cusack et al. 2014).

In general, the nucleocapsid proteins of sNSVs are structurally diverse; they do not share similar folds in their domains. However, they do all possess at least 2 separate domains; an RNA binding domain, and another domain that mediates oligomerisation between neighbouring monomers. These domains are linked by highly flexible hinge regions that allows N proteins to adopt a number of different conformations, enabling the formation of RNP assemblies with variable geometries (Reguera, Cusack et al. 2014).

The bunyavirus N protein is multifunctional, and in addition to encapsidating viral RNA, N has other critical functions within the virus replication cycle; N interacts with both the viral L protein during RNA synthesis (Eifan and Elliott 2009), and the virus glycoproteins (G_N and G_C) during virus assembly and budding (Overby, Pettersson et al. 2007; Overby, Popov et al. 2007; Ribeiro, Borst et al. 2009). It is thought that interaction of the L protein with N induces local conformational changes in N that allow it to release the RNA genome temporarily, so it can be used as a template for RNA synthesis by the L protein. As the L protein then continues downstream RNA synthesis, it is thought N re-encapsidates the viral RNA to avoid the formation of dsRNA intermediates (Green, Cox et al. 2014).

The S segment of members of other bunyavirus genera encode additional non-structural proteins, either using a negative sense coding strategy and a second ORF (orthobunyaviruses) or the NSs protein ORFs are transcribed as mRNAs from positive sense templates (phleboviruses and tospoviruses) (Elliott 2014). Bunyavirus NSs proteins are generally thought to be interferon antagonists that enhance virus pathogenicity (Weber, Bridgen et al. 2002; Hollidge, Weiss et al. 2011; Eifan, Schnettler et al. 2013). There is no evidence of the CCHFV or HAZV S segment expressing any non-structural protein product, although the CCHFV S segment does encode an additional open reading frame in the positive sense orientation that is conserved in almost all CCHFV strains (Hewson, Chamberlain et al. 2004).

1.4.1 The crystal structure of CCHFV N

The crystal structure of several different strains of the CCHFV N protein have been determined. The YL04057 strain was solved to a resolution of 2.3 Å by (Guo, Wang et al. 2012), the Baghdad-12 strain to a resolution of 2.1 Å by (Carter, Surtees et al. 2012), and the IbAr10200 strain to 3.1 Å by (Wang, Dutta et al. 2012). In all cases the structure of the full length protein was determined (residues 1-482) and consistent with other sNSVs, CCHFV N possesses two domains; a globular domain and an arm domain; the arm domain is flexible and free to adopt a number of different positions. In the Baghdad-12 strain the globular domain comprises 23 alpha helices from both the N and the C terminus of N (α1 to α11 of the N-terminus, and α18 to α29 of the C terminus), where the C terminal alpha helices are located in the core of the globular domain, and are surrounded by alpha helices from the N-terminus (helices are numbered according to Figure 1.10 (B)) (Carter, Surtees et al. 2012). The arm domain comprises two long alpha helices (α15 and α17) that extend away from the globular domain and have a loop at the apex that is supported by a small 3-helix bundle (α12 to α14) (Figure 1.10) (Carter, Surtees et al. 2012). Figure 1.10 (B) also shows the primary sequence alignment between strains YL04057 and Baghdad-12, and the conserved caspases-3 cleavage motif on the apex of the arm domain is highlighted in (B).

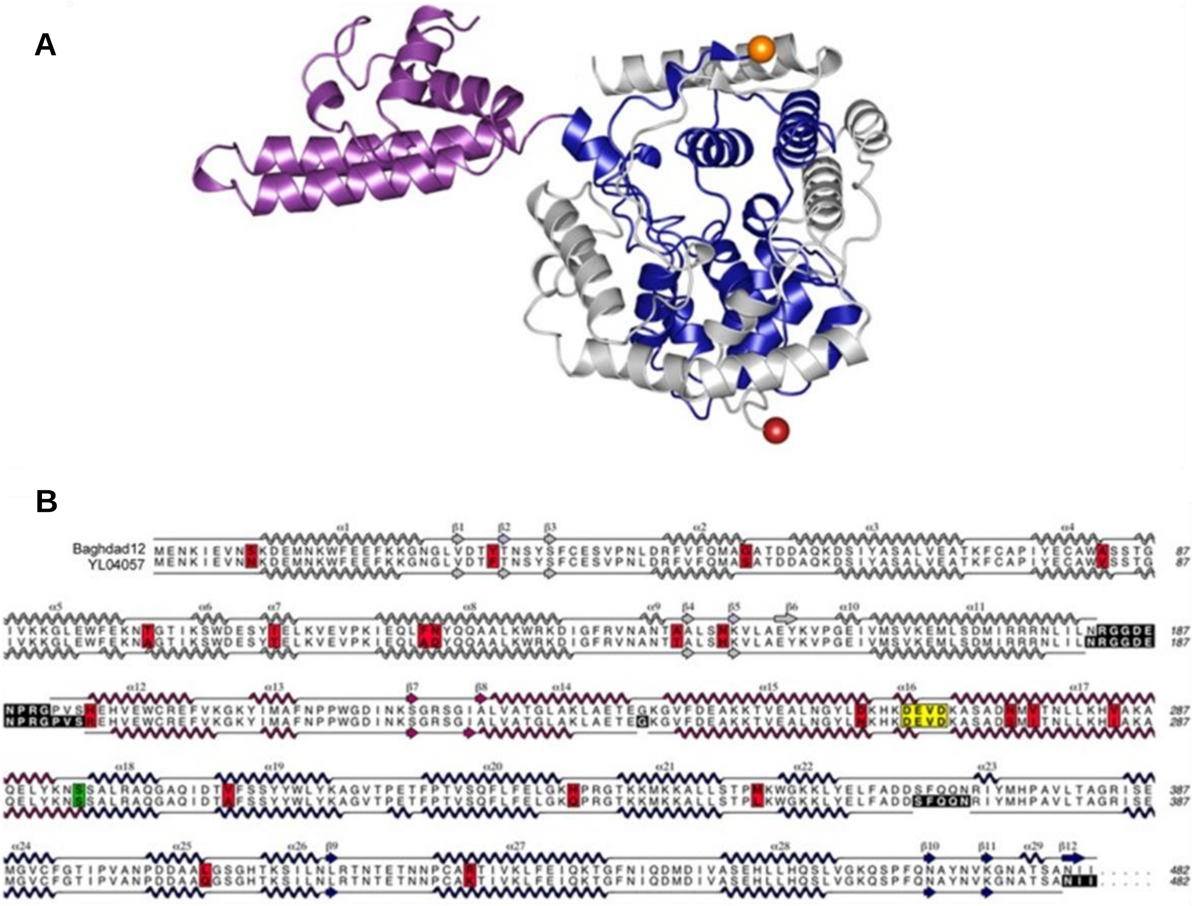


Figure 1.10 Crystal structure of CCHFV N Baghdad-12 strain, and primary sequence alignment of YL04057 and Baghdad-12 strains

Taken from (Carter, Surtees et al. 2012). (A) Crystal structure of CCHFV N Baghdad-12 strain with the N terminus shown in grey, the C-terminus blue, and the extended arm domain in purple. The N and C-termini are highlighted by red and yellow balls, respectively. (B) primary sequence alignment of YL04057 and Baghdad-12 strains with alpha helices numbered accordingly.

In both the YL04057 and Baghdad-12 strain only 474 of the 482 residues were able to be built into the model; residues 183-191 form a disordered loop that is missing in electron density. In the IbAr10200 structure this flexible loop is visible in only one of the three monomers in the asymmetric unit, in the other two it remains disordered, with no detectable electron density (Wang, Dutta et al. 2012). This disordered loop links the arm domain to the globular domain, and it has been hypothesised that the loop could become ordered upon RNA binding, and therefore the arm domain may ‘gate’ RNA binding, in a similar manner to the RVFV and LASV gate (Ferron, Li et al. 2011; Hastie, Liu et al. 2011; Carter, Surtees et al. 2012). Analysis of the electrostatic surface potential of all three structures revealed two positively charged surfaces that were proposed to be involved in RNA binding (Figure 1.11). One positively charged patch is on the surface of N, adjacent to the arm domain and is termed the platform. The other is a crevice that channels between the base of the globular domain and the beginning of the arm domain, in a region similar to where RNA binds in the LASV N terminal domain (Figure 1.13) (Wang, Dutta et al. 2012). However, these structures were all devoid of RNA and attempts to co-crystallise purified CCHFV N with short sequences of RNA were unsuccessful, thus the RNA binding domain of CCHFV N remains unidentified.

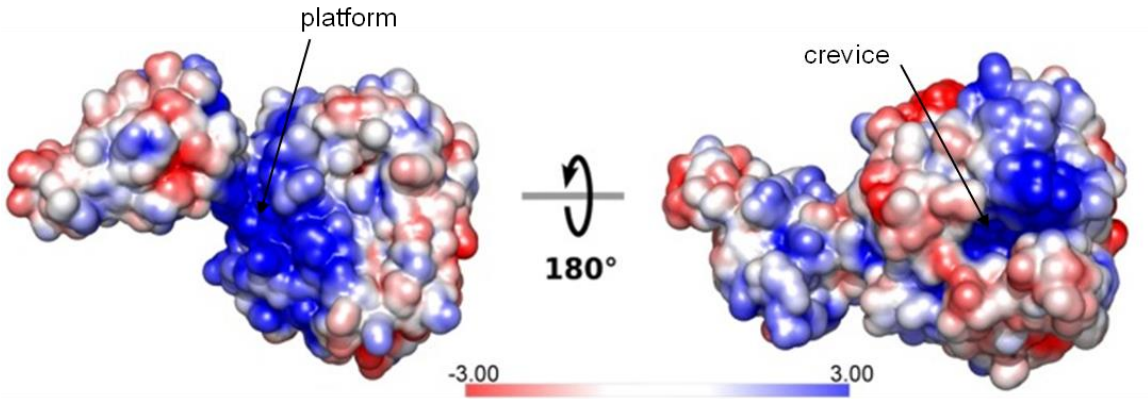


Figure 1.11 Surface electrostatic potential of CCHFV N

Taken from (Carter, Surtees et al. 2012). Surface electrostatic potential of CCHFV N reveals areas of positive charge (blue) and negative charge (red), and two possible RNA binding regions termed the 'platform' and the 'crevice'.

The globular domains of all three CCHFV N structures are very similar, for example comparison of the YL04057 and Baghdad-12 strains reveals the globular domains superpose very closely on one another; the Cα root mean square deviation (RMSD) is 0.975 Å for 357 residues with 92.16 % sequence identity (Carter, Surtees et al. 2012). The arm domains of these strains also contain the same fold and superpose closely on one another (Cα RMSD of 0.642 Å, 99 residues, 77.78 % identity), however the position of the arm domain is different in the two structures. The arm is rotated by about 180 °, and the apex of the loop is shifted by a distance of 39.59 Å in the YL04057 strain compared to the Baghdad-12 strain (Carter, Surtees et al. 2012). This change in the arm domain position could be a result of amino acid differences between the strains. Alternatively, given that the arm domain and globular domain are linked by a single α helix, which is unlikely to be rigid, and a disordered loop not visible in electron density, the arm domain could be fairly flexible, and might be free to adopt several different conformations (Carter, Surtees et al. 2012). It is not known whether the change in the position of the arm domain is involved in N multimerisation, or whether the RNA binding ability of the respective proteins is affected by the position of the arm domain, as this might alter the distribution of electrostatic surface potential (Carter, Surtees et al. 2012).

Structural alignment (superposition) of CCHFV N with RVFV N (a member of the *Phlebovirus* genus), revealed very little structural similarity between their globular domains, and no similarity at all was detected between their arm domains. Instead, the greatest structural similarity was observed between CCHFV N and the N terminal domain of Lassa virus (LASV) N, a member of the *Arenaviridae* family. The N terminal domain of LASV N, comprising residues 1-340 (the RNA binding domain of LASV N), when structurally aligned with the globular domain of CCHFV N, had a Cα RMSD of 3.38 Å for 237 residues with 8.01% sequence identity (Carter, Surtees et al. 2012). Interestingly this suggests that despite CCHFV and RVFV both being classified as members of the *Bunyaviridae* family, CCHFV is more closely related to the arenavirus LASV than RVFV. This suggestion is supported both by a phylogenetic analysis of the L and N sequences from sNSVs, which suggests that nairoviruses are more closely related to arenaviruses then other members of the *Bunyaviridae* family (Carter, Surtees et al. 2012), and also the fact that nairoviruses and arenaviruses share elements of their replication cycles that are unique to nairoviruses within the *Bunyaviridae* family, such as their reliance on cellular SKI-1/S1P protease to process their glycoprotein precursors (Lenz, ter Meulen et al. 2001; Vincent, Sanchez et al. 2003).

1.4.2 Potential multimerisation mechanism of CCHFV N

In all three studies, purified CCHFV N was mainly monomeric, although (Wang, Dutta et al. 2012) also reported the existence of higher order structures. These higher ordered oligomers (apparent MW over 300 kDa) had OD_{260}/OD_{280} ratios of 1.49, which is consistent with them containing nucleic acid (presumably RNA) (Wang, Dutta et al. 2012). In contrast the reported monomeric N (apparent MW approximately 53 kDa) was apparently devoid of bound nucleic acid, having an OD_{260}/OD_{280} ratio of 0.69. Wang et al did not crystallise the higher ordered structures with bound nucleic acid, however they did observe antiparallel double superhelical polymers of CCHFV N within crystals of monomeric N. Crystallisation of monomeric N in solution and the subsequent formation of higher ordered structures and oligomers in crystals is not uncommon – for example the monomeric form of RVFV N once crystallised forms hexameric rings, as well as other oligomeric structures (Ferron, Li et al. 2011). It is thought that while free in solution the presence of RNA may be required to stabilise the interactions between monomers, however the much greater concentration of N during crystallisation trials and within crystals may drive N proteins to form larger, more stable oligomers (Wang, Dutta et al. 2012).

CCHFV N formed long helical structures in crystals where N monomers were organised in a ‘head-to–tail’ manner, with the loop on the end of the arm domain directly interacting with the base of the globular domain of an adjacent monomer (Wang, Dutta et al. 2012). Although these polymeric structures could just be the result of crystallographic packing artefacts, they could also represent the physiological organisation of CCHFV RNPs. Wang et al also propose a model for the mechanism of RNA transcription from the RNP templates. They were able to show that incubation of CCHFV N with short, primer length (16-21 nucleotides) ssRNA induced a conformational change in N, causing the arm domain to rotate away from the globular domain such that N could no longer take part in multimerisation. They propose that this disruption causes the release of monomeric CCHFV N from higher ordered structures. Wang et al further propose that during transcription the presentation of capped, primer length ssRNA stolen from host cell mRNAs, to CCHFV RNPs would induce a similar conformational change in N, causing the release monomeric N and exposing vRNA to the RdRp for transcription (Wang, Dutta et al. 2012).

1.4.3 Cap-binding and endonuclease potential of CCHFV N

CCHFV N does not bind mRNA cap analogues (m7G, m7Gp or m7Gppp), and is therefore unlikely to be responsible for cellular mRNA cap snatching during the virus replication cycle (Guo, Wang et al. 2012). Guo et al did however show that CCHFV N bound single stranded RNA (ssRNA) with a very weak affinity, and that N possessed divalent-cation dependent, DNA specific, endonuclease activity (Guo, Wang et al. 2012). This is somewhat surprising; CCHFV N is so far the only sNSV N protein that has been attributed with DNA specific endonuclease activity, although the C-terminal domain of LASV N (not the domain that CCHFV N shares a strong structural similarity with) has been shown to possess a dsRNA specific 3’-5’ exonuclease activity (Qi, Lan et al. 2010; Hastie, Kimberlin et al. 2011). The possible function of the endonuclease activity of CCHFV N within the replication cycle of CCHFV remains to be determined. The weak binding affinity of NSV nucleocapsid proteins for ssRNA *in vitro* has previously been reported despite the function of N being RNA binding during the virus replication cycle (Ferron, Li et al. 2011).

1.4.4 Cleavage of CCHFV N by active caspase-3

The crystal structure of CCHFV N also reveals the position of the DEVD motif which is cleaved by activated caspase-3 during apoptosis in CCHFV infected cells (Karlberg, Tan et al. 2011). The DEVD motif is located on the loop at the apex of the arm domain, in a position that is theoretically easily accessible by proteases. Recombinant CCHFV N is cleaved *in vitro*, resulting in the generation of 2 fragments of N of approximately 30 kDa and 24 kDa. *In vitro* cleavage of N is not very efficient; incomplete cleavage of N is observed after overnight incubation of N with recombinant caspase-3 in equimolar ratios, and this reflects cleavage in infected cells, where only a small proportion of N is cleaved (Karlberg, Tan et al. 2011; Carter, Surtees et al. 2012). It is possible that oligomerisation of CCHFV N completely obscures the DEVD motif, as it is thought the loop of the arm domain directly interacts with the base of the adjacent globular domain, suggesting that only monomeric N would be an available substrate for activated caspase-3. This might explain the slow rate of caspase-3 cleavage; indeed incubation of CCHFV N with primer length ssRNA, which disrupts the multimeric structure and releases monomeric N, increases the efficiency of CCHFV N cleavage by caspase-3 (Wang, Dutta et al. 2012).

Following cleavage by activated caspase-3, size exclusion chromatography analysis indicates that the two fragments of CCHFV N remain bound to each other, and the monomer remains intact as a single unit (Carter, Surtees et al. 2012). This suggests that caspase cleavage of CCHFV N may not affect protein function; alternatively cleaved CCHFV N may have an altered tertiary or quaternary structure which may affect its function (Carter, Surtees et al. 2012). Analysis of N function in a CCHFV mini-genome replicon system, using mutated CCHFV N possessing a DEV(E/G) motif that cannot be cleaved by caspase-3, has conflicting results. (Wang, Dutta et al. 2012) report an increase in RdRp function, indicating cleavage by caspase-3 is detrimental to CCHFV replication, (Carter, Surtees et al. 2012) report no difference in the level of reporter gene expression. The exact function of caspase-3 cleavage of N during the replication cycle of CCHFV remains to be determined.

1.4.5 Cellular binding partners of CCHFV N

CCHFV N has been shown to interact with the interferon (IFN) inducible protein Mx1 and the cytoskeletal protein actin (Andersson, Bladh et al. 2004; Andersson, Simon et al. 2004). During infection N localises to perinuclear regions in infected cells; initially N is expressed in a small granular pattern in the cytoplasm, and is then targeted to perinuclear regions in the first 8 hrs after infection (Andersson, Simon et al. 2004). This perinuclear staining was not associated with the Golgi membrane, and was shown to be dependent on the presence of intact actin filaments; disruption of actin filaments with cytochalasin D resulted in a much more granular cytoplasmic distribution of N (Andersson, Simon et al. 2004). Actin also co-immunoprecipitated with N from CCHFV infected cells. An 80-97 % reduction in viral titre was observed when cells were treated with cytochalasin D (1 or 10 µg/mL) implying that actin filaments may be involved in or enhance virus assembly or release (Andersson, Simon et al. 2004).

Mx1 binds to CCHFV N during infection; Mx1 co-immunoprecipitates with N from infected cells, and co-localises with N in large perinuclear inclusions (Andersson, Bladh et al. 2004). Mx1 inhibits the replication of CCHFV; expression of Mx1 in Vero cells results in a significant reduction in CCHFV titre, and during IFN-mediated inhibition of CCHFV replication, Mx1 was shown to be an important effecter protein (Andersson, Lundkvist et al. 2006). Mx1 belongs to the dynamin family of large GTPases and its expression is induced exclusively by IFN α and IFN β. Mx1 expression also has an inhibitory effect on the replication of viruses from multiple different families; within the *Bunyaviridae* family Mx1 was shown to inhibit the replication of Dugbe virus, a member of the Nairobi sheep disease serogroup, as well as members of the *Hantavirus* and *Phlebovirus* genera (Frese, Kochs et al. 1996; Kanerva, Melen et al. 1996).

In the case of CCHFV it is thought the interaction of Mx1 with N might sequester N away from sites of active virus replication, preventing N from fully participating in the virus replication cycle (Andersson, Bladh et al. 2004). Alternatively Mx1 might interact with N in the context of the RNPs or capsid-like structures and prevent their functions. Mx1 did not interfere with the early stages of CCHFV entry, S segment transcription and N protein translation (up to 12 hrs post infection), but rather seemed to inhibit virus genome replication. RT-PCR analysis revealed a reduction in vRNA, and a similar loss in virus titre as new infectious particles could not form (Andersson, Bladh et al. 2004).

1.4.6 Comparison between CCHFV N and HAZV N

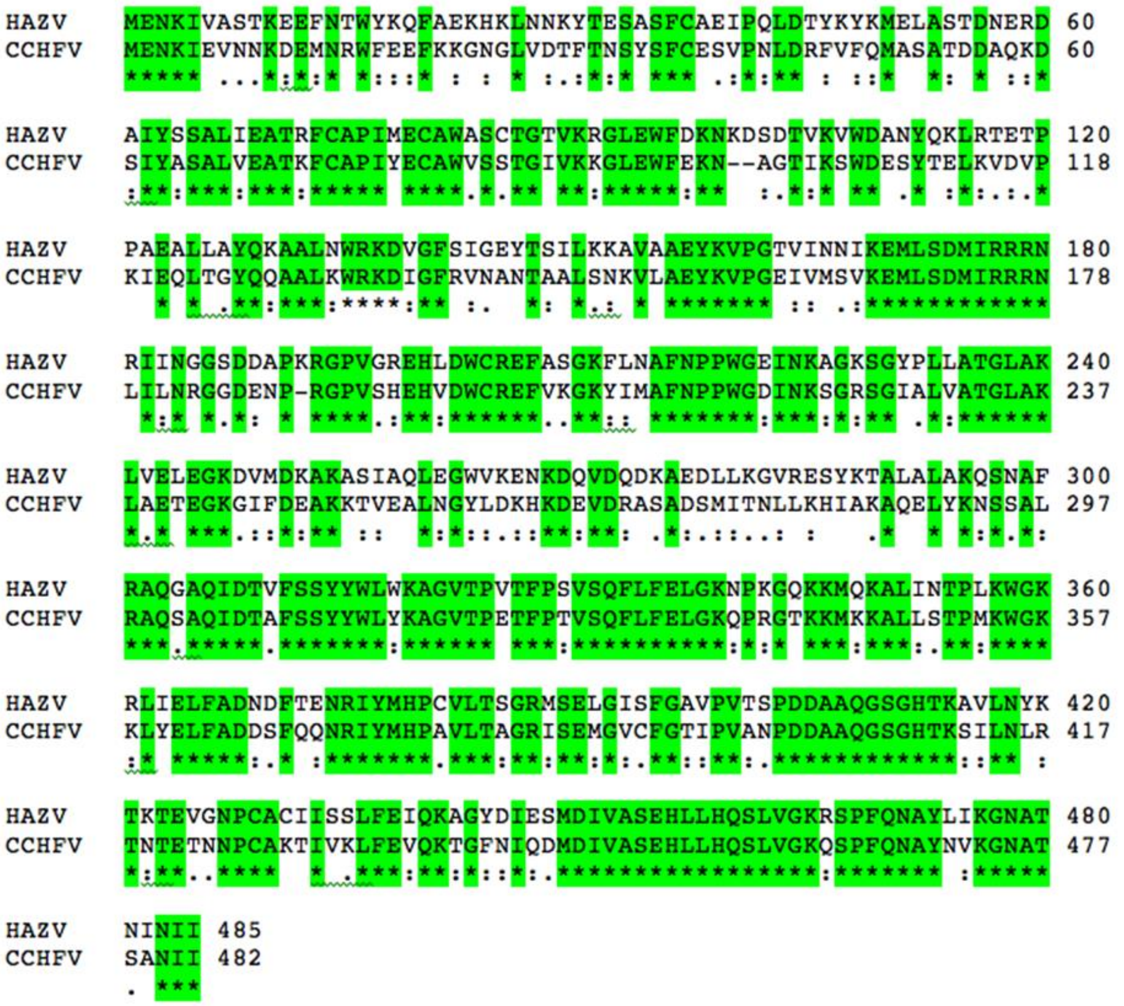
Alignment of the nucleotide sequences of CCHFV and HAZV S segments reveals slight differences in their overall length and the length of their 3' and 5' NTRs (Table 2). The nucleotide sequence homology between CCHFV N and HAZV N ORFs is 49.7 %.

Table 2. Comparison between CCHFV and HAZV S segments

Comparison between CCHFV and HAZV S segments reveals slight differences in segment and NTR length. Data was acquired from (Marriott and Nuttall 1992).

	Overall segment length (nts)	5' NTR length (nts)	3' NTR length (nts)	Size of N protein
CCHFV	1672	55	171	482 aa (53.966 kDa)
HAZV	1677	82	140	485 aa (54.186 kDa)

The amino acid sequences of CCHFV N and HAZV N proteins are more highly conserved, sharing an overall 60% amino acid sequence identity (Figure 1.12); HAZV N encodes 3 extra amino acids in comparison to CCHFV N. It is thought HAZV N and CCHFV N perform the same function in their virus replication cycles; although very few studies have investigated HAZV N, HAZV N has been found associated with RNA in purified viral RNPs, presumably encapsidating viral RNA in a similar manner to CCHFV N (Clerx, Casals et al. 1981; Foulke, Rosato et al. 1981). Interestingly in the study performed by Foulke et al, HAZV RNPs co-sedimented with small amounts of G_C, possibly implying the interaction between the RNPs and G_C is stronger than the interaction between the RNPs and G_N.



```
HAZV      MENKIVASTKEEFNTWYKQFAEKHKLNNKYTESASFCAEIPQLDTYKYKMELASTDNERD 60
CCHFV     MENKIEVNNKDEMNRWFEEFKKGNGLVDTFTNSYSFCESVPNLDRFVFQMASATDDAQKD 60
          ***** ...*:*:* *:::* : : * :.:*:* *** .:*:** : ::*  *: * ::*

HAZV      AIYSSALIEATRFCAPIMECAWASCTGTVKRGLEWFDKNKDSDTVKVWDANYQKLRTETP 120
CCHFV     SIYASALVEATKFCAPIYECAWVSSTGIVKKGLEWFEKN--AGTIKSWDESYTELKVDVP 118
          :**:***:***:***** ****.*.** **:*****:**  :.*:* ** .* :*:.:.*

HAZV      PAEALLAYQKAALNWRKDVGFSIGEYTSILKKAVAAEYKVPGTVINNIKEMLSDMIRRRN 180
CCHFV     KIEQLTGYQQAALKWRKDIGFRVNANTAALSNKVLAEYKVPGEIVMSVKEMLSDMIRRRN 178
            * * .**:***:****:** :.  *: *.: * ******* :: .:************

HAZV      RIINGGSDDAPKRGPVGREHLDWCREFASGKFLNAFNPPWGEINKAGKSGYPLLATGLAK 240
CCHFV     LILNRGGDENP-RGPVSHEHVDWCREFVKGKYIMAFNPPWGDINKSGRSGIALVATGLAK 237
           *:* *.*: * ****.:**:******..**:: *******:***:*:** .*:******

HAZV      LVELEGKDVMDKAKASIAQLEGWVKENKDQVDQDKAEDLLKGVRESYKTALALAKQSNAF 300
CCHFV     LAETEGKGIFDEAKKTVEALNGYLDKHKDEVDRASADSMITNLLKHIAKAQELYKNSSAL 297
          *.* ***.::*:** ::  *:*::.::**:**: .*:.::..: :   .*  * *:*.*:

HAZV      RAQGAQIDTVFSSYYWLWKAGVTPVTFPSVSQFLFELGKNPKGQKKMQKALINTPLKWGK 360
CCHFV     RAQSAQIDTAFSSYYWLYKAGVTPETFPTVSQFLFELGKQPRGTKKMKKALLSTPMKWGK 357
          ***.*****.*******:****** ***:**********:*:* ***:***:.**:****

HAZV      RLIELFADNDFTENRIYMHPCVLTSGRMSELGISFGAVPVTSPDDAAQGSGHTKAVLNYK 420
CCHFV     KLYELFADDSFQQNRIYMHPAVLTAGRISEMGVCFGTIPVANPDDAAQGSGHTKSILNLR 417
          :* *****:.* :*******.***:**:**:*:.**::**:.************::** :

HAZV      TKTEVGNPCACIISSLFEIQKAGYDIESMDIVASEHLLHQSLVGKRSPFQNAYLIKGNAT 480
CCHFV     TNTETNNPCAKTIVKLFEVQKTGFNIQDMDIVASEHLLHQSLVGKQSPFQNAYNVKGNAT 477
          *:**..****  * .***:**:*::*:.*****************:******* :*****

HAZV      NINII 485
CCHFV     SANII 482
          . ***
```

Figure 1.12 Amino acid sequence alignment of CCHFV N and HAZV N.

Conserved amino acid residues between CCHFV N and HAZV N are highlighted in green. These proteins share an overall 60% amino acid sequence identity; HAZV N encodes 3 extra amino acids. Alignment was generated with ClustalW.

1.5 THE N PROTEIN STRUCTURE OF RELATED VIRUSES

The crystal structures of the N proteins of many sNSVs have been determined, both with and without RNA bound. These include N proteins from the orthomyxoviruses; Influenza A and B (Ye, Krug et al. 2006) (Ng, Lam et al. 2012), and Infectious Salmon Anaemia virus (Zheng, Olson et al. 2013). Within the *Bunyaviridae* family, the structure of N proteins from members of the *Phlebovirus* genus RVFV (Raymond, Piper et al. 2010; Ferron, Li et al. 2011; Raymond, Piper et al. 2012) and Severe fever with thrombocytopenia syndrome virus, SFTSV (Jiao, Ouyang et al. 2013; Zhou, Sun et al. 2013)) have been solved. Within the *Orthobunyavirus* genus the structure of the N protein of BUNV (Ariza, Tanner et al. 2013), LACV (Reguera, Malet et al. 2013), Schmallenburg virus (SHMV) (Ariza, Tanner et al. 2013; Dong, Li et al. 2013), and Leanyer virus (Niu, Shaw et al. 2013)) have been solved. The structure of CCHFV N from the *Nairovirus* genus has also been determined. The crystal structure of the N protein of several arenaviruses, including Lassa fever virus (LASV) (Qi, Lan et al. 2010; Hastie, Liu et al. 2011) and Junin virus (Zhang, Li et al. 2013), have also been determined. The following section briefly summarises what is known about the LASV N protein structure, as the crystal structure of CCHFV N showed greatest homology to the N terminal domain of LASV N.

LASV N consists of 2 domains, an N and C terminal domain joined by a flexible linker region. The N terminal domain is compact and mostly α-helical, consisting of head and body regions that contain 4 and 8 helices respectively, whereas the C terminal domain possesses a single β-sheet surrounded by α helices (Hastie, Kimberlin et al. 2011; Hastie, Liu et al. 2011). The crystal structures of LASV N with RNA bound and without RNA bound have been solved, and reveal that LASV N possesses a novel gating mechanism for RNA binding. Six to seven ssRNA nucleotides bind per monomer of LASV N, in a crevice located in the N terminal domain. Although most of the RNA sequence bound to LASV N seemed to be non-specific, LASV N did show some preference for a purine base at position 3 in the RNA strand (Hastie, Liu et al. 2011).

RNA binding to LASV N is mostly mediated though interactions of the phosphate residues and backbone of the RNA with arginine and lysine residues present on α helices α12 and α14, and a hydrogen bond network between the phosphate backbone and several serine and threonine residues of N (Hastie, Liu et al. 2011). Mutational analysis of several of these residues revealed their absolute necessity to the function of a LASV mini-genome replicon system, presumably because they are key RNA binding residues. The gating mechanism of LASV RNA binding is mediated by two α helices, helix α5 and helix α6, as well as the flexible loop that links these two helices (Figure 1.13). In the RNA free structure helix α5 is 10 residues longer, and both the longer helix α5 and helix α6 lie across the RNA binding crevice, blocking RNA from binding. In contrast, in the RNA bound monomer, these 10 residues are flexible and disordered, and as a result helix α5 does not extend across the crevice and helix α6 is rotated away from the crevice. In addition the loop connecting helix α5 and helix α6 is moved away from the crevice in the presence of RNA and seems to form a ‘gate’; in the RNA free monomer (the closed conformation) this loop sterically clashes with the RNA, whereas in the RNA bound ‘open conformation' several residues in this loop interact with the RNA backbone and sugar residues (Hastie, Liu et al. 2011).

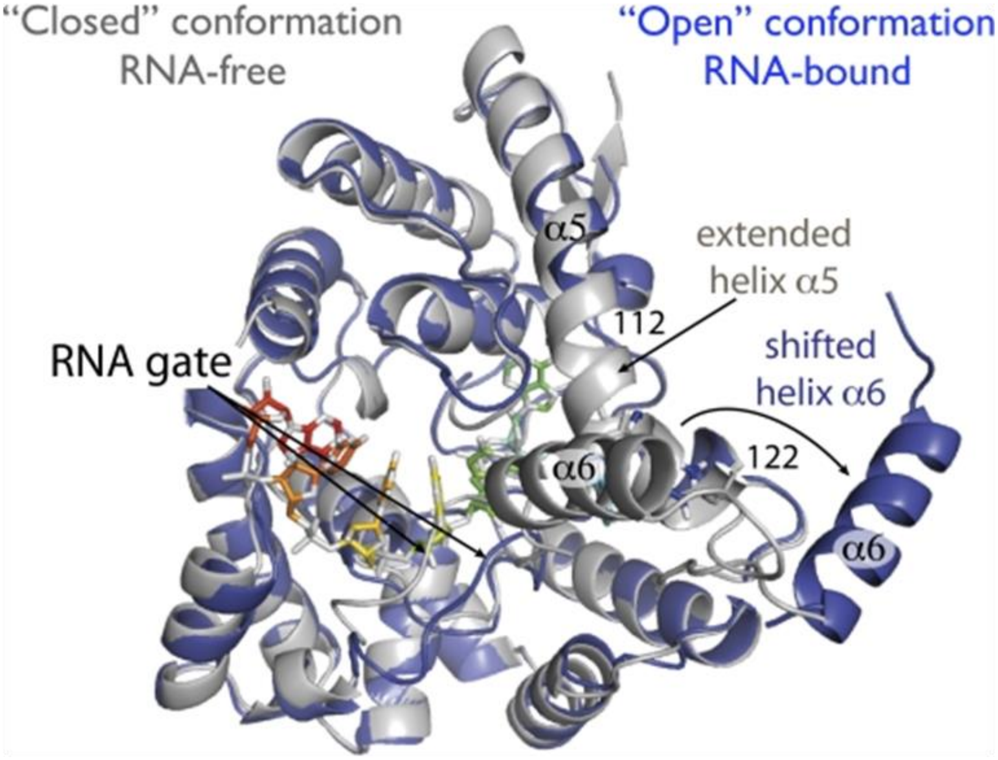


Figure 1.13 The LASV N 'gating' mechanism of RNA binding

Taken from (Hastie, Liu et al. 2011). Ribbon diagram depicting the novel LASV N gating mechanism of RNA binding. The N terminal domain is shown in both the open conformation (blue) with RNA bound, and the closed conformation (silver), which does not bind RNA.

It has been hypothesised that in the 'closed conformation' (in which N is trimeric) LASV N cannot bind RNA, and this closed conformation therefore performs one of the functions of the P protein (the polymerase co-factor present in other NSVs but not bunyaviruses) and prevents N polymerisation and non-specific encapsidation of cellular RNA. The factor that catalyses the shift from the closed conformation to the open conformation is currently unknown, however the L protein is thought to be involved. The fact that the RNA is bound in such a deep crevice and is inaccessible for base pairing during replication supports the hypothesis that during RNA synthesis the L protein induces a conformational change in N that causes N to transiently release the RNA for replication and transcription. N then re-encapsidates the RNA after the L protein has continued downstream RNA synthesis (Hastie, Liu et al. 2011).

In contrast to previous hypotheses, LASV N did not bind mRNA cap analogues (m7GTP), similar to the lack of cap binding ability displayed by CCHFV N. The C terminal domain of LASV N is a dsRNA specific 3'-5' exonuclease, which is thought to be an important part of the LASV strategy to evade detection by the innate immune system. The exonuclease activity of the C terminal domain of LASV N is thought to eliminate any dsRNA products resulting from RNA replication that could be recognised the cellular RNA helicases MDA5 or RIG-I (Hastie, Kimberlin et al. 2011). CCHFV N does not have an equivalent domain, despite that fact Guo et al reported DNA specific endonuclease activity of CCHFV N.

1.6 INTERACTION OF CCHFV WITH THE INNATE IMMUNE RESPONSE

CCHFV is classified as an IFN sensitive virus. IFNs are key signalling molecules expressed upon detection of viral (or other microbial) pathogen associated molecular patterns (PAMPs). Induction of IFNs results in the expression of a range of viral restriction factors such as Mx1, 2'5'OAS and IFITM (IFN inducible transmembrane proteins) which not only induces an antiviral state that reduces viral replication, but also initiates the adaptive immune response (Medzhitov and Janeway 1999, Habjan, Andersson et al. 2008). Pre-treatment of cells with IFNα prevents replication of CCHFV (Andersson, Lundkvist et al. 2006), and siRNA experiments indicate this antiviral effect is mediated for the large part by Mx1. Inhibition of CCHFV replication is not complete when expression of Mx1 is experimentally reduced, therefore other components of the innate antiviral response are also thought to play a role in the inhibition of CCHFV replication. To counter this inhibition, CCHFV infection has been shown to delay activation of the innate antiviral response, partially through inhibition of IRF-3 translocation to the nucleus (Andersson, Karlberg et al. 2008). Minor amounts of IFNβ mRNA are detected 24 hrs post-infection and secreted IFN is only detected 48 hrs post infection. Interestingly if cells are treated with IFNα after infection with CCHFV, even as little as 6 hrs post infection, there is no effect on progeny virus titre compared to mock treated cells (Andersson, Karlberg et al. 2008). This suggests that CCHFV delays activation of the innate antiviral response until sufficient viral protein synthesis and replication has occurred, and at this point the induction of the IFN response has no effect on virus replication.

These observations are reflected in the IFN-α/βR$^{-/-}$ and STAT-1 knock out mouse models for CCHFV infection. All IFN-α/βR$^{-/-}$ mice exhibited 100% mortality when infected intraperitoneally with 10 ffu (focus forming units) CCHFV, whereas wild type mice survived doses as high as 10^6 ffu (Bereczky, Lindegren et al. 2010). Similarly STAT-1 knockout mice challenged with CCHFV doses as low as 10 ffu exhibited 100% mortality in 3-5 days whereas wild type mice exhibited neither morbidity nor mortality (Bente, Alimonti et al. 2010). Wild type (WT) mice with functioning IFN pathways do not develop CCHF; CCHFV can replicate within these animals, but they do not succumb to disease, and are asymptomatic (Bente, Alimonti et al. 2010; Bereczky, Lindegren et al. 2010; Bowick, Airo et al. 2012) whereas STAT-1 and IFN-α/βR$^{-/-}$ knockout mice exhibit key features of the disease seen in human cases, and die within 3-5 days.

CCHFV infection delays activation of the innate antiviral response in STAT-1 and IFN-α/βR$^{-/-}$ knockout mice compared to WT mice. At 24 hrs post infection WT mice showed a greater transcriptional expression of genes involved in the inflammatory response compared to knockout mice (Bowick, Airo et al. 2012). By 48 hrs post infection, there was no obvious difference in the expression profiles of these genes in WT compared to STAT1 knockout mice, and actually the transcriptional levels of many of the IFN target genes were increased in knockout mice compared to WT. However expression of these genes 48 hrs post infection is not sufficient to enable these mice to survive CCHFV infection, and indeed may have a detrimental effect: CCHF pathogenesis is thought to be partly contributed to by high expression levels of pro-inflammatory cytokines. It could be that the delay in induction of antiviral genes is critical to CCHFV replication, and while the IFN response is crucial in the control of CCHFV, the kinetics of the IFN response is in fact one of the most important factors in determining the outcome of infection (Bowick, Airo et al. 2012).

CCHFV has several ways of avoiding detection by the innate antiviral response, thereby preventing induction of IFN expression. The main pattern recognition receptors (PRRs) thought to be responsible for CCHFV detection in the cytosol and subsequent expression of IFNα/β are retinoic acid inducible gene 1 (RIG-I) and melanoma differentiation factor 5 (MDA5). Rig-I and MDA5 are both constitutively expressed RNA helicases found in most cell types, whose expression is strongly induced by IFNs in a positive feedback loop of virus detection (Takeuchi and Akira 2010). The agonist for RIG-I has been defined as an RNA with both a triphosphate moiety (PPP) and a blunt-ended base-paired region of ~20 nt at the 5’ end, and as such should recognise the panhandle structures possessed by the genomic RNA segments of CCHFV (Goubau, Deddouche et al. 2013). To counteract this and avoid RIG-I detection, the 5’ termini of CCHFV genomic segments are processed such that the terminal nucleotide possesses a 5’-monophosphate group rather than a triphosphate group (Habjan, Andersson et al. 2008). Therefore CCHFV genomic segments do not bind to and are not recognised by RIG-I, and RIG-I dependent signalling is not induced by the presence of viral genomic segments (Habjan, Andersson et al. 2008).

Although the genomic segments of CCHFV do not induce innate antiviral signalling, other components of the virus particle may be capable of activating other PRRs and inducing IFN expression. Other mechanisms possessed by CCHFV to combat the innate antiviral response reside in the OTU-like cysteine protease encoded by the L segment (the vOTU), as previously described in section 1.3.6. The vOTU of CCHFV deconjugates ubiquitin and ISG15 from their protein targets, thereby potentially interfering with any innate signalling pathways regulated by ubiquitin or ISG15, as well as directly antagonising the antiviral effects of ISG15, which are mediated though the ability of ISG15 to be conjugated to cellular proteins.

1.7 THE INDUCTION OF APOPTOSIS BY CCHFV

Apoptosis is a tightly regulated mechanism of programmed cell death that leads to destruction of cells in a manner that is distinct from necrosis, and that does not result in the activation of an inflammatory response. Hallmarks of apoptosis are DNA fragmentation, membrane ‘flipping’ (the expression of the normally inward facing phosphatidylserine on the plasma membrane outer leaflet), expression of Annexin V on the outer leaflet of the plasma membrane, cell shrinkage, membrane ‘blebbing’ and extensive protein cross linking (Elmore 2007). Caspases are also activated during apoptosis by cleavage of their inactive procaspase form, which is constitutively expressed in the cytoplasm. Active caspases are aspartic acid residue specific cysteine proteases that cleave proteins at specific aspartic acid motifs during apoptosis.

Two distinct apoptotic pathways exist, the extrinsic pathway, which is activated in response to extracellular stimuli, and the intrinsic pathway, which is activated by intracellular stimuli and mediated by mitochondria. Many different stimuli result in the induction of apoptosis, including the detection of viral PAMPs; apoptosis is induced as a result of IFN signalling, cell entry of viruses, activation of protein kinase R (PKR), activation of the transcription factor p53, and ER stress (Everett and McFadden 1999). These stimuli ultimately lead to an increase in the permeability of the mitochondrial outer membrane, and the release of normally sequestered pro-apoptotic proteins. A diagrammatic summary of the intrinsic apoptotic pathway is shown in Figure 1.14.

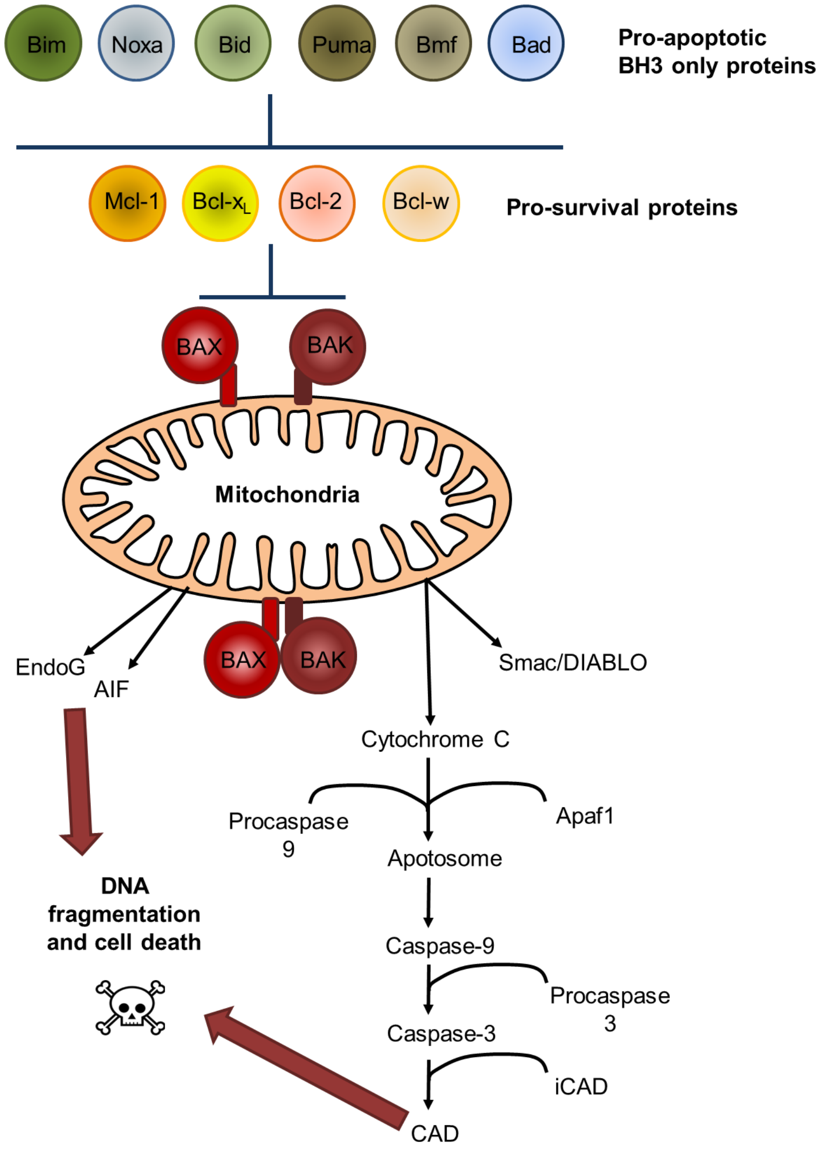


Figure 1.14 Diagrammatic summary of the intrinsic apoptotic pathway

Adapted from (Postigo and Ferrer 2009). The intrinsic apoptotic pathway is regulated by pro-apoptotic and pro-survival members of the Bcl-2 family of proteins. An increase in mitochondrial outer membrane permeability results in the release of cytochrome C from the intermembrane space, and activation of caspase-9 and caspase-3, which ultimately leads to cell death.

The Bcl-2 family of proteins govern mitochondrial membrane permeability and they can either be pro-apoptotic or anti-apoptotic. Some of the pro-apoptotic members include Bcl-10, BAX, BAK, Bid, Bad, Noxa and PUMA, and some of the anti-apoptotic members include Bcl-2, Bcl-x, Bcl-x_L, Bcl-XS, Bcl-w, BAG (Figure 1.14). These proteins have special significance as they can determine whether a cell commits to or aborts the apoptotic process, and the ratio of BAX:Bcl-2 or Bcl-x_L is thought to be able to determine whether a cell lives or dies (Elmore 2007; Rodrigues, Paranhos-Baccala et al. 2012). The transcription factor p53 also has a critical role to play in the regulation of expression of the Bcl-2 family of proteins (Schuler and Green 2001; Elmore 2007).

Apoptotic stimuli can result in dimerisation of two proteins embedded in the outer mitochondrial membrane, BAX and BAK, and an increase in outer mitochondrial membrane permeability. This results in the release of two main groups of pro-apoptotic proteins into the cytosol from the mitochondrial intermembrane space (Saelens, Festjens et al. 2004). This first group of protein includes cytochrome C and SMAC/DIABLO, amongst others, and the second group is released later on during apoptosis, after the cell has committed to die and includes apoptosis inducing factor (AIF) and endonuclease G (EndoG) (Elmore 2007). Cytochrome C and SMAC/DIABLO activate the caspase dependent apoptotic pathway; cytochrome c binds to and activates Apaf 1 as well as procaspase-9, forming the apoptosome (Hill, Adrain et al. 2004). The local concentration of procaspase-9 in the apoptosome results in self cleavage and activation of procaspase-9, and SMAC/DIABLO promotes apoptosis by inhibiting inhibitors of apoptosis (IAPs).

Caspase-8, -9, and -10 are initiator caspases whose role once activated is to cleave and activate the executioner procaspases, procaspase-3, -6, and -7. Activation of the executioner caspases begins the execution phase, the final pathway leading to cell death. Active caspase-3 cleaves a variety of substrates including poly(ADP)ribose polymerase (PARP), inhibitor of caspase activated Dnase (ICAD) to release caspase activated Dnase (CAD), and gelosolin, which once cleaved is involved in actin cleavage and cytoskeletal disruption (Elmore 2007).

AIF and endonuclease G are both members of the second group of proteins to be released from the mitochondrial intermembrane space, and they translocate to the nucleus and act independently of caspases. AIF cleaves DNA into ~50-300 bp fragments, and begins condensation of peripheral nuclear chromatin, and endonuclease G cleaves nuclear chromatin into oligonucleosomal DNA fragments. CAD then leads to oligonucleosomal DNA fragmentation and a more pronounced and advanced chromatin condensation (Elmore 2007). DNA fragmentation is the final stage of apoptosis and indicates cell death.

CCHFV induces apoptosis in infected cells at a relatively late time point after infection (Karlberg, Tan et al. 2011; Rodrigues, Paranhos-Baccala et al. 2012). Expression of active caspase-3 and DNA fragmentation can be detected 48 hrs post infection, and increases 72 hrs post infection when major cytopathic effect (CPE) is observed. The ability of CCHFV to induce apoptosis was shown to be dependent on the ability of the virus to replicate as UV-inactivated CCHFV failed to induce apoptosis. CCHFV infection of hepatocytes induced ER stress, as shown by detection of the unfolded protein response (UPR); ER stress can activate the intrinsic apoptotic pathway and CCHFV was shown to modulate both the intrinsic and extrinsic apoptotic pathways and activate inflammatory mediators such as IL-8 (Rodrigues, Paranhos-Baccala et al. 2012).

Mitochondrial dependent apoptosis was thought to occur in CCHFV infected hepatocytes as elevated levels of BAX mRNA was detected, and although there was a minor increase in Bcl-xL expression, overall the BAX:Bcl-xL ratio was positive. Increased mRNA levels of PUMA and Noxa, both pro-apoptotic Bcl-2 family members was also observed; it is thought that CCHFV induces apoptosis through multiple pathways that involve cross talk between the ER stress response and the mitochondrial mediated apoptotic pathway (Rodrigues, Paranhos-Baccala et al. 2012).

CCHFV N is cleaved at a conserved DEVD motif on the apex of the arm domain by active caspase-3 during infection, resulting in the production of a 30 kDa fragment of CCHFV N (Karlberg, Tan et al. 2011). Interestingly, if the DEVD site is mutated and no longer cleaved by caspase-3 the other N cleavage products become more pronounced. This indicates that while the conserved DEVD motif is the major site for caspase-3 cleavage, there are other sites or motifs that are available for cleavage during apoptosis, especially if the DEVD motif is disrupted. Inhibition of caspase-3 was shown to increase virus titres, implying that caspase-3 cleavage of N is not beneficial for the virus replication cycle, and apoptosis may be induced by the host cell immune response as a way of limiting virus replication (Karlberg, Tan et al. 2011). However the reason why CCHFV infection results in apoptosis so late after infection is unclear; it could be that CCHFV has evolved mechanisms of delaying activation of caspase-3 and apoptosis, similar to the way in which CCHFV delays activation of the IFN response, until sufficient virus replication has occurred to ensure the spread of progeny virions.

1.8 PROJECT AIMS

The N protein of CCHFV is indispensable to virus replication, and therefore represents a potential therapeutic target. The crystal structure of CCHFV N has recently been solved; the aim of this project was to determine the crystal structure of the N protein of the closely related HAZV virus. HAZV is not pathogenic to humans, and it was hoped that structural similarities between CCHFV N and HAZV N would be revealed that would support the use of HAZV as a model for CCHFV infection. The identification of areas of structural similarity could also potentially be exploited in the structure guided design of small molecule inhibitors, whose efficacy against HAZV could be tested under containment level 2 conditions, prior to testing against CCHFV.

Given the limited protein coding capacity of CCHFV and HAZV, it is thought their N proteins would have non-structural roles within the cellular environment during virus replication, roles which are mediated through interaction of CCHFV N and HAZV N with cellular proteins. The aim of this project was also to identify cellular proteins that interact with CCHFV N using a mass spectrometry based approach. We also aimed to determine whether a specific cellular protein, heat shock protein 70 (HSP70) interacted with HAZV N during HAZV replication. We aimed to investigate the role of HSP70 during the replication cycle of HAZV through the use of siRNA and small molecule inhibitors of HSP70. Little is known about the HAZV replication cycle and the final aim of this project was to determine whether HAZV infection resulted in apoptosis, and whether HAZV N is a target for proteolytic cleavage by cellular caspases, as is the case during CCHFV infection.

In summary, the aims of this project were to determine the crystal structure of HAZV N, and to identify cellular proteins and pathways that might be important during the replication cycle of CCHFV and HAZV, such that they might become the targets of host-directed antiviral therapeutic strategies.

CHAPTER 2 : MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Vectors

The HAZV N open reading frame (ORF) optimized for bacterial expression was synthesized by Dundee Cell Products, and was provided cloned into a pET28-SUMO expression vector. The HAZV N ORF was subcloned into a pGEX6P2 expression vector (GE Healthcare); this fusion vector encodes glutathione S-transferase (GST) that is used to N-terminally tag the ORF of interest with GST, and a human rhinovirus 3C (HRV) protease (GE Healthcare) cleavage site. The HAZV N nucleotide sequence was confirmed by DNA sequencing; expression of this HAZV N ORF resulted in the production of a protein whose amino acid sequence matched that of HAZV isolate JC280, GenBank accession number; M86624.1.

The CCHFV N ORF was synthesized by Dundee Cell Products, and was provided cloned into a pCAGGs vector. pCAGGs vectors enable high level gene expression in mammalian cells, driven by the synthetic CAG promoter, that contains a hybrid CMV enhancer/chicken β-actin promoter. The CCHFV N ORF was subcloned into a pEGFP-C2 expression vector (Clontech); this fusion vector contains an enhanced green fluorescent protein (EGFP) tag, and enables the expression of ORFs of interest with an N-terminal EGFP tag. Gene expression from pEGFP-C2 is driven by a CMV promoter and results in strong, constitutive expression of N terminally tagged EGFP-fusion proteins in many mammalian cell types. pEGFP-C2 also contains an SV40 poly(A) signal, which increases RNA stability, and enables high level protein expression. The CCHFV N nucleotide sequence was confirmed as correct by DNA sequencing and matched that of CCHFV strain 10200, accession number; U88410. See appendix I for HAZV N and CCHFV N ORF sequences.

2.1.2 Bacterial strains

Plasmid DNA constructs were amplified, and recombinant GST-HAZV N was expressed, through transformation of competent *Escherichia coli* (*E. coli*) cells. Bacterial strains used were: Subcloning efficiency *JM109* Competent cells (Promega) for initial transformation of ligation reactions; *TOP10* cells prepared in the laboratory using the standard Inoue method for routine retransformation; and the BL21 derivative

Rosetta-2 for isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible protein expression.

2.1.3 Continuous human cell lines

HEK293T (human embryonic kidney cell origin) (Graham, Smiley et al. 1977), HUH-7 (human hepatocellular carcinoma origin) (Nakabayashi, Taketa et al. 1982) and SW13 (human adrenal cortex carcinoma origin) (Leibovitz, McCombs et al. 1973) cells were utilised to express recombinant proteins or for virus replication. HEK293T cells are a human kidney epithelial cell line that constitutively expresses the SV40 large T antigen, HUH-7 cells are a hepatocyte derived carcinoma cell line, originally isolated from a liver tumor and SW13 cells are a human adrenal cortex carcinoma cell line, originally derived from a small cell carcinoma tissue biopsy. 293T cells and HUH-7 cells were obtained from the Health Protection Agency Culture Collections (HPACC), and SW13 cells were a kind gift from Dr Stuart Dowall and Dr Roger Hewson, of Public Health England (PHE).

2.1.4 HAZV strain

HAZV strain JC280 was kindly provided by Dr Stuart Dowall and Dr Roger Hewson (PHE, UK) as an infectious cell culture supernatant from SW13 cells. The cRNA sequence of HAZV S, M, and L segments from strain JC280 has been published on GeneBank with the respective accession numbers: M86624.1, DQ813514.1 and DQ076419.1.

2.2 METHODS

2.2.1 Manipulation of recombinant DNA

2.2.1.1 The polymerase chain reaction (PCR)

During subcloning the cDNA representing HAZV N and CCHFV N ORFs were amplified by PCR, incorporating terminal restriction endonuclease recognition sites, to facilitate insertion into expression plasmids. PCR reactions were performed in 50 µL reaction volumes and contained; 100 ng of template DNA, 0.4 µM of each forward and reverse primer, 0.3 mM of each deoxyribonucleotide triphosphate (dNTPs; Roche), 10 x Thermopol buffer (New England Biolabs) and 1 unit of Vent polymerase (New England Biolabs). Reaction cycles were carried out in a thermocycler (Microfuge) as follows; initial denaturation at 95 °C for 5 minutes; 30 cycles of denaturation at 95 °C for 30 seconds, annealing at 50 °C - 65 °C for 30 seconds (dependent on primer sequence), and elongation at 72 °C for 90 seconds, then final extension at 72 °C for 5 minutes. PCR products were kept at 4 °C until removal from thermocycler; PCR products were then isolated and purified by agarose gel electrophoresis followed by gel extraction.

2.2.1.2 Restriction digest

Plasmid DNA was subjected to restriction enzyme digestion for cloning and diagnostic purposes. Restriction digests were carried out in a 50 µL volume and contained: 1 µg DNA, +/- 1 x bovine serum albumin (BSA; restriction enzyme dependent), 1 x compatible NEBuffer (New England Biolabs), and 2 units of each restriction enzyme (New England Biolabs). Reactions were incubated at 37 °C for 2-3 hrs, or were left at 37 °C overnight. The products of digestion were purified by agarose gel electrophoresis followed by gel extraction using the Zymoclean Gel DNA recovery kit (Zymo Research) as per the manufacturer’s protocol.

2.2.1.3 Agarose gel electrophoresis

Size, integrity and purity of PCR products, plasmid DNA, products of restriction enzyme digestion, and ligations were determined using 1 % agarose gels composed of 0.5 g analytical grade agarose, 50 mL of 1 x TAE buffer (40 mM Tris-acetate, 1 mM ethylene-diamine-tetraacetic acid (EDTA)) and SYBR Safe DNA stain (Life Technologies) diluted 1:10,000. DNA samples were mixed with 1 x GelPilot DNA loading dye (Qiagen) and loaded alongside the molecular marker Hyperladder I (Bioline); gels were then run at 90 V for 30-60 minutes, in 1 x TAE buffer. Blue light transillumination was used to visualise DNA.

2.2.1.4 Ligations

Ligations were carried out in 20 µL reaction volumes containing: 3:1, 6:1 or 10:1 molar ratio of insert:vector DNA, 1 x Ligase Reaction Buffer (30 mM Tris-HCl (pH 7.8), 10 mM MgCl2, 10 mM DTT and 1 mM ATP; Promega) and 1 unit T4 DNA Ligase (Promega). Ligase reactions were incubated at 4 °C overnight, then transformed into DH5α cells.

2.2.1.5 Bacterial transformations

When using subcloning efficiency DH5α competent cells (Life technologies), transformations were carried out according to the manufacturer’s instructions, or transformations were carried out as follows for laboratory-made cells: 1 µL of plasmid DNA at a concentration of 25 - 50 ng/µL, or 1 µL ligation mix were added to 50 µL competent cells that had been thawed on ice. Transformation mixes were incubated on ice for 30 minutes, heat shocked at 42 °C for 45 seconds, and rapidly cooled on ice for 2 minutes. 450 µL of pre-warmed LB (Luria-Bertani) media (10 g tryptone, 5 g yeast extract, 10 g NaCl in 1 L water) was added to transformation mixes, which were then incubated at 37 °C for 1 hr with shaking. Transformed bacteria were pelleted by centrifugation at 300 x *g* for 2 minutes at room temperature, the supernatant was removed, and bacteria were resuspended in 50 µL LB media prior to being spread onto antibiotic-selective LB agar plates and incubated at 37 °C overnight.

2.2.1.6 Starter cultures

Starter cultures of 5 mL LB medium (small-scale) or 50 mL (large scale) were grown from single colonies picked from transformation plates. Single colonies were allowed to grow overnight in LB medium containing the appropriate antibiotic (final concentration 50 µg/mL kanamycin or 100 µg/mL ampicillin); these starter cultures were then used to inoculate overnight cultures either for protein expression or plasmid DNA amplification. Aliquots of starter cultures were mixed in a 70:30 ratio with glycerol and stored frozen at -80 °C as glycerol stocks.

2.2.1.7 Plasmid DNA amplification

Overnight cultures of 200 mL LB medium containing the appropriate concentration of antibiotic were inoculated with 200 µL starter cultures, and the following morning, bacteria were pelleted by centrifugation at 4000 x *g* at 4 °C for 20 minutes. Plasmid DNA was isolated from pelleted bacteria using Plasmid Maxiprep kits (Qiagen) according to the manufacturer's instruction, where isolation of plasmid DNA is based on the alkaline lysis method followed by isopropanol precipitation. Plasmid DNA was also isolated from starter cultures using Spin Miniprep kits (Qiagen) according to the manufacturer's instruction, where isolation of plasmid DNA is based on adsorption to a silica membrane under high-salt conditions. A NanoDrop 1000 (Thermo Scientific) was used to quantify DNA by spectrophotometry.

2.2.2 Protein expression and purification in *E. coli.*

2.2.2.1 Culture growth, IPTG induction and lysis

Large-scale 50 mL starter cultures were grown from stabs of glycerol stocks of pGEX-GST-HAZV N transformed *E. coli* Rosetta-2 cells, 50 mL starter cultures were then used to inoculate 1 L overnight cultures of LB medium containing 100 µg/mL ampicillin. Cultures were grown to OD_{600} 0.6 - 0.8 at 37 °C, then were induced with 100 µM IPTG for 16 hrs at 18 °C. Bacterial cells were recovered by centrifugation at 4000 x *g* at 4 °C for 30 minutes, the supernatant was removed, and cells were either frozen at -20 °C, or resuspended immediately in lysis buffer containing: 100 mM NaCl, 20 mM Tris pH 8, 1 % (v/v) Triton X-100, 1 mg/mL chicken egg white lysozyme (Sigma Aldrich), 1 mM $MgCl_{2,}$ 1x complete protease inhibitor cocktail EDTA-free (Roche), 1 unit (U) DNase and 1 U RNase. 25 mL lysis buffer was used per 1 L bacterial culture. Cells were incubated in lysis buffer on ice for 30 minutes, then lysis was completed by sonication on ice with 15 cycles of 10 seconds on, 5 seconds off, at 8 micron amplitude. Lysates were clarified by centrifugation at 18,000 x g for 30 minutes at 4 °C and the supernatant was retained for GST-HAZV N purification by glutathione affinity chromatography.

2.2.2.2 Purification by glutathione affinity chromatography

Clarified bacterial supernatant containing soluble GST-HAZV N protein was applied to glutathione sepharose 4B resin (GE Healthcare) in batch at room temperature for 1 hr with rotation. Resin with bound GST-HAZV N was transferred to a gravity flow column attached to a peristaltic pump, and unbound bacterial lysate was removed from the resin. Resin was then washed once in 4 volumes of binding buffer (100 mM NaCl, 20 mM Tris pH 8), once in 4 volumes binding buffer containing 1.5 M NaCl to remove RNA derived from the bacterial expression host to generate Apo-HAZV N, (or this step was excluded when RNA-bound HAZV N was purified), then resin was washed once more in 4 volumes of binding buffer. Elution buffer (40 mM reduced glutathione, 5 % glycerol, 50 mM NaCl, 15 mM HEPES pH 7, 1 mM DTT) was used to dissociate GST-HAZV N from the resin in 3 successive elutions. Solutions containing eluted GST-HAZV N were pooled and protein concentration was determined by adding 1 µL protein to a 1:5 dilution of Protein Assay reagent (Bio-Rad) in ddH_2O, and reading the absorbance at 595 nm. Eluted GST-HAZV N was then incubated overnight at 4 °C with HRV 3C protease to cleave the GST tag from HAZV N; HRV 3C protease was expressed and purified in house from glycerol stocks.

2.2.2.3 Size exclusion chromatography

Size exclusion chromatography was used analytically to examine the oligomeric status of HAZV N. A 320 mL HiLoad 26/600 Superdex 75 pg column (GE Healthcare) was used with an Akta Prime pump system at 4 °C, and absorbance at 280 nM was used to monitor protein elution. All buffers used were filtered through 0.2 µM filters and degassed. The column was equilibrated with running buffer (500 mM NaCl, 20 mM Tris pH 7.5, 1 mM EDTA, 1 mM DTT) and the protein was concentrated to a just over 5 mL volume with 10 kDa molecular weight cut off (MWCO) Vivaspin centrifugal concentraters (Sartorius Stedim). The protein sample was then passed through a 0.2 µm syringe filter prior to injection onto the column with a 5 mL loop. The cloumn was run at approximately 2 mL/minute until the void volume (~90 mL) had eluted, then 3 mL fractions were collected until a full column volume of buffer had eluted (320 mL). The column was calibrated using a gel filtration standard kit (Bio-Rad) containing thyroglobulin (670,000 Da), γ-globulin (158,000 Da), ovalbumin (44,000 Da), myoglobin (17,000 Da), vitamin B_{12} (1,350 Da).

2.2.2.4 Ion exchange chromatography

A 1 mL Resource S cation exchange column (GE Healthcare) was used at pH 7 (below the isoelectric point (pI) of HAZV N (8.73)) with an Akta purifier pump system at 4 °C, and absorbance at 280 nM was used to moniter protein elution. All buffers used were 0.2 µM filtered and degassed. The Resource S coumn was equilibrated in buffer A (50 mM NaCl, 15 mM HEPES pH 7, 1 mM DTT); cleaved HAZV N was dialysed into 5 L buffer A overnight at 4 °C using snakeskin dialysis tubing (10,000 Da MWCO, Thermo Scientific), prior to concentration to just over 5 mL volume. The column was run at a flow rate of approximately 1 mL/minute and the protein sample was injected onto the column using a 5 mL loop. 1 mL fractions were collected from this point until the 280 nm trace returned to baseline. To elute bound HAZV N, the NaCl concentration was linearly increased by applying a gradient of 0-100 % buffer B (1 M NaCl, 15 mM HEPES pH 7, 1 mM DTT). Appropriate fractions were pooled and concentrated to 10-12 mg/mL using 10,000 Da MWCO centrifugal concentrators (Sartorius Stedim) and stored either at 4 °C or at -80 °C.

2.2.3 Crystallographic techniques

2.2.3.1 Sparse matrix screening

Initial crystallisation trials were carried out using commercially available sparse matrix screens. The following kits were used: Wizard I and II, Wizard III and IV (Emerald Biosystems), Salt RX, Matrix, Index, and Crystal Screen I and II (Hampton Research). Sitting drops were set up in 96-well MRC-2 drop plates (Molecular Dimensions) using an Oryx 6 robot (Douglas Instruments) with 60 µL reservoire solution. Protein concentrated to 10-12 mg/mL was dispensed into 1 µL droplets containing protein:reservoire ratios of 50:50 or 70:30. Plates were sealed with Viewseal pressure adhesive clear seals (Greiner Bio-One) and incubated at 11 °C, 18 °C or 25 °C.

2.2.3.2 Crystal optimisation

Sparse matrix screen sitting drops were monitered for crystal growth 24 hrs after they were set up, and then periodically for 2 weeks afterwards. Conditions that resulted in crystal growth were further optimised in 24-well plates in hanging drop orientations, using 1-2 µL drops at 50:50 and 70:30 protein:reservoire ratios. Matrices were set up to vary concentration of precipitants, type of precipitants, pH of buffer, protein concentration, and temperature of crystal growth. An additive screen (Hampton Research) was also used in an attempt to improve crystal growth.

2.2.3.3 Crystal harvesting and data collection

Appropriately sized loops (Hampton Research) were used to pick crystals, which were then soaked in a cryprotectant solution containing 1:1 reservoir solution and one of the following cryoprotectants (60 % glycerol, 60 % PEG-400, 60 % PEG-200, 60 % ethylene glycol, or 60 % 2-Methyl-2, 4-pentanediol (MPD)) prior to cyro-cooling in liquid nitrogen. All data were collected at -173 °C at the Diamond Light Source on beamlines I02 and I03. Crystals were screened for diffraction by taking test images at Φ=0° and 90°, using 12,658 eV X-rays (0.9795 Å). Full datasets were collected from crystals that diffracted to better than 4 Å.

2.2.3.4 Data processing and structure solution

X-ray data were integrated using XDS (Kabsch 2010) or iMOSFLM (Leslie 2006; Battye, Kontogiannis et al. 2011) from the Collaborative Computer Project 4 (CCP4) suite (Potterton, Briggs et al. 2003; Winn, Ballard et al. 2011). POINTLESS was used to confirm the space group identity, and to test for twinning, then AIMLESS (Evans 2006; Evans 2011) was used to scale the reduced data from all crystals. Datasets from both Apo-HAZV-N and RNA-bound HAZV N were solved via molecular replacement (MR) with the globular domain of CCHFV N (Protein Data Bank (PDB) accession number 4AKL) residues 1-182, and 296-482 using PHASER (McCoy, Grosse-Kunstleve et al. 2007). BUCANNEER (Cowtan 2006) was used to build an initial model into the calculated electron density maps; this was followed by consecutive cycles of manual building in COOT (Emsley and Cowtan 2004) and structure refinement in REFMAC5 (Murshudov, Skubak et al. 2011). Refined models were analysed with the MolProbity server, models were examined in COOT and figures were made using PyMOL.

2.2.4 Cell culture methods

2.2.4.1 Continuous cell culture

All cells were maintained at 37 °C with 5 % CO_2 in humidified incubators. HEK293T and SW13 cells were grown in Dulbecco's modified Eagle's medium (DMEM; Sigma) supplemented with 10 % (v/v) foetal bovine serum (FBS; Invitrogen), 100 IU penicillin/mL, and 100 µg streptomycin/mL. This is referred to as complete media. HUH-7 cells were grown in complete media supplemented with 1 % non-essential amino acids (NEAA). Cell lines were passaged using trypsin-EDTA (Sigma) 3 times per week as they approached confluency. Cells were seeded at the following densities 24 hrs prior to manipulation to achieve 60-70 % confluency: 96 well plate (10^4 cells/well), 12 well dish (10^5 cells/well), 6 well dish (2 x 10^5 cells/well) and 10 cm^2 dish (10^6 cells/well). Autoclaved glass coverslips (VWR) were placed in wells prior to seeding for immunofluorescence (IF) experiments.

2.2.4.2 Freezing and thawing cells

Cells were frozen for long term storage in liquid nitrogen. Cells were trypsinised and washed 3 times in ice-cold PBS, with centrifugation at 1000 x *g* at 4 °C for 5 minutes between each wash step. Cells were counted, then resuspended in ice-cold FBS supplement with 10 % DMSO, to a final concentration of 10^6 cells/mL. 1 mL cell suspension was aliquoted in each cryovial, which were cooled at a controlled rate in an isopropanol chamber to -80 °C overnight prior to being transferred to liquid nitrogen. Frozen cells were rapidly thawed at 37 °C, centrifuged once at 600 x *g* at 4 °C for 5 minutes, the supernatant was removed and the cell pellet resuspended in complete media and transfered to a 25 cm^2 flask. Cell cultures were expanded when cells reached 80 % confluency.

2.2.4.3 Stable isotope labelling of amino acids in cells culture (SILAC)

Frozen HEK293T cells were thawed, then cultured in isotopically distinct labelled SILAC-DMEM (Dundee Cell Products Ltd) for a minimum of seven population doublings (approximately 2 weeks) to ensure >95 % labelling. SILAC-DMEM recipes used were R0K0 (Light), which contained ^{12}C L-arginine and ^{12}C L-lysine, and R6K4 (Medium) which contained ^{13}C L-arginine and 2D L-lysine. SILAC-DMEM was supplemented with 10 % dialysed foetal calf serum (FCS; Dundee Cell Products Ltd), and passaged using a PBS-based cell dissociation buffer (Invitrogen) instead of trypsin, to avoid the introduction of unlabelled amino acids.

2.2.5 Transfections

2.2.5.1 Calcium phosphate mediated transfection of HEK293T cells

In order to identify cellular interacting partners of EGFP-CCHFV N, SILAC labelled HEK293T cells were transfected via calcium phosphate ($CaPO_4$) with 10 µg plasmid DNA encoding EGFP-CCHFV N or EGFP per 10 cm^2 dish, for immunoprecipitation (IP) experiments (4 x 10 cm^2 dishes were transfected with each construct to ensure the IP of sufficient amounts of EGFP-CCHFV N to identify cellular proteins that co-IP by mass spectrometry). Unlabelled 293T cells were also transfected with EGFP-CCHFV N and EGFP, as well as pCAGGs-CCHFV N via $CaPO_4$. The appropriate amount of plasmid DNA was added to a solution containing 244 mM $CaCl_2$, this was added dropwise to 2 x HBS (274 mM NaCl, 302 µM $(Na_2HPO_4).12H_2O$, 55 mM HEPES pH 7). $CaCl_2$ and 2 x HBS were 0.2 µm filtered. Solutions of calcium chloride, DNA and 2 x HBS were incubated at room temperature for 30 minutes to allow calcium phosphate-DNA precipitates to form, prior to being adding dropwise to HEK293T cells; $CaPO_4$ facilitates binding of the DNA to the cell surface, where it then enters the cell by endocytosis. Fresh complete media was added to cells immediately prior to transfection. Complete media containing transfection mixes were incubated with HEK293T cells overnight; this was replaced with fresh complete media the following morning, or cells were harvested.

2.2.5.2 Lipofectamine mediated transfection of HUH-7 and SW13 cells with pCAGGs-CCHFV N, and siRNA

HUH-7 cells were transfected with pCAGGs-CCHFV N for IF experiments and SW13 cells were transfected with siRNA (small interfering RNA) for protein depletion experiments. siRNA targeting HSP70 (gene HSPA1A; Life Technologies) was a duplex of 21 nucleotides, the sequence of the siRNA from 5’->3’ is as follows: 5’-AAUUUAUACUGCCAUCUUATT-3’. Scrambled siRNA (MISSION universal negative control 1; Sigma Aldrich) was used as a negative control; HSPA1A siRNA and scrambled siRNA were obtained as lyophilised RNA, and stock solutions were made up to the same concentration (10 µM) in nuclease free ddH_2O. SW13 cells (12-well plate) were transfected with 80 pmole HSP70 siRNA per well to deplete HSP70, or 80 pmole scrambled siRNA was transfected as a negative control for off-target effects due to siRNA transfection. RNAiMAX (Life Technologies) was used to transfect SW13 cells with siRNA, and Lipofectamine 2000 (Life Technologies) was used to transfect HUH-7 cells with pCAGGs-CCHFV N.

RNAiMAX and Lipofectamine 2000 are both cationic lipid-based transfection reagents that facility entry of DNA and siRNA into cells, and their transfection protocols are the same; cells are seeded in 12-well plates 24 hrs before transfection such that their confluency is 60-70 % when transfected. 80 pmole siRNA, or 0.71 µg pCAGGs-CCHFV N was added to 100 µL Opti-MEM. In a separate microfuge, 3 µL RNAiMAX or Lipofectamine 2000 was added to 100 µL Opti-MEM; both solutions were mixed gently and allowed to equilibrate for 5 minutes. The Opti-MEM-DNA solution was dropwise added to the Opti-MEM-RNAiMAX/Lipofectamine 2000 solution, mixed gently, and incubated at room temperature for 20 minutes to allow complexes to form between the positively charged cationic lipids and the negatively charged DNA/RNA backbone. The cationic lipids bind to and fuse with the plasma membrane, and entry of DNA into cells is thought to occur through endocytosis. Complete media was removed from cells, and was replaced with serum- and antibiotic-free media (SFM), after cells were washed once in SFM. OptiMEM containing DNA/RNA-cationic lipid complexes was dropwise added to cells in SFM, complexes were removed the following morning and replaced with complete media, or cells were harvested.

Depletion of HSP70 was optimised by varying incubation times and ratios of siRNA:RNAiMAX. Efficacy of HSP70 depletion was analysed by western blot 24 hrs, 48 hrs, and 72 hrs after transfection; western blot membranes were probed with an anti-HSP70/HSP72 mAb (monoclonal antibody) (C92F3A-5; Enzo Life Sciences) that specifically detected the HSPA1A gene product. When assessing the effect of HSP70 depletion on HAZV replication, media containing transfection complexes was removed from cells 20 hrs after transfection, cells were washed once with PBS, then cells were infected with HAZV MOI 1, in 250 µL SFM. Cells were incubated at 37 °C with rocking to allow HAZV to adsorb to plasma membranes, then virus containing SFM was removed and replaced with DMEM containing 2 % FBS. HAZV was allowed to replicate for 6 hrs or 24 hrs prior to harvesting cells.

2.2.6 Analysis of protein expression

2.2.6.1 Preparation of whole cell lysate

Cells were harvested from cell culture dishes and flasks by scraping whilst in media. Cell pellets were recovered by centrifugation at 4 °C at 2000 x *g* for 5 minutes, washed 3 x in PBS, with the same centrifugation step between each wash, then resuspended in lysis buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.5% NP40, 1x EDTA-free complete protease inhibitor (Roche)) for IP, or radioimmunoprecipitation assay (RIPA) buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 % (v/v) NP40 alternative, 0.5 % (w/v) sodium deoxycholate, 0.1 % and 1 x EDTA-free complete protease inhibitor cocktail) for western blot analysis. Cell pellets were incubated in lysis or RIPA buffer for 30 minutes on ice, then cell lysates were clarified by centrifugation at 13,000 x *g* at 4 °C for 10 minutes, and the supernatant containing total protein was either directly utilised or stored at -80 °C for future use.

2.2.6.2 Determining protein concentration by BCA assay

Total protein concentration was determined from whole cell lysates using the micro bicinchoninic acid (BCA; Pierce) protein assay system according to the manufacturer's 96-well plate protocol. Bovine serum albumin (BSA) standards were prepared to create a standard curve ranging from 25 µg/mL - 2000 µg/mL BSA. Whole cell lysates were diluted 1:25 and 50 µL of both standards and diluted sample were mixed with 200 µL assay working reagent (created by diluting reagent B 1:50 with reagent A) then transferred to 96-well plates and incubated at 37 °C for 30 minutes. Protein concentration of samples was determined by measuring their absorbance at 570 nm in a TECAN infinite F50 plate reader and comparing them to the standard curve. Protein samples of the appropriate concentration (2-10 µg) were mixed with 4 x lithium dodecyle sulphate (LDS) buffer (Invitrogen) supplemented with 50 mM DTT reducing agent, prior to denaturation and resolution by SDS PAGE.

2.2.6.3 SDS polyacrylamide gel electrophoresis (SDS PAGE)

5 mL SDS PAGE gels were made with a 12 % (or 15 %) resolving gel (4 mL (5 mL) 30 % bis-acrylamide, 2.5 mL 1.5 M Tris-HCl pH 8.8, 3.3 mL (2.3 mL) ddH$_2$O, 100 µL 10 % SDS, 100 µL 10 % ammonium persulphate (APS), 10 µL TEMED) and a 5 % stacking gel (0.83 mL 30 % bis-acrylamide, 0.63 mL 1 M Tris-HCl pH 6.8, 3.4 mL ddH$_2$O, 50 µL 10 % SDS, 50 µL 10 % APS, 5 µL TEMED). Protein samples were either mixed 1:1 with 2 x denaturing sample buffer (60 mM Tris pH 6.8, 25 % (v/v) glycerol, 2 % (w/v) SDS, 5 % β-mercaptoethanol, 0.01 % (w/v) bromophenol blue) or 2-10 µg protein sample was mixed with 4 x LDS sample buffer (Invitrogen), supplemented with 50 mM DTT reducing agent. Protein samples were denatured by heating at 95 °C for 5 minutes, or at 70 °C for 10 minutes, immediately prior to loading onto gels. Samples were loaded alongside 7.5 µL ColorPlus prestained protein marker (New England Biolabs) or SeeBlue Plus 2 prestained protein marker (Invitrogen), and electrophoresis was performed in SDS running buffer (25 mM Tris, 192 mM glycine, 0.1 % (w/v) SDS) at 200 V for 40-60 minutes.

2.2.6.4 Protein visualisation by Coomassie and silver stain

Proteins resolved by SDS PAGE were directly visualised by incubating gels in Coomassie stain (0.25 % (w/v) Coomassie R-250, 50 % (v/v) methanol, 10 % (v/v) glacial acetic acid) at room temperature for 1 hr, followed by destaining for 1 hr at room temperature in destain solution (40 % (v/v) methanol, 10 % (v/v) glacial acetic acid), and rehydration of gels in ddH_2O prior to imaging. Gels were silver stained using a SilverQuest staining kit (Invitrogen) according to the manufacturer's instructions.

2.2.6.5 Western blot analysis

Proteins were transferred from SDS PAGE gels (section 2.2.6.3) to polyvinylidene fluoride (PVDF) membranes (Immobilon-P Transfer membrane; Millipore), or fluorescence compatible PVDF membranes (FL-PVDF) (Immobilon-FL Transfer membranes; Millipore) using a Trans-Blot semi-dry cell (Bio-Rad) in Towbin buffer (25 mM Tris, 192 mM glycine, 20 % (v/v) methanol) for 1 hr at 15 V. PVDF membranes were blocked for 1 hr at room temperature in 10 % (w/v) non-fat powdered milk (Sigma) in Tris buffered saline supplemented with Tween-20 (TBS-T) (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1 % Tween-20). FL-PVDF membranes were blocked for 1 hr at room temperature in 50 % (v/v) Odyssey blocking buffer (LiCor) in TBS (50 mM Tris-HCl pH 7.5, 150 mM NaCl). Blocking buffer was then replaced with primary antibody diluted to the appropriate concentration in 5 % (w/v) non-fat powdered milk in 1 x TSB-T, or in 50 % Odyssey blocking buffer in TBS, either for 1 hr at room temperature or at 4 °C overnight. A full list of primary antibodies used to detect immobilised proteins is provided in Table 3.

PVDF membranes were then extensively washed in TBS-T, before incubating with the appropriate secondary antibody conjugated to horseradish peroxidase (HRP), under the same conditions and dilutions as for the primary antibody incubation. FL-PVDF membranes were extensively washed in TBS, then incubated with the appropriate fluorescently labelled secondary antibody under the same conditions and dilutions as for the primary antibody incubations. A full list of secondary antibodies used is provided in Table 4. PVDF membranes were washed in TST-T, once in water, and proteins were visualised using enhanced chemiluminescence on photographic film. Enhanced chemiluminescence solution (ECL) was made with a 1:1 mix of each of the following solutions, and applied to the membrane: solution 1, 25 mM Luminol (3-aminophthalydazide), 0.3 mM *p*-coumaric acid, 100 mM Tris-HCl pH 8.5; solution 2, 0.01 % (v/v) hydrogen peroxide, 100 mM Tris-HCl pH 8.5. After incubation of membranes for 1 minute in ECL, they were exposed to film (CL-Xposure film; Thermo Scientific) and proteins were visualised after fixation and development of film with a Konica SRX-101A processor. FL-PVDF membranes were washed in TBS, once in water, then fluorescently labelled secondary antibodies were directly visualised using a LiCor Odyssey Sa Infrared imaging system (LiCor).

2.2.6.6 Immunofluorescence (IF)

IF experiments were carried out to visualise the subcellular localisation of various proteins. Cells were grown on 19 mm glass coverslips housed in 12 well plates prior to manipulation and subsequent fixation. Media was removed from cells, which were washed twice in PBS, prior to fixation of cells either in formalin (37 % formaldehyde (BHD Chemicals) diluted to 4 % in PBS)) for 10 minutes at room temperature, or 4 % (w/v) paraformaldehyde (4 % PFA) for 20 minutes at room temperature. The fixative was removed from cells following the appropriate incubation, cells were washed 3 times with PBS, then either stored in PBS at 4 °C, or cells were permeabilised in 0.5 % (v/v) Triton-X 100 in PBS for 12 minutes at room temperature. Permeabilised cells were washed 3 times in PBS, then blocked in 2 % (v/v) FBS in PBS for 30 minutes.

Primary and secondary antibody incubations were both performed with antibodies diluted to the appropriate concentration in 2 % (v/v) FBS in PBS; after blocking, cells were incubated with 50 µL of the appropriate primary antibody for 1 hr at room temperature. Cells were then washed in PBS and incubated with 50 µL of the appropriate fluorescently labelled secondary antibody for 1 hr at room temperature. A list of the primary and secondary antibodies used for IF, and their concentrations, is provided in table 3 and 4. Cells were washed in PBS prior to mounting onto glass slides using ProLong Gold anti-fade reagent (Invitrogen) containing DAPI (4',6-diamidino-2-phenylindole) to allow visualisation of cell nuclei. Mounted coverslips were allowed to cure at room temperature for 24 hrs prior to visualisation, and kept at 4 °C in the dark for long term storage. Slides harbouring fixed and stained cells were analysed by confocal microscopy; confocal images were captured on either an inverted LSM (laser scanning microscope) 510 META Axiovert 200M, or an upright LSM 510 META Axioplan microscope (Carl Zeiss Ltd). The excitation and emission spectra of the lasers used are as follows; for the detection of DAPI, excitation wavelength 405 nm, emission wavelength 450 nm, for the detection of Alexa Fluor 488 conjugated secondary antibodies, excitation wavelength 488 nm, emission wavelength 518 nm, and for the detection of Alexa Fluor 568 conjugated secondary antibodies, excitation wavelength 543 nm, emission wavelength 590 nm. All pinholes were set to allow optical sections of 1-2 µm to be captured.

2.2.6.7 Live cell imaging

Cells were seeded into glass bottom plates prior to transfection with EGFP-CCHFV N or EGFP (as described in section 2.2.5.1). 24 hrs after transfection, complete media was replaced with CO_2 independent, phenol red-free media (Life Technologies) and images of cells expressing EGFP or EGFP-CCHFV N were captured on the inverted LSM 510 META confocal microscope (Carl Zeiss). Cells were maintained at 37 °C with a heated stage throughout imaging. EGFP allowed direct visualisation of EGFP-CCHFV N in transfected cells when illuminated with light of wavelength 488 nm, the excitation wavelength of EGFP.

2.2.7 Immunoprecipitations

2.2.7.1 GFP-Trap IPs

Immunoprecipitation (IP) was used to enrich EGFP-CCHFV N and EGFP from SILAC labelled 293T cells using a GFP-Trap (chromotek). GFP-Traps are composed of agarose beads covalently conjugated to camelid antibodies (only contain the antibody ‘heavy’ chain) that specifically recognise and bind to EGFP. Lysates from 4 x 10 cm^2 dishes containing SILAC labelled cells transfected with EGFP or EGFP-CCHFV N were prepared as described in section 2.2.6.1. EGFP was transfected into Medium (R6K4) labelled cells and EGFP-CCHFV N was transfected into Light (R0K0) labelled cells. Two IPs were performed separately, each using 30 µL GFP-Trap bead slurry. GFP-Trap beads were pre-equilibrated in ice-cold dilution buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, and 1 x protease inhibitor cocktail (Roche)) as follows; GFP Trap beads were sedimented by centrifugation at 2700 x *g* at 4 °C for 2 minutes, the storage buffer (20 % ethanol) was removed, and GFP-Trap beads were washed twice in dilution buffer with the above centrifugation step repeated between each wash, then resuspended in 500 µL dilution buffer. Cell lysates from EGFP and EGFP-CCHFV N transfected cells were adjusted to a total volume of 1 mL with dilution buffer, equilibrated GFP-Trap beads were centrifuged at 2700 x *g* for 2 minutes at 4 °C, dilution buffer was removed, and GFP-Trap beads were resuspended in 1 mL cell lysates.

Cell lysates were incubated with GFP-Trap beads for 2 hrs with rotation at 4 °C to allow anti-GFP antibodies to bind EGFP and EGFP-CCHFV N. EGFP or EGFP-CCHFV N bound GFP-Trap beads were centrifuged at 2700 x *g* to sediment GFP-Trap beads, then supernatant containing unbound cellular proteins was removed. GFP-Trap beads were then washed once in dilution buffer, once in dilution buffer containing 275 mM NaCl and proteins were eluted from GFP-Trap beads by heating at 95 °C for 10 minutes in the presence of 50 µL 2 x LDS-sample buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 x protease inhibitor cocktail, 1 x LDS and 50 mM DTT). Equal volumes (15 µL) of material from EGFP and EGFP-CCHFV N IPs were combined and sent to Dundee Cell Products Ltd for MS analysis. GFP-Trap IPs were also performed on unlabelled cells transfected with EGFP or EGFP-CCHFV N.

2.2.7.2 Dynabead IPs

Recombinant protein G-coupled Dynabeads (Novex) were used to IP untagged CCHFV N and HAZV N from cell lysates using anti-CCHFV N or anti-HAZV N antibodies (section 3.2.4 describes the generation of these antibodies). Recombinant Protein G is a bacterial protein that binds antibodies via their Fc region. Dynabeads are uniform, 2.8 µm superparamagnetic beads with recombinant protein G covalently conjugated to their surface. Dynabead IPs were carried out using a Dynabead Protein G immunoprecipitation kit (Life Technologies) according to the manufacturer's instructions. 50 µL Dynabeads were transferred into 1.5 mL microfuge tubes, and placed against a magnet; beads migrated to the side of the tube facing the magnet, the supernatant was removed, and replaced with 10 µL anti-CCHFV N or anti-HAZV N diluted in 200 µL Ab Binding and washing buffer (Life Technologies). Dynabeads were incubated with anti-CCHFV N or anti-HAZV N for 30 minutes with rotation at room temperature to allow antibodies to bind Protein G, then tubes were placed next to a magnet and supernatants containing unbound antibodies were removed. Dynabeads were washed once in Ab binding and washing buffer, then anti-CCHFV N or anti-HAZV N bound Dynabeads were incubated with cell lysates prepared as described in section 2.2.6.1, for 15 minutes at room temperature with rotation.

Dynabeads with antigen-antibody complexes bound were washed three times in wash buffer (Life Technologies). Dynabeads with bound antigen-antibody complexes were transferred to a new microfuge tube, and proteins were eluted by heating at 95 °C for 10 minutes in 2 x LDS buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 x protease inhibitor cocktail, 1 x LDS and 50 mM DTT).

2.2.8 Mass spectrometry (MS) analysis of SILAC labelled GFP-Trap IPs

2.2.8.1 SDS PAGE, tryptic digest, and NanoLC MS ESI MS/MS analysis

MS analysis was performed by Dundee Cell Products Ltd. Protein samples generated by EGFP immunoprecipitations were separated by one- dimensional SDS-PAGE (4-12% bis-Tris Novex mini-gel, Invitrogen). The resulting separated proteins were cut from the gel in 10 slices and subjected to in-gel digestion with trypsin. Trypsin digested peptides were separated using an Ultimate U3000 nanoflow LC-system (Dionex Corporation) consisting of a solvent degasser, micro and nanoflow pumps, flow control module, UV detector and a thermostated autosampler. A sample volume of 10 µL (comprising 2 µg) was loaded with a constant flow of 20 µL/min onto a PepMap C18 trap column (0.3 mm x 5 mm, Dionex Corporation). After trap enrichment peptides were eluted onto a PepMap C18 nano column (75 µm x 15 cm, Dionex Corporation) with a linear gradient of 5-35 % solvent B (90 % acetonitrile with 0.1% formic acid) over 65 min at a constant flow rate of 300 nL/min. The HPLC system was coupled to a LTQ Orbitrap XL (Thermo Fisher Scientific Inc) via a nano ES ion source (Proxeon Biosystems). The spray voltage was set to 1.2 kV and the temperature of the heated capillary was set to 200 ^{o}C. Full scan MS survey spectra (m/z 335-1800) in profile mode were acquired in the Orbitrap with a resolution of 60,000 after accumulation of 500,000 ions. The five most intense peptide ions from the preview scan in the Orbitrap were fragmented by collision-induced dissociation (normalized collision energy 35%, activation Q 0.250 and activation time 30 ms) in the LTQ after the accumulation of 10,000 ions. Maximal filling times were 1,000 ms for the full scans and 150 ms for the MS/MS scans. Precursor ion charge state screening was enabled and all unassigned charge states as well as singly charged species were rejected. The dynamic exclusion list was restricted to a maximum of 500 entries with a maximum retention period of 90 seconds and a relative mass window of 10 ppm. The lock mass option was enabled for survey scans to improve mass accuracy. The data were acquired using Xcalibur software.

2.2.8.2 Peptide quantification

Quantification was performed with MaxQuant version 1.0.7.4 (Cox and Mann 2008) and was based on two-dimensional centroid of the isotope clusters within each SILAC pair. To minimize the effect of outliers, protein ratios were calculated as the median of all SILAC pair ratios that belonged to peptides contained in the protein. The generation of peak list, SILAC- and extracted ion current-based quantitation, calculated posterior error probability, false discovery rate based on search engine results, peptide to protein group assembly, and data filtration and presentation was carried out using MaxQuant. The derived peak list was searched with the Mascot search engine (version 2.1.04; Matrix Science, London, UK) against a concatenated database combining 80,412 proteins from International Protein Index (IPI) human protein database version 3.6 (forward database), and the reversed sequences of all proteins (reverse database). Proteins were quantified if at least one MaxQuant-quantifiable SILAC pair was present. Identification was set to a false discovery rate of 1 % with a minimum of 2 quantifiable peptides.

2.2.9 Virological techniques

2.2.9.1 Virus infection and propagation

Infectious HAZV was used as the clarified cell culture supernatant from infected SW13 cells. HAZV was diluted to the desired MOI in SFM, then complete media was removed from appropriately seeded cells, which were washed once in PBS. PBS was removed from cells and replaced with SFM containing HAZV. Cells were incubated with HAZV at 37 °C with gentle rocking for 1 hr to allow HAZV to adsorb to cell surfaces, SFM containing HAZV was then removed from cells, cells were washed once in PBS, then incubated with DMEM supplemented with 2 % FBS at 37 °C, 5 % CO_2 in humidified incubators. In all experiments using HAZV, mock infected cells were included as a control; mock infections were carried out as described above with the exception that the SFM did not contain HAZV.

HAZV was propagated in SW13 cells seeded in T175 cm^2 flasks, when cells reached 60-70 % confluency. Cells were infected as described above with HAZV at an MOI 0.001. HAZV was allowed to replicate for 4 days, then the supernatant containing infectious HAZV was harvested, and clarified by centrifugation at 4000 x *g* at 4 °C for 20 minutes. Clarified supernatant was directly used to infect SW13 cells. The titre of infectious HAZV was estimated by plaque assay (described below); typical yields from virus propagation experiments were between 1 x 10^5 and 1 x 10^7 plaque forming units per mL (pfu/mL).

2.2.9.2 HAZV titre determination by plaque assay

Plaque assays were performed in duplicate in 12-well plates; 25 µL HAZV was diluted by a factor of 10 in 225 µL SFM (10^{-1} dilution), which was further serially diluted to generate 10^{-2} – 10^{-6} dilutions. These HAZV dilutions were used to infect SW13 cells as described in section 2.2.9.1. 2 x 10^5 SW13 cells were plated out in 12 well plates for plaque assays, ensuring cells were ~90-95 % confluent when infected. After HAZV adsorption, virus was removed from cells and replaced with complete media diluted 1:1 with 1.6 % (w/v) high viscosity carboxy-methyl cellulose (CMC; Sigma). The CMC/complete media overlay prevented HAZV from migrating through media, and ensured that following secretion from an infected cell, HAZV particles could only infect cells that immediately neighbour the previously infected cell. HAZV is cytopathic, therefore the CMC/complete media overlay ensured that over the course of the assay (6 days), plaques were formed as HAZV infection resulted in cell death in very defined regions.

Cells were incubated for 6 days, then fixed in 20 % (v/v) formalin for 1 hr. Cells were washed 3 time in ddH_2O, then plaques were visualised by staining cells with crystal violet solution (0.1 % (w/v) crystal violet, 20 % ethanol) for 15 minutes. The chromophore containing component of crystal violet is positively charged at neutral pH and binds to the negatively charged cell membrane; the cell monolayer in a plaque assay therefore appears purple, and areas where HAZV has replicated and killed cells (plaques) are not stained and appear transparent. Virus titre was estimated using the following equation:

$$Virus\ titre\ (pfu\ per\ mL) = \frac{number\ of\ plaques}{(dilution\ factor\ x\ volume\ (mL))}$$

2.2.9.3 Purification of HAZV by density gradient centrifugation

HAZV was purified from clarified cell culture supernatants through the use of density gradient centrifugation to remove the culture media and the majority of cellular components. A continuous 5 % - 25 % iodixanol gradient was prepared at least 24 hrs prior to use by sequentially layering 1 mL of 25 %, 20 %, 15 %, 10 % and 5 % (v/v) iodixanol diluted in TNE buffer (0.01 M Tris pH 7.4, 0.1 M NaCl, 1 mM EDTA) into a 13 x 51 mm PolyClear open-top centrifuge tube (SETON). Between each layer application the gradient was frozen on dry ice. Following application of the final layer the gradient was stored at -80 °C. Before use, the gradient was thawed at room temperature overnight to allow the layers to diffuse into one another. Iodixanol supplied as a 60 % solution in water (‘OptiPrep’; density 1.32 g/mL) was used as the density gradient medium, as iodixanol is non-ionic, non-toxic to cells, and is metabolically inert.

Clarified supernatant containing infectious HAZV was mixed with 50 % (w/v) PEG 6000 in TNE buffer, such that the final concentration of PEG 6000 was 10 % (w/v). HAZV was precipitated overnight at 4 °C as this mixture was continually stirred. Precipitated HAZV was recovered by centrifugation at 4000 x *g* for 30 minutes at 4 °C, and the resulting pellet was resuspended in 1 mL of TNE buffer and layered over the 5 % - 25 % continuous iodixanol gradient. The resuspended HAZV pellet and gradient were ultracentrifuged at 250,000 x *g* for 2.5 hrs at 4 °C in an AH650 swinging rotor. After ultracentrifugation an opalescent band was observed ~ 2/3 way down the gradient; this was removed using a 1 mL syringe and a 25 gauge needle (~600 µL), mixed gently, and frozen at -80 °C. This fraction containing infectious HAZV was plaque assayed to estimate infectious virus titre, and BCA assayed to estimate protein concentration. A sample containing ~11 µg protein was denatured by heating for 10 minutes at 95 °C in the presence of LDS and DTT and sent for MS analysis.

2.2.10 Mass spectrometry analysis of the fraction containing infectious HAZV

MS analysis of the fraction containing infectious HAZV was performed to identify cellular proteins that co-purified with infectious HAZV. MS analysis was carried out by Dr Stuart Armstrong at the University of Liverpool, and is described below.

2.2.10.1 SDS-PAGE and Tryptic digestion

The viral protein preparation and controls were run approximately 1 cm into a NuPAGE® (Life Technologies) precast 4-12% gel Tris-Bis gradient gel and the proteins were stained with colloidal Coomassie Blue G-250 (Simply Blue^{TM}, Life Technologies) according to the manufacturer's instructions. The entire gel lane was excised and cut into smaller pieces (approx. 1mm^3). Gel pieces were destained in 50 % acetonitrile/50 mM ammonium bicarbonate (pH ~8), reduced for 30 minutes at 37 °C with 10 mM dithiothreitol (Sigma) in 50 mM ammonium bicarbonate and alkylated with 55 mM iodoacetamide (Sigma) in 50 mM ammonium bicarbonate for 30 minutes in the dark at room temperature. Gel pieces were washed for 15 minutes in 50 mM ammonium bicarbonate and dehydrated with 100% acetonitrile. Acetonitrile was removed and the gel plugs rehydrated with 0.01 µg/µL proteomic grade trypsin (Sigma) in 50 mM ammonium bicarbonate. Digestion was performed overnight at 37 °C. Peptides were extracted from the gel plugs using successive 15 minute incubations of 3% (v/v) acetonitrile 0.1% (v/v) TFA. Peptide extracts were pooled and reduced to dryness using a centrifugal evaporator (Microfuge concentrator plus) and re-suspended in 3% (v/v) acetonitrile, 0.1% (v/v) TFA for analysis by mass spectrometry.

2.2.10.2 NanoLC MS ESI MS/MS analysis

Peptide mixtures (2 µL) were analyzed by on-line nanoflow liquid chromatography using the nanoACQUITY-nLC system (Waters MS technologies, Manchester, UK) coupled to an LTQ-Orbitrap Velos (ThermoFisher Scientific, Bremen, Germany) mass spectrometer equipped with the manufacturer's nanospray ion source. The analytical column (nanoACQUITY UPLCTM BEH130 C18 15cm x 75µm, 1.7µm capillary column) was maintained at 35 ^{0}C and a flow-rate of 300 nL/min. The gradient consisted of 3-40 % acetonitrile in 0.1 % formic acid for 150 minutes then a ramp of 40-85 % acetonitrile in 0.1 % formic acid for 5 minutes. Full scan MS spectra (m/z range 300-2000) were acquired by the Orbitrap at a resolution of 30,000. Analysis was performed in data dependant mode. The top 20 most intense ions from MS1 scan (full MS) were selected for tandem MS by collision induced dissociation (CID) and all product spectra were acquired in the LTQ ion trap.

Spectral data were transformed to .mgf files with Proteowizard (version3.0, (Chambers, Maclean et al. 2012) and used for peptide identification using the Mascot (version 2.3.02, Matrix Science) search engine. Tandem MS data were searched against the HAZV and human predicted proteomes (UniProt Release 2014_02). Mascot search parameters were as follows; precursor mass tolerance set to 10ppm and fragment mass tolerance set to 0.8 Da. One missed tryptic cleavage was permitted. Carbamidomethylation (cysteine) was set as a fixed modification and oxidation (methionine) set as a variable modification. Mascot search results were further processed using the machine learning algorithm Percolator. The false discovery rate was < 1%. Individual ion scores > 13 indicated identity or extensive homology ($p < 0.05$). Approximate, label-free, spectral counting based quantitation of the proteins was determined using the Exponentially Modified Protein Abundance Index (emPAI) calculated automatically by the mascot search engine (Ishihama, Oda et al. 2005).

2.2.11 Inhibitor studies

In all assays the inhibitors remained in the medium throughout the experiment, and the final concentration of DMSO in the medium did not exceed 0.2 % (v/v).

2.2.11.1 Inhibition of HSP70 with VER155008 and pifithrin-µ

The small molecule inhibitors VER155008 (Sigma) and pifithrin-µ (PIF; Sigma) (a commonly used synonym for PIF is: 2-Phenylethynesulfonamide (PES)) were used to assay the effect of inhibition of HSP70/HSC70 on HAZV replication. VER and PIF were both dissolved in dimethyl-sulphoxide (DMSO; Sigma) to make a 50 mM stock, which was aliquoted and stored at -20 °C. Immediately prior to use, 50 mM stocks were thawed and diluted with DMSO to make 10 mM aliquots; these were further diluted in DMEM supplemented with 2 % FBS to the appropriate concentration (between 40 µM and 0.04 µM). 1 x 10^5 SW13 cells were seeded in 12 well plates such that cells were 60-70 % confluent when complete media was removed from cells, cells were washed once in PBS, and media containing the appropriate concentration of PIF and VER was added to cells. After 1 hr, cells were infected with HAZV at MOI 1 as described in section 2.2.9.1; after adsorption HAZV was removed from cells and replaced with fresh media containing the appropriate concentrations of PIF and VER. 24 hrs after infection cell culture supernatants and whole cell lysates were harvested and plaque assayed (section 2.2.9.2), or analysed by western blot (section 2.2.6.5).

2.2.11.2 Inhibition of caspase-3 and executioner caspases with Z-DEVD-FMK and Z-FA-FMK

Z-DEVD-FMK (BD Biosciences) and Z-FA-FMK (Santa Cruz Biotechnology) were used to assess the effect of inhibiting caspase-3 and the executioner caspases on HAZV replication. Z-DEVD-FMK and Z-FA-FMK were reconstituted in DMSO to make a stock concentration of 20 mM. Immediately prior to use aliquots were thawed and diluted in DMEM supplemented with 2 % FBS to a final concentration of 20 µM or 40 µM. SW13 cells were seeded at 1 x 10^5 cells per well in 12-well plates, 24 hrs later cells were infected with HAZV at an MOI 1 as described in section 2.2.9.1. After adsorption of HAZV to cell membranes, SFM containing HAZV was removed from cells and replaced with DMEM supplemented with 2 % FBS, and the appropriate concentration of either Z-DEVD-FMK, or Z-FA-FMK. Cell culture supernatants and whole cell lysates were harvested 24, 48, or 72 hrs after infection and plaque assayed (section 2.2.9.2) or analysed by western blot (sections 2.2.6.5).

2.2.12 MTT assay

MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assays were used to assess the cytotoxity of siRNA, Z-DEVD-FMK, Z-FA-FMK, PIF, and VER treatments on SW13 cells. The MTT assay is a colorimetric assay to determine cell viability, that is based on the cleavage of yellow tetrazolium salt (MTT) by mitochondrial enzymes to form a soluble, blue formazan product; the amount of formazan produced is directly proportional to the number of living, not dead cells, and can be detected by absorption of light at 570 nm. 1 x 10^4 SW13 cells were seeded in 96 well plates 24 hrs prior to treatment with siRNA, Z-DEVD-FMK, Z-FA-FMK, PIF and VER; 24 hrs, 48 hrs, or 72 hrs after treatment, MTT assays were performed using the CellTiter 96 Non-radioactive cell proliferation assay kit (Promega), according to the manufacturer's instructions. A dye solution containing MTT was added directly to cells in 96 well plates, cells were incubated at 37 °C for 4 hrs, prior to the addition to the solubilisation/stop solution to wells to solubilize the formazan product. After overnight incubation, absorbance at 570 nm was recorded with the TECAN infinite F50 plate reader. In each case viability of cells treated with siRNA, Z-DEVD-FMK, Z-FA-FMK, PIF or VER was compared to cells treated with either 0.1 % or 0.2 % (v/v) DMSO. Cell viability at these concentrations of DMSO did not vary from untreated cells.

2.2.13 TUNEL (terminal dUTP nick-end labelling) assay

DNA fragmentation and apoptosis induced by HAZV was detected with an *In Situ* Cell Death Detection Kit, Fluorescein (Roche) according to the manufacturer's instructions. SW13 cells were seeded onto glass coverslips housed in 6 well plates 24 hrs prior to infection. Cells were infected with HAZV at MOI 1 or mock infected (section 2.2.9.1), then 24, 48, or 72 hrs after infection, cells were fixed in 4 % (w/v) PFA (section 2.2.6.6). Cells were permeabilised in 0.1 % (v/v) Triton-X 100 diluted in 0.1 % (w/v) sodium citrate for 5 minutes on ice. DNA strand breaks were then identified by labelling free 3'-OH DNA ends using the *In Situ* Cell Death Detection Kit, Fluorescein (Roche); this kit contains the enzyme terminal deoxynucleotidyl transferase (TdT) which catalyzed the polymerization of fluorescently labelled nucleotides to free 3'-OH DNA ends in a template independent manner, when cells were incubated with the TUNEL reaction mix for 1 hr at 37 °C. Fluorescein labelled nucleotides incorporated into nucleotide polymers were detected and quantified by fluorescence microscopy using the upright LSM 510 META confocal microscope (Carl Zeiss Ltd).

2.3 LIST OF ANTIBODIES

2.3.1 Primary antibodies used in western blot analysis and/or IF

Table 3 Primary antibodies used in western blot analysis and/or IF staining

Target	Vendor	Catalogue number	WB dilution	IF dilution
EGFP	Santa Cruz	sc-8334	1:1000	N/A
vimentin	Santa Cruz	sc-966	1:1000	N/A
actin	Sigma Aldrich	A3853	1:1000	N/A
B23	Santa Cruz	sc47725	1:1000	N/A
p53	Cell Signaling	9282	1:1000	N/A
GAPDH	Abcam	ab8245	1:5000	N/A
CCHFV N	in house	N/A	1:10,000	1:1000
HAZV N	in house	N/A	1:5000	1:500
HSP70/HSC70	Abcam	Ab6535	1:1000	1:100
HSP70	Enzo Life sciences	ADI-SPA-810-D	1:1000	N/A
HSC70	Santa Cruz	sc-7298	1:1000	N/A
Cleaved caspase-3 (Asp175)	Cell Signaling	9664	1:1000	N/A
Caspase-3	Cell Signaling	9665	1:1000	N/A
PARP	Cell Signaling	9542	1:1000	N/A
Cleaved PARP	Cell Signaling	5625	1:1000	N/A
Caspase-9	Cell Signaling	9508	1:1000	N/A
Cleaved caspase-9 (Asp330)	Cell Signaling	7237	1:1000	N/A
Caspase-7	Cell Signaling	12827	1:1000	N/A
Cleaved caspase-7 (Asp198)	Cell signaling	8438	1:1000	N/A

2.3.2. Secondary antibodies used in western blot analysis and/or IF

Table 4 Secondary antibodies used in western blot or IF analysis.

Secondary antibodies were conjugated to either horse radish peroxidase (HRP) or a fluorescent label. These are detailed below.

Target-primary antibody species	Conjugation	Vendor	Catalogue number	WB dilution	IF dilution
Mouse	HRP	Sigma	A4116	1:5000	N/A
Rabbit	HRP	Sigma	A6154	1:5000	N/A
Rat	HRP	Sigma	A5795	1:5000	N/A
Sheep	HRP	Sigma	A3415	1:5000	N/A
Sheep (Light chain specific)	HRP	Jackson ImmunoResearch	213-032-177	1:10,000	N/A
Mouse	Alexa Fluor 568	Invitrogen	A11061	N/A	1:250
Sheep	Alexa Fluor 488	Invitrogen	A11015	N/A	1:250
Mouse	IRDye 680RD	LiCor	926-68072	1:10,000	N/A
Rabbit	IRDye 800CW	LiCor	926-32213	1:10,000	N/A
Sheep	IRDye 800CW	LiCor	926-32214	1:10,000	N/A

CHAPTER 3: DETERMINING THE CRYSTAL STRUCTURE OF HAZV N

3.1: CHAPTER INTRODUCTION

High resolution X-ray crystallography was used to determine the structure of ~90 % of all deposits currently in the Protein Data Bank (PDB). During an X-ray crystallography experiment X-rays are diffracted by electrons within a crystal; collection of this data allows the calculation of an electron density map, into which an atomic model of the molecule can be built. Structure solution by X-ray crystallography poses a number of difficulties; some of these are fundamental (e.g the ‘phase problem’, described in section 3.3.2, along with the method used in this study to address it in section 3.3.4), and other difficulties are practical (e.g. crystal growth and manipulation). However X-ray crystallography remains the prefered option for high resolution structure solution due to the ability to achieve resolutions that are theoretically only limited by crystal quality and X-ray wavelength (< 1 Å).

Other techniques for macromolecular structural analysis, such as nuclear magnetic resonance (NMR) and single particle reconstruction of cryo-electron microscopy (EM) images, although they have advantages and disadvantages over X-ray crystallography, can only achieve resolutions equivalent to 3-3.5 Å and > 10 Å, respectively. These resolutions are not adequate for structure-guided drug design. New techniques are being developed to overcome some of the current limitations of X-ray crystallography, such as neutron diffraction (uses the diffration of neutrons from protons within the crystal) and X-ray free-electron lasers (X-FEL) (produces intensely brilliant, ultrashort X-ray pulses, over 10 times the brightness of synchrotron radiation), however these are not yet in widespread use.

This chapter describes the development of a high yield expression system for full length HAZV N, its purification using GST-resin and ion exchange chromatography, and crystallisation trials that result in the determination of the crystal structure of HAZV N. As a result both apo-HAZV N (RNA free) and RNA bound HAZV N structures were solved. In addition purified HAZV N was used to generate polyclonal antibodies targeting HAZV N, the uses of these antibodies are described in chapter 4 and chapter 5.

3.1.1 Introduction to bacterial expression of HAZV N

Structural studies require the abundant production of homogeneous protein; over-expression of recombinant proteins using prokaryotic *E. coli* based expression systems is the most common method of achieving this. The plasmid used here (Figure 3.1) is a pGEX-based vector in which expression of the target protein is under the control of a *tac* promoter, a hybrid promoter which contains *lac*UV5 elements. In this system IPTG induction directly alleviates repression of the target gene, which results in its transcription by the RNA polymerase, as the target protein itself is under lac repression. There are several advantages associated with bacterially expressing proteins, including homogeneity, low cost, high yield and ease of scale up. One of the disadvantages of expressing eukaryotic or viral genes in bacteria is the problem of biased codon usage; if the codon bias of the gene to be expressed differs significantly from that used by *E. coli*, there may not be sufficient concentrations of *E. coli* tRNA targeting rarely used codons to allow optimal translation. This can be overcome by co-expressing rarely used tRNAs, or by altering the target gene to change rare to more commonly used codons, optimal for bacterial expression. This section outlines the process of optimising bacterial expression and purification of full length HAZV N protein.

3.2 RESULTS: EXPRESSION AND PURIFICATION OF HAZV N

3.2.1 Cloning of HAZV N ORF into the pGEX6P2 expression vector

The HAZV N ORF optimised for bacterial expression was synthesised by Dundee Cell Products, and was provided cloned into a pET 28-SUMO expression vector, using 5' BamHI and 3' XhoI restriction sites. This construct did not provide optimal yields of HAZV N, so to improve expression the HAZV N ORF was subcloned into the pGEX6P2 expression vector, which has an N terminal glutathione S-transferase (GST) tag and a human rhinovirus (HRV) 3C protease cleavage site (Figure 3.1 (B)). The GST tag allows purification by glutathione affinity chromatography and is known to help the solubility of recombinant proteins.

The HAZV N fragment was subcloned by direct digestion from the pET 28-SUMO vector and ligation into the pGEX6P2 vector using BamHI and XhoI restriction enzymes. Following double digestion of both the pET 28-SUMO vector and empty pGEX6P2 vector, the HAZV N ORF fragment was isolated by agarose gel electrophoresis and gel extracted, then ligated into the digested, empty pGEX6P2 vector and transformed into *JM109* cells. The correct orientation and sequence of the insert was confirmed by DNA sequence analysis. This construct will be referred to as pGEX-HAZV N and resulted in the expression of an N terminally GST tagged HAZV N (GST-HAZV N). Following cleavage of the GST tag with HRV 3C protease, five non-native residues (Gly-Pro-Leu-Gly-Ser) remained at the N-terminus of HAZV N. A schematic representation of the GST-HAZV N fusion protein is shown in Figure 3.1 (A) and a schematic representation of the pGEX-HAZV N vector used to express it in Figure 3.1 (B).

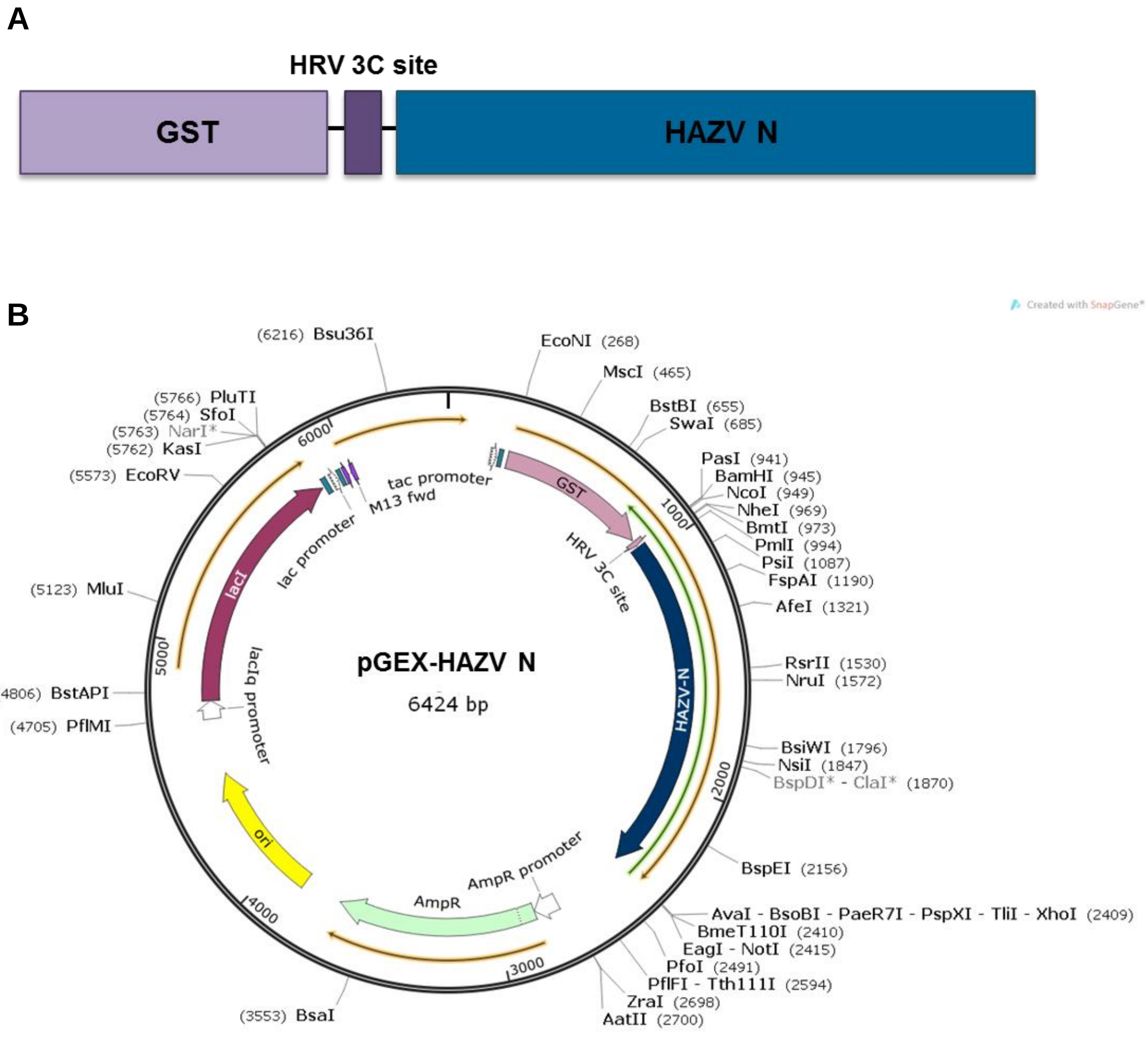


Figure 3.1 Schematic representation of GST-HAZV N, and plasmid map of the HAZV N ORF in the pGEX6P2 vector.

HAZV N was fused at the N terminus to GST, which could be cleaved from HAZV N at the HRV 3C site using HRV 3C protease (A). The plasmid map (B) was generated using SnapGene software.

3.2.2 GST-HAZV N expression and purification

3.2.2.1 GST-HAZV N expression

Initial expression of GST-HAZV N from pGEX-HAZV N was tested using the optimal conditions achieved for CCHFV N (Carter, Barr et al. 2012) as a starting point. His-SUMO-CCHFV N was expressed as follows: overnight induction at 18 °C in Rosetta 2 cells with 100 µM IPTG; cell recovery by centrifugation and resuspension in lysis buffer, sonication of lysates, then clarification of lysates by centrifugation. When using these conditions to express GST-HAZV N from pGEX-HAZV N, SDS-PAGE analysis showed the GST-HAZV N fusion protein was well expressed, and was soluble, although there was also a large amount of insoluble GST-HAZV N. This was addressed by increasing the percentage of Triton-X 100 in the lysis buffer from 0.1 % to 1 % (v/v), and including a 30 minute incubation stage on ice prior to sonication. The predicted molecular weight of GST-HAZV N is 80 kDa; when analysed by SDS PAGE GST-HAZV N actually has an apparent molecular weight of slightly lower than 80 kDa, however this may just be an artefact of electrophoresis (Figure 3.2).

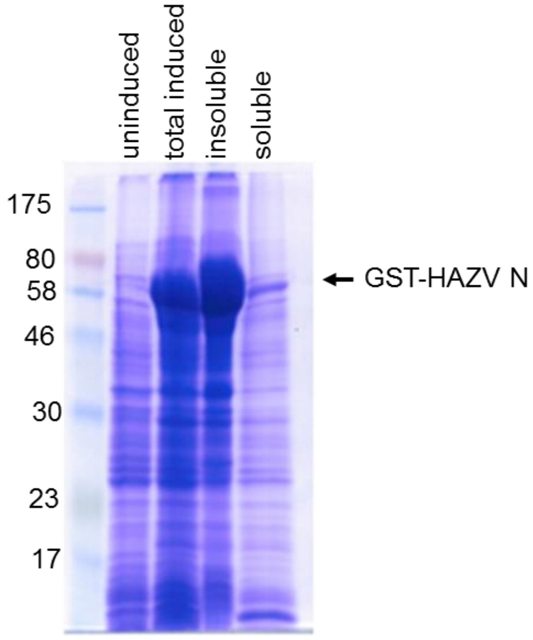


Figure 3.2 Initial expression of GST-HAZV N

Expression of GST-HAZV N from pGEX-HAZV N in Rosetta-2 cells overnight at 18 °C with 100 µM IPTG, followed by pellet resuspension in lysis buffer, sonication of lysates, and clarification of lysates by centrifugation.

Following expression of GST-HAZV N, glutathione sepharose affinity chromatography was used to purify GST-HAZV N from bacterial cell lysates. Clarified bacterial cell lysate was combined with GST resin in batch, and incubated with rotation at room temperature for 1 hr to allow GST-HAZV N to bind. A considerable amount of GST-HAZV N remained unbound after 1 hr, therefore overnight binding of GST-HAZV N to GST resin at 4 °C was also tested. However this resulted in more protein precipitating out of solution without any apparent increase in HAZV-N yield, therefore in all subsequent experiments clarified lysates were incubated with GST resin for 1 hr at room temperature. GST resin with GST-HAZV N bound was washed once in binding buffer, then to dissociate *E. coli* derived RNA from GST-HAZV N, resin was washed once in buffer containing 1.5 M NaCl. This resulted in the purification of 'RNA free' HAZV N that co-eluted with less bacterially derived RNA than RNA bound-HAZV N; the 1.5 M NaCl wash step was excluded when RNA-bound HAZV N was purified. Resin with GST-HAZV N bound was then equilibrated in another binding buffer wash, prior to three successive elutions in buffer containing reduced glutathione to dissociate GST-HAZV N from the resin (Figure 3.3). Eluted GST-HAZV N was incubated with HRV 3C protease to cleave the GST tag from HAZV N; this resulted in the production of a strong band of ~54 kDa (the correct predicted molecular weight for HAZV N) when visualised by SDS PAGE and Coomassie staining (Figure 3.3 (B)).

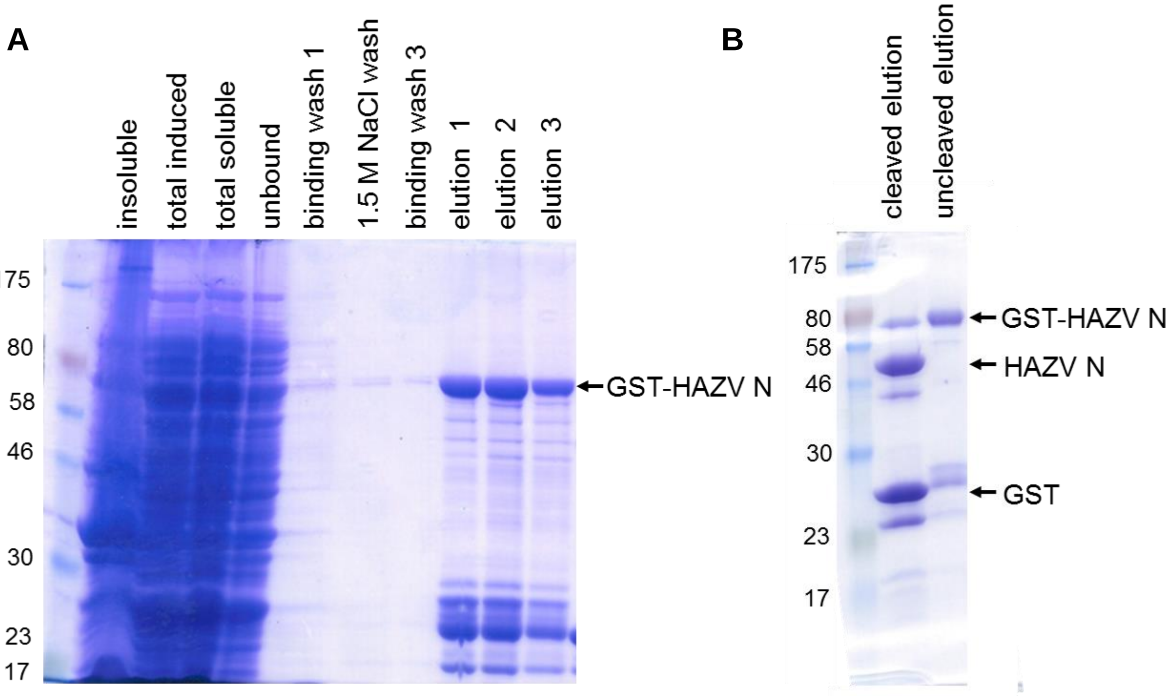


Figure 3.3 Expression of GST-HAZV N and removal of the GST tag

Purification and HRV 3C cleavage of GST-HAZV N resulted in the release of soluble HAZV N with an apparent molecular weight (MW) of 54 kDa. (A) SDS PAGE gel showing the purification of GST-HAZV N by glutathione sepharose affinity chromatography. (B) SDS PAGE gel showing cleavage of GST-HAZV N with HRV 3C protease to release HAZV N.

3.2.2.2 Size exclusion chromatography

After removal of the GST tag with HRV 3C protease, further purification of HAZV N was needed to isolate fully cleaved HAZV N from uncleaved GST-HAZV N and GST. In order to achieve this, and to analyse the oligomeric status of HAZV N, size exclusion chromatography was carried out (section 2.2.2.3). Size exclusion chromatography can separate protein mixtures on the basis of protein size, due to the differential exclusion of larger proteins from pores in the chromatography beads that make up the column. Larger proteins and multimers are eluted first, as they cannot enter the pores in the beads and pass straight through the column; smaller proteins are eluted afterwards, as they interact with the pores in the beads and take a longer time to pass through the column. Given the difference in predicted molecular weight of

GST-HAZV N (80 kDa), HAZV N (54 kDa), and GST (26 kDa), these protein species should be seperated by size exclusion chromatography.

The size exclusion column was equilibrated in gel filtration buffer overnight, then the cleaved GST-HAZV N protein mixture was concentrated to a 5 mL volume and injected onto the column with a 5 mL loop. The column was run at a flow rate of 2 mL/min until the void volume (V_0, ~90 mL) had eluted; fractions were collected from this point onwards, until a full column volume of buffer had eluted (320 mL). The A_{280} and A_{260} traces are shown in Figure 3.4 (A), and SDS PAGE gel analysis of the corresponding fractions are shown in Figure 3.4 (B). Cleaved HAZV N seemed to be eluted in all fractions analysed between 105 mL and 160 mL. Peak 3 consisted mostly of nucleic acid and was not analysed by SDS PAGE. Peak 1 contained a mixture of multimerised HAZV N and uncleaved GST-HAZV N. Complete cleavage of GST-HAZV N could not be achieved despite increasing the ratio of HRV 3C protease:elution volume, possibly due to aggregation or misfolding of a small proportion of GST-HAZV N. Consistent with the observations of (Carter, Barr et al. 2012) and (Wang, Dutta et al. 2012) for CCHFV N, the A_{260}/A_{280} ratio appeared to be greater in the multimeric HAZV N peak (peak 1), implying that this species co-eluted with more nucleic acid than the monomeric species in Peak 2.

Size exclusion chromatography confirmed that the majority of HAZV N was monomeric in solution, although higher ordered structures did form that contained a greater proportion of nucleic acid, in a similar fashion to purified CCHFV N. Unfortunately GST in solution (26 kDa) multimerises to form a dimer of 52 kDa, which is very similar to the molecular weight of HAZV N (54 kDa). Protein species so close in molecular weight cannot be resolved by size exclusion chromatography, therefore an alternative method of HAZV N purification was attempted. This is described in the next section, where ion exchange chromatography was utilised in order to separate HAZV N from uncleaved GST-HAZV N and GST.

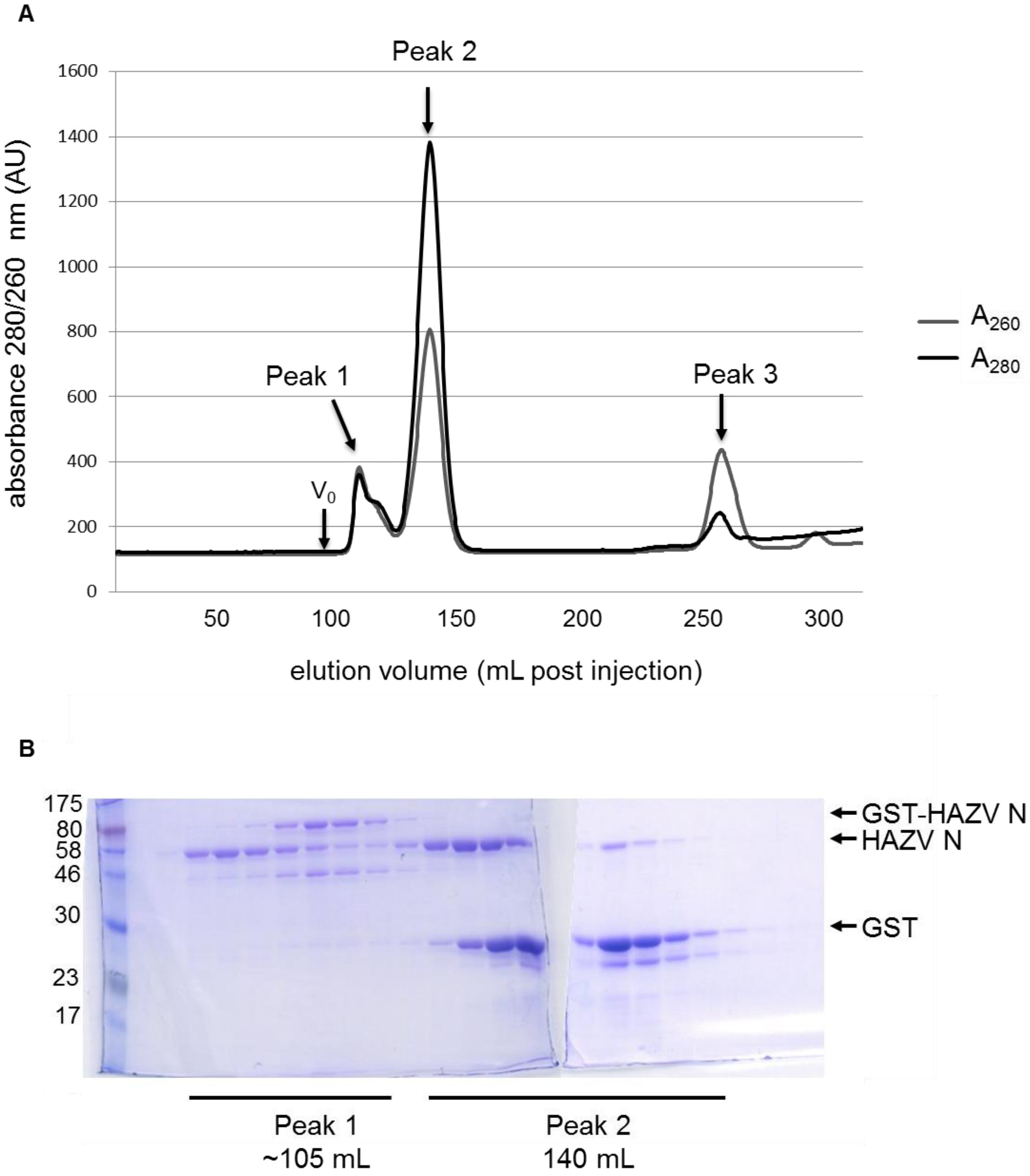


Figure 3.4 Size exclusion chromatography of cleaved HAZV N.

(A) The A_{280}/A_{260} absorbance trace indicated protein eluted mainly in 2 peaks; most protein eluted in peak 2 at 140 mL, which represented proteins of ~54 kDa. (B) SDS PAGE analysis of fractions from size exclusion chromatography, with the corresponding peaks labelled. HAZV N, 54 kDa, appeared in all fractions analysed, although the majority of monomeric HAZV N co-eluted with dimeric GST, therefore size exclusion chromatography could not be used to purify HAZV N.

3.2.2.3 Ion exchange chromatography

HAZV N and GST have a significantly different isoelectric point (pI), of 8.73 and ~6 respectively, which raises the possibility that they could be resolved using ion exchange chromatography. A cation exchange column was used (Resource S column) with all buffers at pH 7 so that HAZV N would be positively charged and bind to the column, whilst GST (and other contaminants) would be negatively charged and pass straight through the column without binding. After elution from the GST resin and cleavage with HRV 3C protease, HAZV N was dialysed into the initial ion exchange buffer, buffer A, which contained 50 mM NaCl. The low NaCl concentration in buffer A allowed the positively charged HAZV N to bind to the negatively charged cation exchange column through ionic interactions. After GST and other contaminants flowed through, the cation exchange column was washed several times with buffer A. HAZV N was then eluted by increasing the NaCl concentration in a linear gradient until a sufficiantly high NaCl concentration was reached to disrupt the ionic interactions between HAZV N and the column. A linear NaCl gradient was generated by gradually mixing buffer A (contained 50 mM NaCl) with buffer B (contained 1 M NaCl) using an Äkta purifier pump. HAZV N eluted relatively rapidly after the NaCl concentration started increasing, which indicated the ionic interaction between HAZV N and the Resource S column was relatively weak (Figure 3.5 (A)).

SDS PAGE analysis (Figure 3.5 (B)) of fractions collected during ion exchange chromatography confirmed that the majority of contaminants were removed during this procedure, including residual GST-HAZV N and GST, as they flowed through the cation exchange column without binding (Figure 3.5 (A) Peak 1). In contrast HAZV N bound to the column, and was eluted when the NaCl concentration was increased. This resulted in the isolation of >95 % pure cleaved HAZV N, Peak 2 in Figure 3.5.

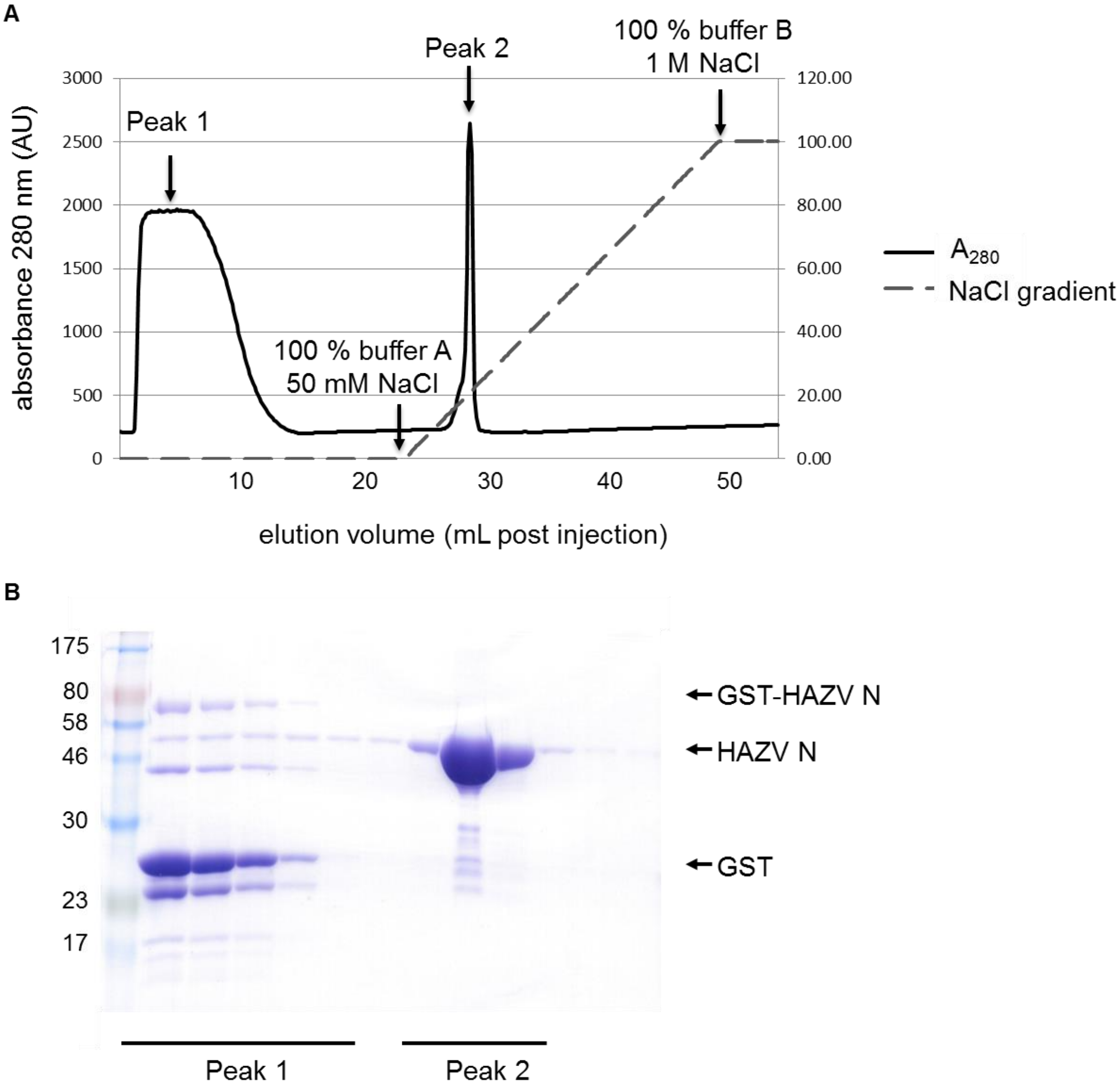


Figure 3.5 Purification of HAZV N using ion exchange chromatography

Ion exchange chromatography resulted in the isolation of HAZV N from GST-HAZV N and GST. (A) A_{280} absorbance trace shows protein eluted in 2 peaks; Peak 1 represents protein that flowed through the column without binding, and Peak 2 represents protein that bound to the column, and was then eluted by an increasing NaCl gradient (dashed grey line). (B) SDS PAGE analysis of fractions from ion exchange chromatography with the corresponding peaks labelled. The majority of HAZV N eluted in Peak 2, whereas GST-HAZV N and GST flowed through the column without binding, Peak 1.

Expression and purification of HAZV N by ion exchange chromatography produced 4-7 mg > 95 % pure HAZV N per 4 L bacterial culture. Attempts at further optimisation were unsuccessful, and this was sufficient to proceed with crystallisation trials, therefore no further optimisation of expression or affinity purification was undertaken. Preparations of HAZV N that were washed with 1.5 M NaCl whilst bound to the GST resin had an OD_{260}/OD_{280} ratio of 0.84, indicating that a low level of RNA was present; the crystal structure derived from these preparations is referred to as Apo-HAZV N. HAZV N protein preparations that did not have this 1.5 M NaCl wash step included resulted in purification of HAZV N with an OD_{260}/OD_{280} ratio of 1.04, indicating more nucleic acid (presumably RNA) co-eluted with this protein. The crystal structure that resulted from this preparation is referred to as RNA bound-HAZV N.

3.2.3 Section summary

This section presents the optimisation of HAZV N expression and purification from bacterial cells, ultimately resulting in the purification of 4-7 mg HAZV N per 4 L bacterial culture that was >95 % pure. The HAZV N ORF was subcloned and expressed as a GST fusion protein from pGEX-HAZV N, as GST is a well documented aid for recombinant protein solubility. GST proteins are enzymes that bind reversibly and with high specificity to their their substrate glutathione, a tripeptide of glutamate-cysteine-glycine. The ability of GST to bind reversibly to glutathione with high specificity and high affinity makes GST ideal for affinity chromatography, and often yields high purity protein.

The 26 kDa GST tag encoded by pGEX vectors is derived from *Schistosoma japonicum* (Smith and Johnson 1988), and forms homodimers in solution. This is one of the disadvantages of using a GST tag for purification of 50-60 kDa recombinant proteins; GST cannot be removed from the recombinant protein mixture by size exclusion chromatography, after HRV 3C cleavage of the fusion protein. As shown in section 3.2.2.2, HAZV N co-eluted with GST homodimers during size exclusion chromatography, resulting in no further purification of HAZV N. To circumvent this problem, HAZV N was seperated from uncleaved GST-HAZV N and GST by cation exchange chromatography, which yielded >95 % pure HAZV N. A schematic summarising the final optimised purification procedure of HAZV N is shown in Figure 3.6.

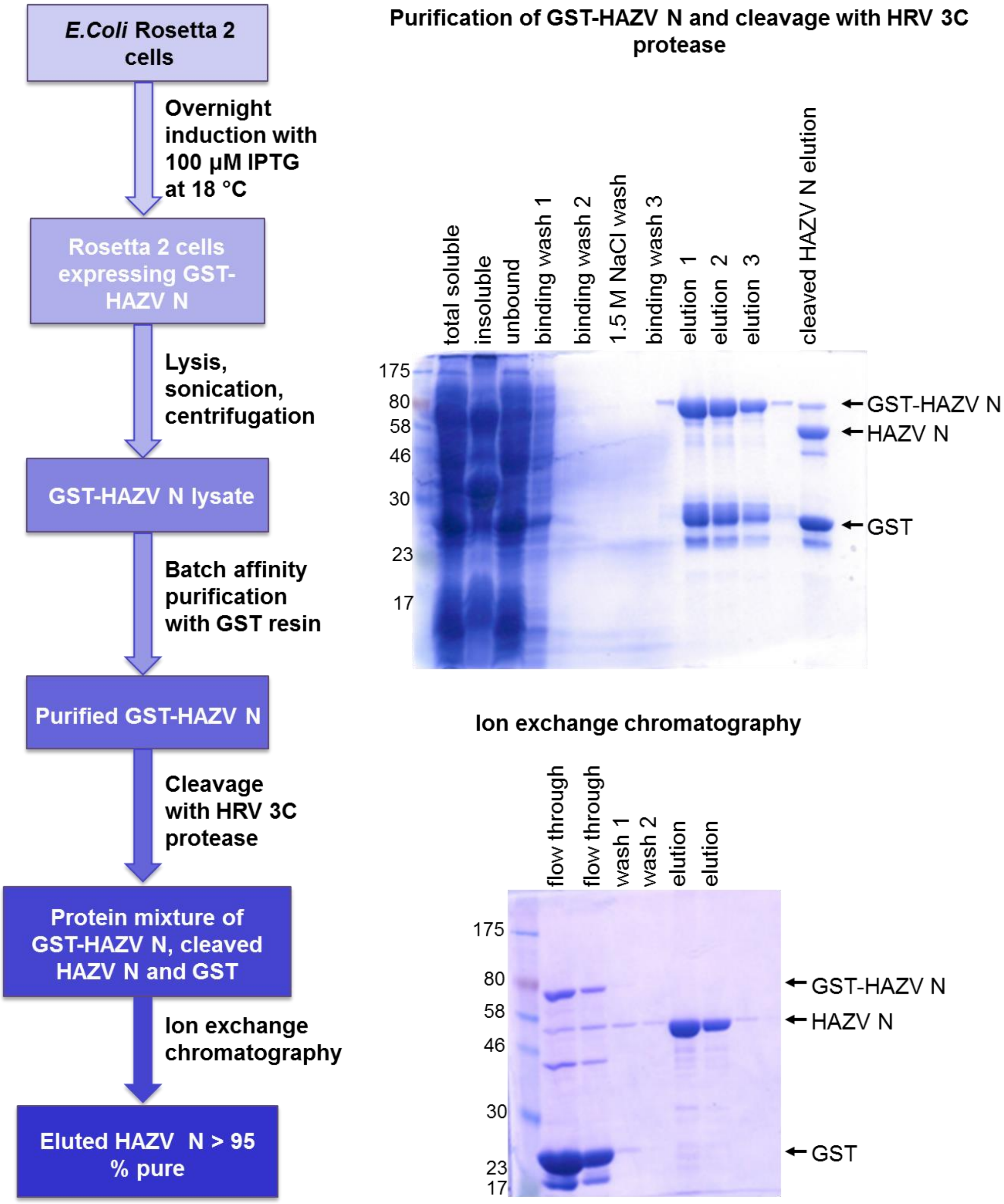


Figure 3.6 Schematic summary of final optimised HAZV N expression and purification.

GST-HAZV N was expressed in Rosetta 2 cells, purified by glutathione affinity chromatography, then cleaved with HRV 3C protease to release HAZV N. HAZV N was further purified by ion exchange chromatography.

Purified HAZV N was used in two downstream applications; the primary use of HAZV N was for crystallisation trials and structure solution; in addition HAZV N was used to generate polyclonal antibodies that bind HAZV N. The strategy used to generate these antibodies is described in section 3.2.4, and downstream uses of these antibodies are described in chapter 4 and chapter 5.

3.2.4 Generation of polyclonal antibodies that bind HAZV N

Purified HAZV N was used to generate polyclonal antibodies that recognised native HAZV N; briefly 1 mg HAZV N protein was sent to Altabioscience, a company that specialises in custom antibody production. Altabioscience performed an immunisation programme (which is UK home office and USA National Institutes of Health compliant) that involved inoculation of a sheep with purified HAZV N, followed by successive bleeds at one, two, and three months post inoculation. SDS PAGE analysis of HAZV N sent to Altabiosciences confirmed it was >95 % pure (Figure 3.7), therefore the serum harvested from these bleeds should contain sheep antibodies that specifically recognise HAZV N. Applied uses of anti-HAZV N antibodies are described in chapter 4 and chapter 5.

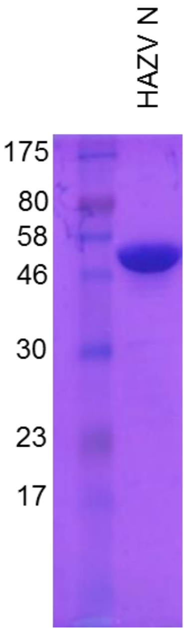


Figure 3.7 Purified HAZV N used to generate polyclonal anti-HAZV N antibodies

SDS PAGE analysis confirmed that HAZV N sent to Altabioscience was >95 % pure.

3.3 RESULTS: SOLUTION OF THE HAZV N CRYSTAL STRUCTURE

3.3.1 Crystal growth

3.3.1.1 Introduction to crystal growth

X-ray crystallography requires an abundant source of homogeneous, >95 % pure, soluble, concentrated protein. Once this has been achieved, using a methodology similar to that described for HAZV N purification in the previous section, the next challenge of crystallography is crystal growth. One of the previously mentioned difficulties associated with crystallography is the necessity for well ordered, high quality, well-diffracting protein crystals. Highly automated systems, such as the use of robots for setting up crystal screens, making buffers required for crystallisation, and monitoring crystal growth, go someway towards standardising experiments, such that crystallisation conditions can be reproduced more accurately. However, even though high throughput systems are becoming more common, crystal growth still remains one of the more challenging aspects of X-ray crystallography.

It is necessary to grow crystals because the intensity of diffracted X-rays from a single protein molecule are too weak to be detected by even the most sophisticated equipment. Instead, crystals which contain many copies of a protein, arranged in a periodic, well-ordered lattice, amplify the signal of the diffracted X-rays and enable their detection. Protein crystals are grown in the following manner; soluble pure protein is mixed with a solution containing a precipitant, and is then concentrated to supersaturation, beyond the concentration at which the protein is soluble (the solubility curve), such that the phase of the solution moves into the supersaturated labile zone (nucleation zone). At this point spontaneous nucleation (ordered aggregation) of protein crystals can (sometimes) occur. The concentration of protein and precipitants in solution, is not however, allowed to reach the point of precipitation (disordered aggregation). A solubility phase diagram is shown in Figure 3.8, which depicts this relationship between precipitant and protein concentrations.

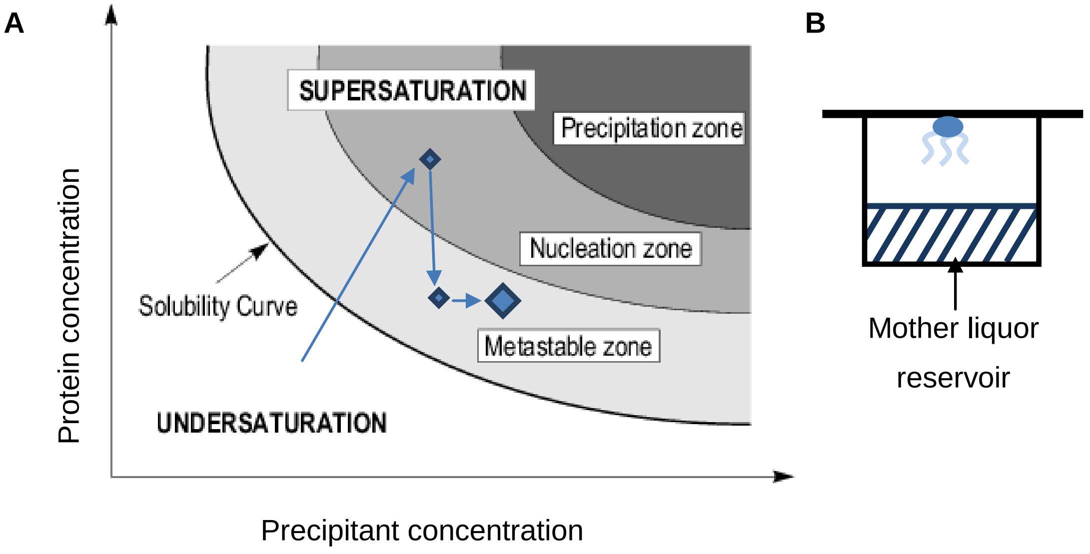


Figure 3.8 A solubility phase diagram.

Adapted from: http://www.xray.bioc.cam.ac.uk/xray_resources/whitepapers/xtal-in-action/node3.html. (A) The phase moves from undersaturated to supersaturated as the concentration of the precipitant and protein increases; once this concentration is sufficient to enter the supersaturated labile zone (nucleation zone), spontaneous nucleation of protein crytals occurs, thus lowering the concentration of protein in solution. At this point the metastable zone is entered, which supports sustained crystal growth. In the precipitation zone disordered aggregation of protein occurs, and no crystals grow. The blue arrows represent protein in solution. (B) Crystal growth by vapour diffusion in hanging drop format.

Proteins and precipitants are concentrated to supersaturation by suspending the protein:precipitant solution over a reservoire of more concentrated precipitant solution (the crystallant or ‘mother liquor’ solution) in a closed system (Figure 3.8 (B)). As water evaporates from the protein:precipitant solution, and equilibrium is approached between the concentration of precipitants in the protein:precipitant solution and the mother liquor reservoire, the concentration of both precipitant and protein increases, supersaturation is achieved and (sometimes) spontaneous nucleation (ordered aggregation) occurs. This decreases the concentration of protein in the protein:precipitant solution, allowing the phase to enter the metastable zone, which supports sustained crystal growth either until impurities are incorporated into the lattice at the surface of the growing crystal, preventing further ordered crystal growth, or until the protein is depleted. This method of growing crystals is known as vapour diffusion; droplets of protein:precipitant solutions are either placed on platforms over a mother liquor reservoire (‘sitting drop’), or suspended underneath coverslips over a mother liquor reservoire (‘hanging drop’) (Figure 3.8 (B)). In both techniques the coverslips or platforms are sealed, to create a closed system. On average it takes between 3-7 days for sufficiently large protein crystals to grow, although they can grow in a matter of hours, or can take weeks or months.

3.3.1.2 HAZV N crystallisation screens and crystal growth

Commercially available sparse matrix screens were used to screen purified HAZV N protein for initial crystallisation conditions. The mother liquor solutions in the sparse matrix screens generally contained a combination of precipitants (organic polyalcohols e.g polyethylene glycol (PEG)), additives (e.g metal ions, detergents, reducing agents), salts (e.g sodium chloride or ammonium sulphate) and buffers of different pHs. The mother liquor compositions in the sparse matrix screens were biased towards conditions that have previously been successful at supporting protein crystal growth. Each screen was supplied in a high throughput 96 well format, which enabled an automated robot system to be used that quickly and reproducibly set up sitting drops with protein:mother liquor ratios of 50:50 and 70:30. The use of multiple sparse matrix screens, each incubated at either 25 °C, 18 °C, or 11 °C allowed hundreds of potential HAZV N crystal growth conditions to be assayed. Several different conditions resulted in the growth of crystals overnight, which increased in size over 2-3 days. The appearance of crystals in several of the initial sparse matrix screening wells enabled the optimisation of these conditions, without further extensive screening, or attempting to proceed with sub-optimal conditions (for example, conditions in which microcrystals grew due to excessive nucleation, or conditions containing crystalline precipitate).

The two conditions in which the best crystals grew (based on the visual apprearance of the crystals e.g sufficient size (>20 µM), single crystals, few splits, sharp edges) were: Wizard 4 A6; 28% (v/v) PEG 400, 0.1 M HEPES pH 7.5, 0.2 M Calcium chloride and Wizard 3 F5; 15 % (w/v) PEG 10,000, 100 mM sodium citrate pH 5.5, 2% (v/v) dioxane. Crystals from the Wizard 3 F5 (WF5) looked most promising visually, therefore this condition was taken forward for further optimisation. Optimisation was performed using 24-well matrices that were set up manually in hanging drop format. The percentage of PEG 10,000 and pH were varied, as was the temperature at which the crystals grew (18 °C, 25 °C and 30 °C). Different types of PEG (PEG 6000, PEG 8000, PEG 20,000), and different salts (sodium chloride, ammonium acetate, lithium sulphate and magnesium chloride) were added to improve crystal growth, as well as using different droplet sizes and different protein: mother liquor ratios. Hanging droplets of 1 µl volume, 70:30 protein:mother liquor ratio generally yielded high quality crystals. Figure 3.9 shows an example of small crystals grown in the protein:mother liquor droplets during optimisation.

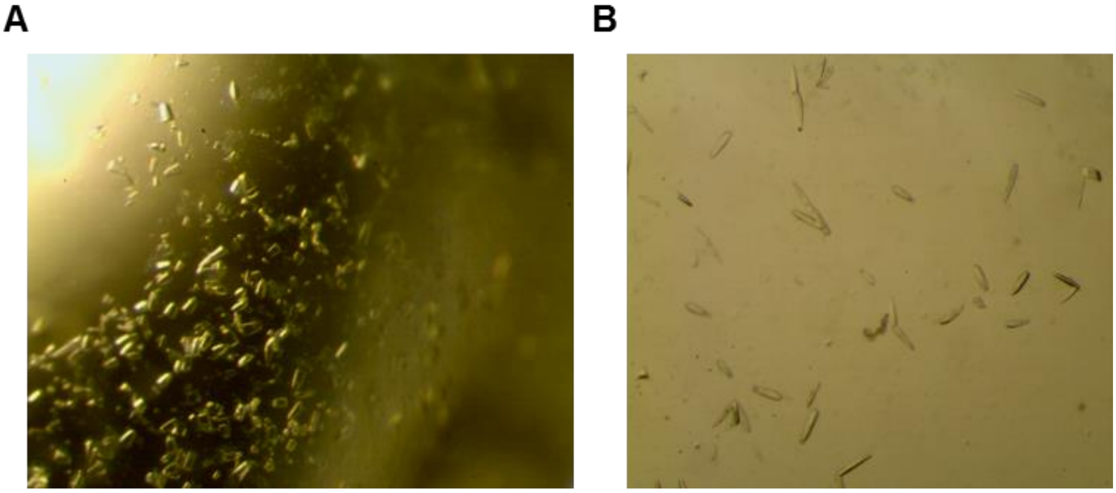


Figure 3.9 Crystals grown in protein:mother liquor droplets.

Although the crystals grown in the protein:mother liquor droplets in (A) and (B) were relatively small, however the crystals in (A) especially, are uniform and substantial, and grow as single crystals. In (B) some of the crystals start growing as multiple crystals from a single point, which made them less likely to diffract well and made it harder to analyse a single crystal.

Further crystal growth optimisation also included the use of a commercially available additives screen (Hampton research). This screen contained small molecules (additives) known to affect crystallisation by disrupting the structure that water molecules form around the target protein or by altering protein solubility. The Oryx 6 robot was used to set up this screen in 96 well sitting drop format at 25 °C, centred around the best diffracting condition; 15% PEG 10,000, 100 mM sodium citrate pH 5, 0.2 M sodium chloride, 2% dioxane. Additives were added to a 110% stock solution of the mother liquor, to a final concentration of 10%. Various conditions resulted in crystal growth, and a few conditions yielded crystals that were visually improved (e.g larger or more cuboidal), however none of the conditions yielded crystals that diffracted to a higher resolution, or that generated diffraction data of any higher quality. Streak seeding was also used to try and improve crystal size and quality, but did not result in any improvement of diffraction resolution or data quality, and so no further optimisation of this condition was undertaken.

Batch to batch variability between different proteins preparations presented significant problems during crystal growth optimisaton. Multiple protein preparations were needed to optimise crystal growth because relatively low yields of HAZV N were achieved during optimisation of HAZV N expression and purification. Different protein preparations varied slightly in terms of final protein concentration, final pH, and in their levels of contaminants. These small, unavoidable differences, significantly impacted both the ability of HAZV N to form crystals and the quality of the crystals that grew. HAZV N was frozen at -80 °C, either in the presence or absence of 5% glycerol, then thawed 2 days later on ice and its ability to form crystals was assayed in known crystal growth conditions. Although freeze-thawed HAZV N could still crystallise, batch to batch variation remained a problem.

Once single crystals that were big enough (>20 µM), and good enough quality grew (no splitting or fracturing), they were manipulated using small loops, in which the crystal was suspended in its mother liquor solution, an example of which is shown in Figure 3.10.

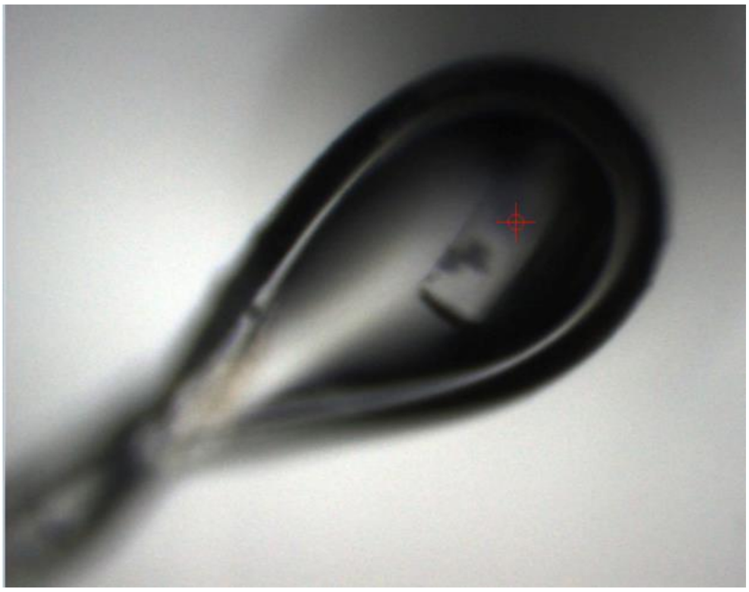


Figure 3.10 Loop containing a cuboidal HAZV N crystal at the synchrotron

This crystal was cryo-cooled in liquid nitrogen, and maintained at -173 °C.

Large (>20 µM), uniform crystals were picked from protein: mother liquor droplets using these small loops, then crystals were cryo-cooled in liquid nitrogen to reduce the radiation damage caused by X-rays. Cryo-cooling can be done directly after picking crystals from the mother liquor droplets, if the mother liquor contains sufficient cryo-protectant (low molecular weight PEG, glycerol, ethelyne glycol, DMSO). As HAZV N did not crystallise in any conditions that contained sufficient cryo-protectant, crystals were instead soaked in cryo-protectant solutions after being picked from mother liquor droplets. Cryo-protectants are necessary to avoid water freezing as ice crystals during cryo-cooling, as ice crystals can disrupt the protein crystal lattice and cause high background diffraction. Cryo-protectants enable water in and around the crystal to instead cool into vitreous ice (an amorphous glass), thus preventing ice crystal formation.

Various different cryoprotectants were used to soak HAZV N crystals during optimisation of HAZV N crystallisation and diffraction, including PEG 200, PEG 400, glycerol, DMSO, ethylene glycol and MPD (2-Methyl-2 4-pentanediol). PEG 200 was determined to be the best cryoprotectant, and was used in all subsequent experiments. Crystals grown in the following condition (15% PEG 10,000, 0.1 M sodium citrate pH 5, 0.1 M NaCl, 2 % dioxane) provided the diffraction data for both the Apo-HAZV N, and the RNA bound-HAZV N crystal structures.

3.3.2 Crystal diffraction and data collection

A HAZV N protein crystal was placed in the path of a beam of monochromatic (single wavelength) X-rays in order to collect X-ray diffraction data. The X-ray source was a synchrotron (a high brilliance, tuneable source of X-rays) located at the Diamond light source, the UK national synchrotron science facility. Data was collected using X-rays because the wavelength of X-rays (1 Å = 0.1 nm) is in the same order of magnitude as the interatomic distances that need to be resolved during a crystallography experiment. When the crystal is placed in the path of a beam of X-rays, the majority of X-rays pass straight through the crystal, however occasionally a photon from the X-ray interacts with the electron density around atoms, and is scattered from the crystal. The scattered X-ray waves emerge from the protein crystal with different phases and scattering angles; depending on the way in which the emerging waves combine with each other, either constructive interference (results in waves with higher amplitudes) or destructive interference (results in waves with lower amplitudes, or no signal) occurs. When these waves hit a detector placed in front of the crystal, a pattern of discrete intensities is recorded; these are known as reflections. The diffraction pattern is formed from these reflections, and is the raw data of an X-ray diffraction (crystallography) experiment.

Bragg's law defines the circumstances that determines when a particular set of emerging waves will combine constructively to generate a reflection. Each unit cell (defined in section 3.3.3) within the crystal lattice can be intersected by a set of parrellel planes (Bragg's planes) at a fixed point along their axis; two waves will only remain in phase after reflecting from these planes if these planes are a distance (d) apart, with an angle of incidence (θ) for two incoming X-ray waves of wavelength λ, and if the additional distance the second wave must travel (d sinθ) is a multiple (n) of the wavelength:

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

This is depicted in Figure 3.11.

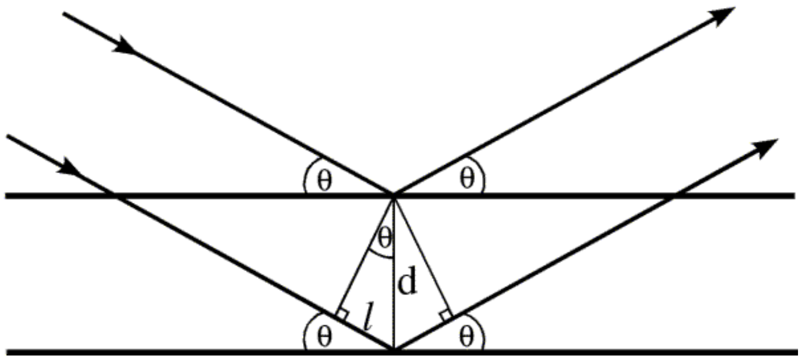


Figure 3.11. Bragg's law

This defines the lattice spacing (d) and diffraction angle (θ) the required for scattered X-rays to remain in phase with each other. These two waves combine constructively, which increases their amplitude, and results in a reflection on the detector. Taken from http://www-structmed.cimr.cam.ac.uk/Course/Basic_diffraction/Bragg_planes.gif.

The resultant wave that generates a reflection on the detector has contributions from each diffracting electron in every unit cell throughout the entire crystal. This is due to the fact that Bragg's planes intersect each unit cell in the same way, therefore each unit cell contributes equivalent constructive interference to the emerging wave, and amplify the emerging signal such that it can be detected. The way in which Bragg's planes intersect each unit cell are described by Miller indices (*hkl*), which are assigned to each plane that contributes to the diffraction pattern. Each reflection in the diffraction pattern can therefore be assigned corresponding Miller indices during indexing of the diffraction data (section 3.3.3). All the contributing waves across the whole crystal that result in a single reflection can be added together in the structure factor equation, which results in a structure factor F_{hkl} being assigned to each individual reflection. The structure factor amplitude $|F_{hkl}|$ is proportional to the square root of the reflection intensity (the reflection intensities and their positions are what is measured during data collection). Each diffraction pattern only provides information about the crystal from a single direction, therefore the crystal must be rotated to collect as much information as possible and to generate a complete dataset. A different set of resultant waves will be detected at each angle as the crystal is rotated, as a different set of planes will satisfy Bragg's law at each angle.

However, many scattering electrons do not lie on Bragg's planes in the unit cell (they lie between the planes at fractional values of *d*). X-ray waves scattered by these electrons are out of phase with the resultant waves, and will result in a degree of destructive interference with the X-ray waves that generate reflections, decreasing the amplitude of these waves, and resulting in a relative decrease in the intensity of the reflections. Information about the arrangement of atoms in the unit cell relative to Bragg's planes is therefore provided by the relative intensities of the reflections. Information about the arrangement of the asymetric units (AU) (the smallest repeating unit of the crystal lattice), and how they are related to each other (i.e the space group), is similarly provided in the symmetry of the reflections in the diffraction patterns.

An example of a diffraction pattern collected from crystals of Apo-HAZV N is shown in Figure 3.12. The intensity of reflections in diffraction patterns and their symmetry together provides almost all the information needed to generate an electron density map from the crystal – with the exception of one vital piece of information; the phase of the resultant waves that generate the reflections. In order to generate an electron density map from the structure factors, a Fourier transform (a way of describing a wave as a sum of all its component waves) is required. It is necessary to know the relative phase of these waves to make sure they are added correctly. The lack of information about the phases of the resultant waves, due to the way diffraction data is currently recorded (photon counts on a detector) leads to the fundamental 'phase problem' that has to be solved in all crystallography (X-ray diffraction) experiments. During analysis of diffraction data from HAZV N crystals, the phase problem was addressed using molecular replacement (outlined in section 3.3.4).

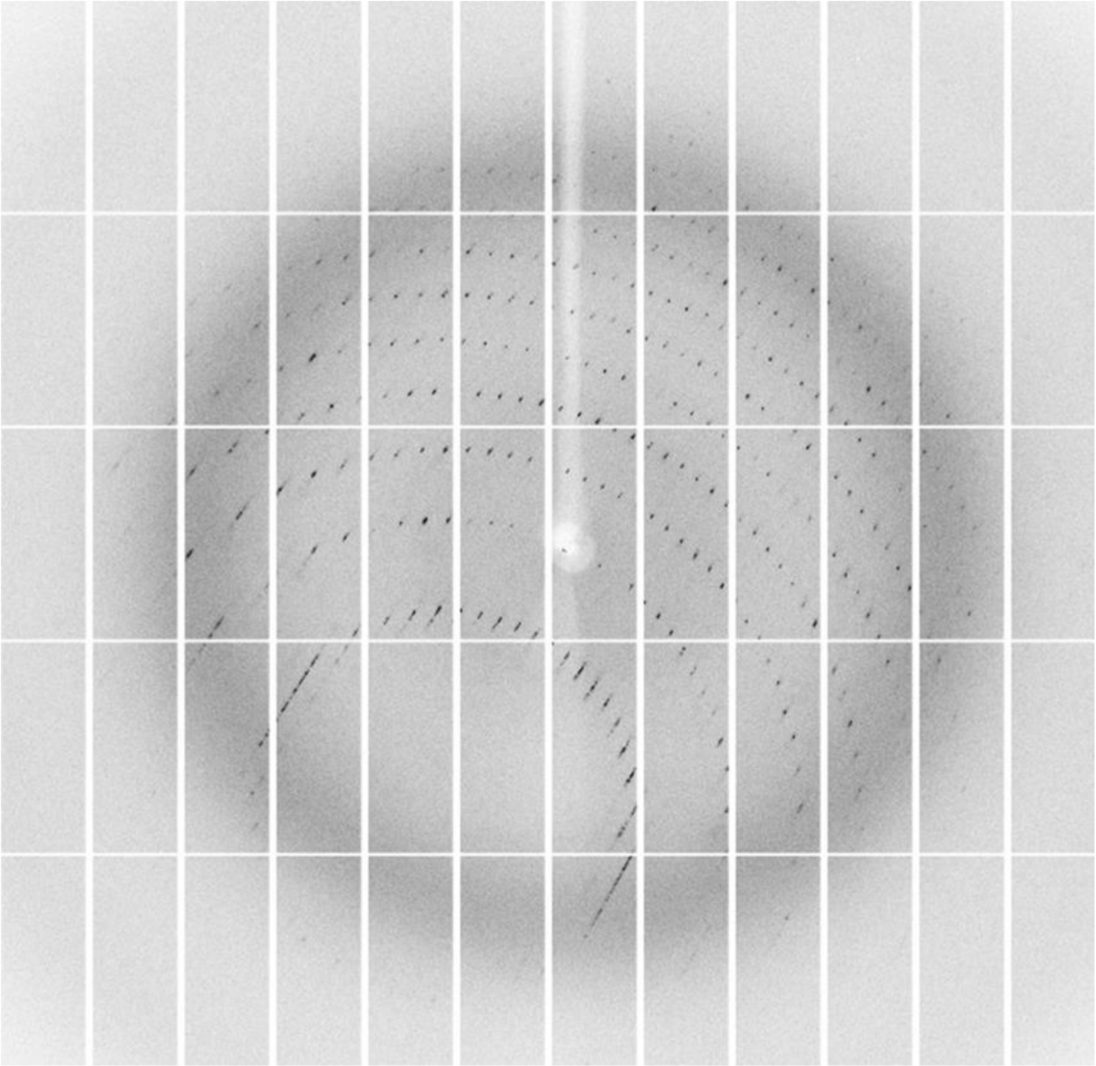


Figure 3.12. Diffraction patterns of Apo-HAZV N .

The symmetry of the reflections together with their intensities provide all the information needed to determine the X-ray structure, apart from the phase of the resultant waves that generate the reflections.

3.3.3 Data processing

The crystal used generate the Apo-HAZV N dataset diffracted to a resolution of 2.7 Å in the $P2_12_12_1$ space group, and the crystal used to generate the RNA bound-HAZV N dataset diffracted to a resolution of 3 Å in the $P2_1$ space group. The space group of the crystal describes how the asymmetric units (AU) are arranged in the unit cell, and it can be determined by the symmetry of reflections in the diffraction pattern. The AU contains the smallest object that can be repeated to make the entire crystal (Figure 3.13 (B)). The unit cell is the identical repeating unit of the crystal that can be stacked together to create the periodically arranged, weakly interacting 3-dimensional lattice of protein and solvent molecules, that is the protein crystal. Unit cells are described by vectors a, b and c, along axes x, y and z and angles α, β and γ, an example of which is shown in Figure 3.13 (A). There are a finite number of space groups (65 for proteins) as there is a finite number of ways in which (protein) molecules can be arranged in space.

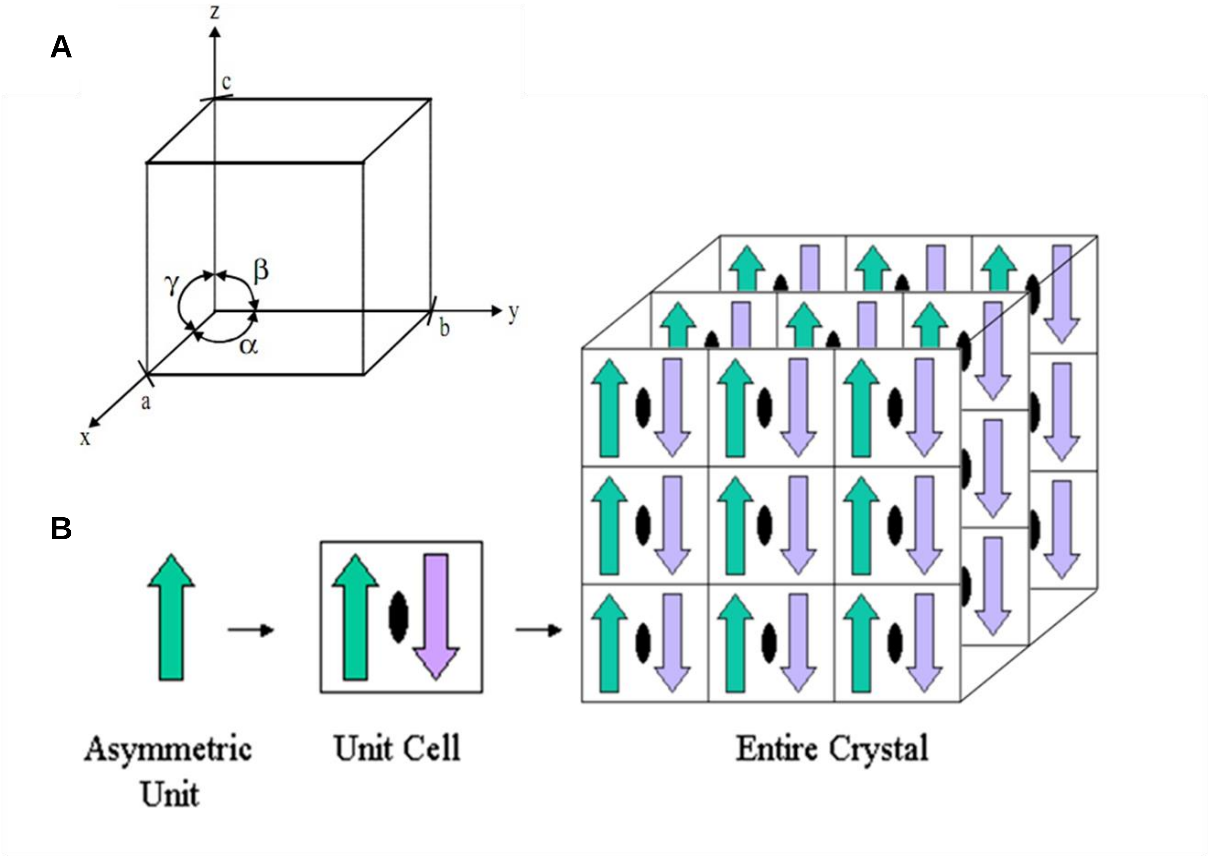


Figure 3.13. The unit cell.

The unit cell is the smallest identical repeating unit of the crystal that can be stacked together to create the crystal (B). The unit cell is annotated by vectors a, b, and c along axes x, y and z, and angles α, β and γ (A). The AU contains the smallest object that can be repeated to make the entire crystal (B). Images are taken from: http://www.rcsb.org/pdb/101/static101.do?p=education_discussion/Looking-at-Structures/bioassembly_tutorial.html and http://www.tutorsglobe.com/homework-help/physical-chemistry/unit-cell-72605.aspx .

The dimensions of the Apo-HAZV N unit cell were a=64.99 Å, b=76.1 Å, c=449.28 Å, α=90 °, β=90 °, and λ=90 °, and the RNA-bound HAZV N unit cell were; a=64.7 Å, b=84.0 Å, c=104.5 Å, α=90 °, β=105.74 °, and λ=90 °.

After data was collected for Apo-HAZV N and RNA-bound HAZV N crystals, data processing was carried out by Dr Antonio Ariza and is briefly described in the following section. X-ray data for both Apo-, and RNA bound-HAZV N crystals was indexed and integrated using XDS (Kabsch 2010). Miller indices (hkl) were assigned to each reflection in the diffraction pattern, and unit cell size and mosaicity was determined during indexing. Integration is the process of recording multiple intensities, then averaging with associated error, for each reflection. The dataset was converted from a series of images into a numerical list of reflection intensities as these two processes were carried out. A programme called AIMLESS was then used to scale and merge the dataset; processes that make the dataset internally consistant by accounting for variation introduced by radiation damage, beam intensity, and detector inconsistancies. AIMLESS then merges multiple observations of each reflection to generate an average intensity for each reflection, then calculates a list of normalised structure factor amplitudes ($|F_{hkl}|$) based on the inverse relationship between the intensities of the reflections and the structure factor amplitudes.

AIMLESS also calculates a number of statistics to objectively assess the quality of the data:

R_{merge} – assesses the agreement between the average intensity of a reflection, and each individual observation of that reflection: R_{merge} is the basic measure of internal consistancy. Higher quality data generally has a lower R_{merge} value, however R_{merge} increases with multiplicity and so is not always an accurate quality indicator.

R_{pim} - the ‘precison indicating R-factor’ is similar to the R_{merge}, but multiplicity is taken into account. The R_{pim} therefore reflects the increase in accuracy as more observations are merged.

Completeness – the number of unique reflections as a percentage of the total number of possible reflections.

Multiplicity – the average number of independent measurements of a reflection. Multiplicity is calculated by dividing the number of total reflections by the number of unique reflections.

I/σ(I) – the intensities average signal to noise ratio.

Table 5 presents these statistics for Apo-HAZV N and RNA bound-HAZV N datasets. These statistics are acceptable for the 2.7 Å and 3.0 Å resolution cut offs that were chosen.

Table 5 Data collection and scaling statistics.

Two datasets are presented here; Apo-HAZV N and RNA-bound HAZV N, both solved by molecular replacement. Parentheses indicates highest resolution shell values, and confirm that the chosen resolution cut off values are statistically acceptable.

Dataset	Apo-HAZV N	RNA-bound HAZV N
Wavelength (Å)	(i02) 0.9795	(i03) 0.9763
Space group	$P2_12_12_1$	$P2_1$
Cell parameters (Å,°)	a=64.99 b=76.1 c=449.30 α=90 β=90 γ=90	a=64.7 b=84.0 c=104.5 α=90 β=105.74 γ=90
Total relections	252499 (17368)	296006 (22737)
Unique reflections	62009 (4410)	21781
Resolution shells (Å)		
Low	74.88-2.70	50.31-3.0
High	2.77-2.70	3.08-3.0
R_{merge}, %	11.6 (63.1)	10.6
R_{pim} %	8.6 (50.8)	3.1
Completeness, %	98.9 (98.2)	99.9 (100)
Multiplicity	4.1 (3.9)	13.6 (14)
$I/\sigma(I)$	9.3 (1.7)	20.8 (3.5)

In order to allow calculation and monitoring of R_{free} during refinement (which occurs later on in the data processing pipeline) approximately 5 % of data were radomly selected in AIMLESS and kept aside to be omitted from refinement.

3.3.4 The use of molecular replacement to solve the phase problem

The Apo-HAZV N and RNA-bound HAZV N datasets were both solved via molecular replacement (MR) with the globular domain of CCHFV N (PDB accession number: 4AKL); residues 1-182, and 296-482, using a programme called PHASER (McCoy, Grosse-Kunstleve et al. 2007). The process of molecular replacement involves the manipulation of the model molecule; in this case the globular domain of CCHFV N, into the same position and orientation in the unit cell as the unknown molecule. If the model is sufficiently similar to the unknown molecule, the orientation and position of the unknown molecule can be determined, and the phases from the model can be used in the structure factor equations of the unknown molecule, along with its experimental structure factor amplitudes. A Fourier transform can then be applied to generate an electron density map. It is necessary to determine the correct orientation and position of the phasing model compared to the unknown structure in the AU because the phases of the atomic structure factors, and hence of the molecular structure factors, depend upon the location of atoms in the unit cell. If the model used for MR is reasonably accurate the calculated electron density map will show features missing from the original model that can be built into the unknown molecule.

Molecular replacement is made possible through the utilisation of Patterson maps. Patterson maps are derived from the Fourier transformation of the structure factor amplitudes, therefore no knowledge of the phase is required, and results in a map not of the position of atoms within a molecule, but of the vectors between the atoms. There are two broad types of vectors, intra-molecular vectors, representing the distance between two atoms within a molecule, and inter-molecular vectors, that represent the distance between two atoms of neighbouring molecules. The way that the vectors accumulate in a Patterson map can provide a signature for the protein structure, and during MR the orientaton and location of the unknown molecule and model are compared through correlation of the intra- and inter-molecular vectors of their corresponding Patterson maps. In order to place the model structure in the correct orientation and location in the AU, 3 rotation angles and 3 translation parameters need to be specified. Intra-molecular vectors are only dependent on orientation, so these can be exploited in the rotation function, and inter-molecular vectors depend on both the location and orientation of the molecule, therefore these can be exploited in the translation function once the orientation is known. The 3 rotation angles can then be computed from the rotation function, and the translation function can place the oriented molecule in the AU.

MR can be used to solve the phase problem if there is a good structural model for a reasonably large fraction of the unknown structure in the crystal. Structural similarities between proteins correlate well with sequence identity, and as a general rule, if the proteins share >30 % sequence identity, the model can be used for MR of the unknown molecule in a straightforward analysis. As the level of sequence identity decreases, solving the phase problem by MR becomes progressively more difficult. In this study, even though the amino acid sequences of HAZV N and CCHFV N share almost 60 % identity, the fact that the arm domain is in a different position in the two structures (Figure 3.24) meant full length CCHFV N could not be used in MR, only the globular domain (residues 1-182, and 296-482) could be used.

3.3.5 Model refinement of Apo-HAZV N and RNA bound-HAZV N

Refinement seeks to minimise the difference between the calculated structure factor amplitudes of the model, and the experimental structure factor amplitudes derived from the diffraction pattern, to generate the most accurate electron density map. Initially electron density maps were built using phases derived from the CCHFV N globular domain for both Apo-HAZV N and RNA bound-HAZV N. A programme called BUCANEER (Cowtan 2006) was then used to auto-build a preliminary structure into the electron density maps. BUCANEER finds likely Cα positions and extends them laterally, through the addition of more Cα atoms, whilst obeying Ramachandran constraints, to create fragments of peptide chains. Sequence information then enables side chains to be assigned and built into the electron density, and gradually the fragments are extended until they begin to join up. Another programme called COOT (Emsley and Cowtan 2004) was then used to manually adjust the position of each residue in the protein chain to improve the electron density fit, as well as torsion angles and bond lengths.

Refinement of the model using a programme called REFMAC5 (Murshudov, Skubak et al. 2011) was then carried out: REFMAC5 tries to minimise differences between the calculated and observed structure factor amplitudes (R_{cryst}) by improving the atomic positions in the model. The process of improving the model also improves the calculated phases, which are used to build an improved electron density map, which is used to improve the model that fits it, in an iterative cycle. The process of refinement can however introduce bias into the model, as both the model along with the diffraction pattern are used to generate the new electron density map at each round of refinement.

Assessment of refinement progress is therefore determined by calculating the R_{cryst} for the 5 % of data that was omitted during scaling. This is the R_{free} and it calculates how well the current model can be used to predict the structure factor amplitudes (the measured intensities) from the original, unrefined dataset as a cross validation. R_{free}, as well as other statistical calculations can be used to assess the quality of a model; some of these quality indicators are briefly summarised in the following section.

R_{cryst} – also refered to as the R-factor, or R-value. A measure of the difference between the calculated structure factor amplitudes, based on the refined model, and the observed structure factor amplitudes. A random set of atoms will have an R_{cryst} of ~0.63, and a perfect match would have a value of 0. In practice the acceptable aim for protein crystals is to have an R_{cryst} that is less the 10 % of the resolution of the data; for a dataset that diffracted to 2.7 Å, the R_{cryst} should be < 0.27, or 27 %.

R_{free} – a measure of how well the refined model can predict the original structure factor amplitudes from the 5 % data that was omitted completely from the refinement process. This is considered to be less biased that the R_{cryst} and in general is slightly higher then the R_{cryst}. A well refined model should have an R_{cryst} and an R_{free} that do not deviate by more than 5 % - 7 %.

B factors – a measurement of the degree of movement around the atomic positions in the model, this reflects the precision of the atomic positions. This could vary for two main reasons: thermal motion (the vibration of an atom around its rest position) and disorder in the crystal from which the structure was determined (an atom, or groups of atoms not residing in the same position in every AU, or in every unit cell, or in every molecule in the AU). In a high quality model the B factors can reflect the mobility and flexibility in various parts of the molecule. In general low B factors indicate little variation in atomic position, and high B factors indicate greater uncertainty about the actual atomic position.

RMSD bond lengths – root mean squared deviation of the models bond lengths from an accepted set of values based on the geometry of small organic molecules. A refined model should exhibit RMSDs of no more then 0.02 Å for bond lengths.

RMSD bond angles – root mean squared deviations of the models bond angles - similar to bond length, these are constrained to accepted values based on the geometry of small organic molecules. For a fully refined model the RMSD for bond angles should not vary by more the 4 ° from the accepted values.

Ramachandran plots – peptide bonds are planar, and if unconstrained during refinement, should settle down to within a few degrees of planar after the final refinement. Other backbone conformational angles (Φ, along the N – Cα bond and Ψ along the Cα – C bond) are also greatly restricted due to steric repulsion between adjacent side chains. The allowed pairs of angles (Φ and Ψ) on either side of the Cα residue are depicted on a Ramachandran diagram, and occupy ‘allowed’ regions;

either regions that do not give steric repulsion (enclosed by light blue polygon) or only give modest repulsion (dark blue polygon) (Figure 3.14). The shapes of the polygons vary slightly for each of the 20 amino acid residues, and after refinement, all of the residues in a crystallographic structure should reside in ‘allowed regions’ in the calculated Ramachandran plot. Ramachadran plots highlight any regions of the model that are conformationally unrealistic, and if any non-glycine residues lie within ‘forbidden’ areas on the Ramachadran plot following refinement, a good explanation needs to exist in terms of structural constraints that overcome the energetic costs of an unusual backbone conformation. Ramachandran plots of the Apo-HAZV N monomer A are depicted in Figure 3.14.

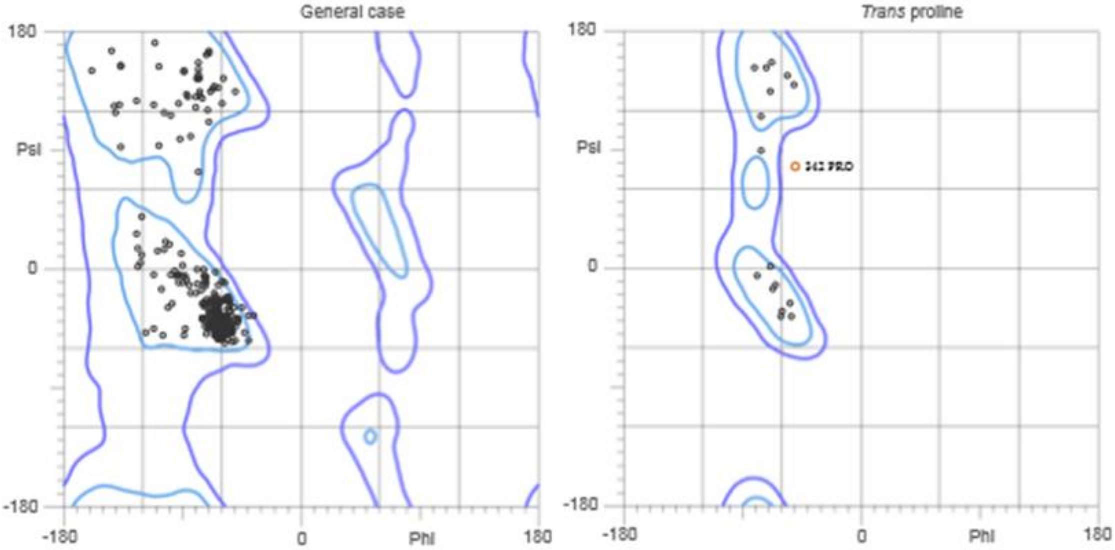


Figure 3.14. Ramachandran plots for Apo-HAZV N monomer A model

Backbone phi and psi angles are plotted for Apo-HAZV N monomer A. ‘Prefered regions’ are light blue, ‘allowed regions’ are dark blue. A single proline residue resides in the ‘forbidden’ region. Plotted using MolProbity.

MolProbity server (Chen, Arendall et al. 2010) – a web based server that analyses the quality of the refined model based on the biophysical criteria described above, and then outputs a series of scores for a particular model. The addition of hydrogen atoms to the model (which are unresolved by crystallography, but whose position is defined by the conformaton of the atoms on which they reside) allows MolProbity to assess unrealistic side chain rotamers, and common misplacement of main chain Cα residues. MolProbity also takes into account side chain clashes and analyses any Ramachandran outliers, then normalises these scores to the resolution of the structure.

Apo-HAZV N was refined to statistically acceptable levels, a summary of the refinement statistics is presented in Table 6. Apo-HAZV N has 4 monomers in the AU, however the electron density for the 4th monomer in the AU was worse then the other monomers (this monomer also had higher B factors). However this did not prevent the model from being completely refined.

After several cycles, refinement of RNA bound-HAZV N became unstable, the R_{cryst} and R_{free} began to diverge. This could be a result of relatively poor quality diffraction images that lead to relatively poor electron density; the overall B factors for RNA bound-HAZV are high (73.54 compared to 44.9 for Apo-HAZV N), and the B factor for the RNA itself is particularly high (95.24). This could indicate a degree of disorder in the crystal that gave rise to the diffraction images, which could occur for a variety of reasons. These include the atoms not occupying the same position in every molecule, or poor cryo-freezing of the crystal, due to suboptimal cryoprotection methods. Refinement of the RNA bound-HAZV N model is not complete, and may well require collection of better quality data; current statistics are presented in table 6.

Table 6 Refinement statistics.

Dataset	Apo-HAZV N	RNA bound-HAZV N
Resolution shells (Å)		
Low	74.88-2.70	50.31-3.00
High	2.77-2.70	3.079-3.00
R_{cryst} %	22.20	20.49
R_{free} %	26.28	29.06
V_M	2.57	2.54
Mol. Per AU	4	2
Reflections working set	58670	20654
Free R-value set (no. of reflections)	5.1% (3125)	5.1 % (1114)
RMSD bond lengths (Å)	0.013	0.009
RMSD bond angles (°)	1.553	1.370
No. atoms used in refinement		
Non-hydrogen atoms	15226	7556
Protein atoms	15156	7496
Water molecules	70	0
RNA atoms	0	60
Mean B factor, Å^2		
Total	44.90	73.54
Protein atoms	44.7	67.24
Water molecules	51.1	N/A
RNA atoms	N/A	95.24
Ramachandran plot statistics, %		
Preferred region	96.49	87.58
Allowed region	2.88	9.24
Outliers	0.27	3.18
Molprobity clashscore for all atoms	2.85	13.4

3.4 RESULTS: THE STRUCTURE OF HAZV N

3.4.1 The HAZV N monomer

To gain mechanistic insights into its multiple functions, we determined the crystal structure of the full length N protein of HAZV strain JC280. The crystal structure of HAZV N was solved by molecular replacement with the globular domain of CCHFV N, residues 1-182, and 296-482; the Apo-HAZV N structure was solved to a maximum resolution of 2.7 Å, and the RNA bound-HAZV N structure to 3 Å. Purified HAZV N was mostly monomeric, and a monomeric view of Apo-HAZV N (monomer A) is shown in Figure 3.15. The description of the HAZV N monomer in the following section applies to both Apo-HAZV N and RNA bound-HAZV N. For all monomers 475 residues were built into the model; residues 187-196 form a disordered loop that is missing in the electron density for all monomers.

Similar to CCHFV N, HAZV N possesses 2 distinct domains, a globular domain and an arm domain. The globular domain contains amino acid residues from both the N- terminus (residues 1-186, silver in Figure 3.15), and the C terminus (residues 298-485, blue in Figure 3.15) of HAZV N. HAZV N is mostly alpha helical; the globular domain contains helix α1 - helix α8 from the N terminus, and helix α13 – helix α20 from the C terminus (alpha helices are numbered according to Figure 3.16). The C-terminal alpha helices form the core of the globular domain, and are surrounded by alpha helices from the N terminus of HAZV N. The arm domain (purple in Figure 3.15) extends away from the globular domain, and is composed of 2 long alpha helices (helix α11 and helix α12), with an exposed loop at the apex, which is supported by a small three-helix bundle (helices α9, α10, and η3). The arm domain is linked to the C-terminus of the globular domain by a single alpha helix (helix α13), however at the N terminus the loop linking the globular domain to the arm domain (residues 187-196) is disordered, and missing in the electron density. This is similar to CCHFV N, where electron density is also missing for the loop linking the arm domain to the N terminus of the globular domain (residues 183-191).

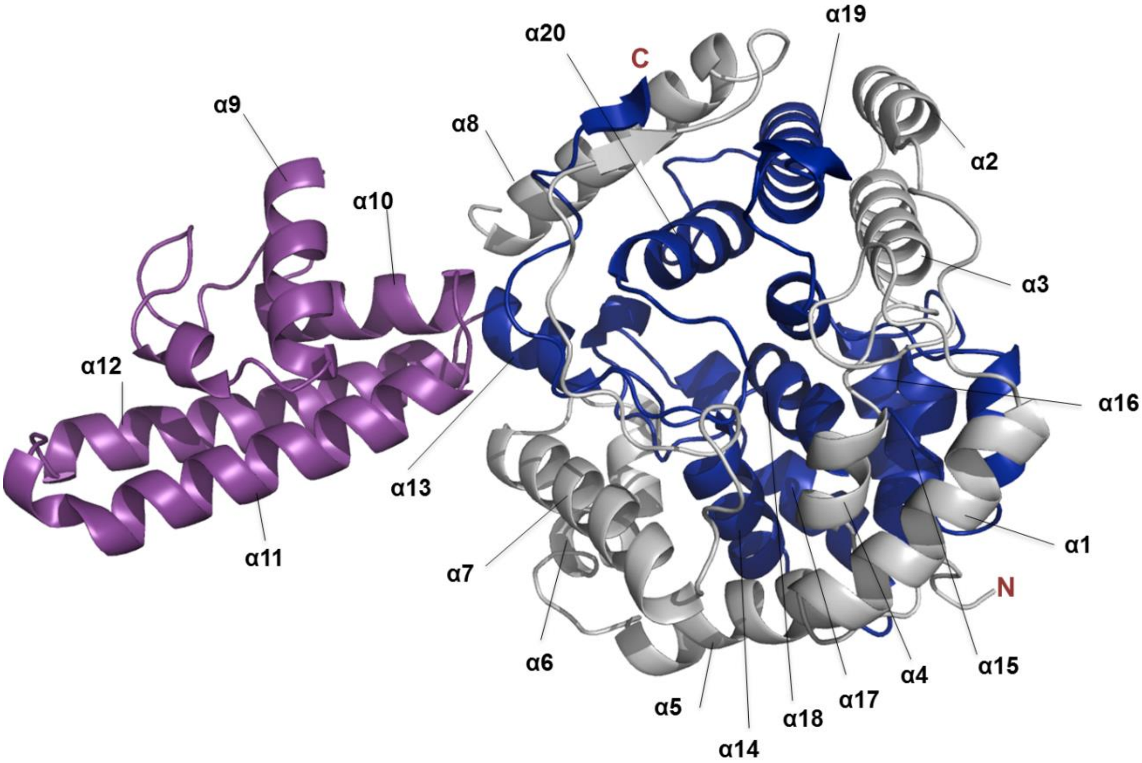


Figure 3.15. The HAZV N monomer.

One monomer from the Apo-HAZV N AU is shown. N- and C-termini are labelled (N and C, respectively). Each monomer contains two domains: the globular domain, composed of residues from the N terminus (residues 1-186, silver), and the C-terminus (residues 298-485, blue), and the arm domain (residues 197-297, purple). The arm domain extends away from the globular domain and is linked to C-terminus of the globular domain by a single alpha helix (helix α13), and to the N terminus by a disordered loop (residues 187-196), which is not visible in this structure. α helices are numbered according to Figure 3.16.

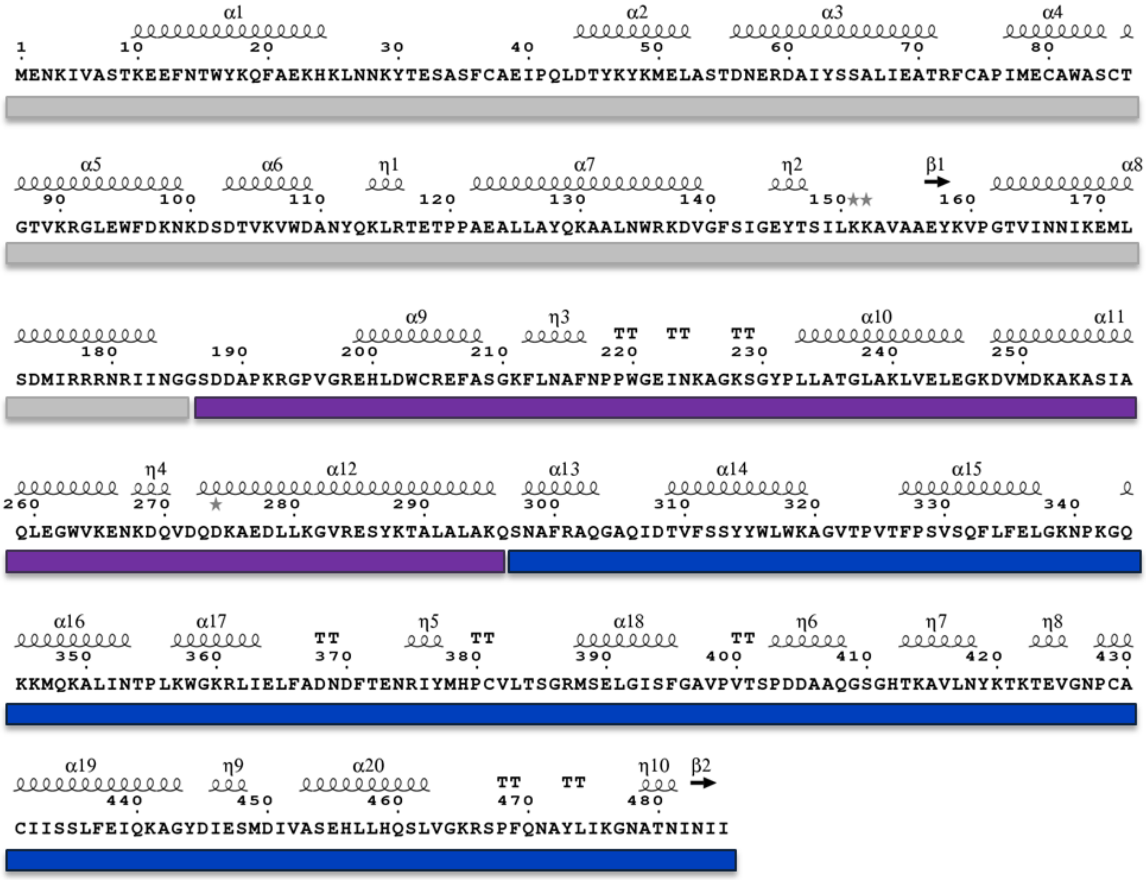


Figure 3.16 Alignment of the secondary structural elements of HAZV N with the primary sequence.

Alpha helices are numbered 1-20 from the N terminus to the C terminus. The N and C termini of the globular domain and the arm domain, are represented by horizontal bars below the sequence coloured according to Figure 3.15. Helix α1 - helix α8 comprise the N terminus of the globular domain, helix α9 – helix α12 the arm domain, and helix α13 – helix α20 the C terminus of the globular domain. (Generated using Espript - http://espript.ibcp.fr *(Gouet, Robert et al. 2003)).* ɳ denotes a helix with 3 residues per turn, α helices have 3.6 residues per turn, and TT denotes a strict β-turn.

Four monomers crystallised in the AU of Apo-HAZV N (Figure 3.17 (B)); the monomers are packed in a head-to-tail arrangement, with two pairs of monomers running anti-parrallel to each other. Contact between the monomers in each pair is between the base of the globular domain of one monomer (residue Pro356) and residues from both the apex of the arm domain, and the supporting 3- helix bundle of the adjacent monomer (residues Leu280, Trp264, Lys276, Val272, Glu271, and Phe217) Figure 3.17 (A). Due to crystal packing, contact between these residues in adjacent monomers (both monomers in the same AU, and monomers in neighbouring AUs) is observed in both Apo-HAZV N and RNA bound-HAZV N. It is thought the 6 residues from the apex of the arm domain and supporting 3-helix bundle form a hydrophobic pocket, in which Pro356 from the adjacent globular domain is buried.

Interestingly, due to the packing of the crystals, the same residues also make contact between the CCHFV N monomers in all three CCHFV N crystal forms (Carter, Surtees et al. 2012; Guo, Wang et al. 2012; Wang, Dutta et al. 2012). The hydrophobic interaction between the arm domain and the base of the globular domain of the adjacent monomer results in the formation of long helical chains of HAZV N and CCHFV N, which can be seen throughout each of the 5 different crystal forms. The conserved nature of this interaction between multiple crystal forms suggests it may be a common feature of nairovirus N proteins, and the structures observed in the crystals may reflect those that form during RNP assembly. Together with the antiparrallel relationship between the pairs of monomers in the Apo-HAZV N AU (which is also observed by (Wang, Dutta et al. 2012) in CCHFV N crystals), this potential multimerisation arrangement resembles that of Influenza virus RNPs. Influenza virus RNPs are double helical, and are composed of 2 N strands of opposite polarity, associating with each other along the helix. Both strands are connected by a short loop at one end of the particle, and associate with the polymerase complex at the other end (Arranz, Coloma et al. 2012).

Mutational analysis would however, have to be performed to determine whether any of the residues involved in the formation of the hydrophobic pocket are also involved in CCHFV N and HAZV N function and transcriptional activity. The conservation of the head-to-tail interaction between the globular domain of one monomer and the arm domain of the adjacent monomer is in contrast to the lateral interactions between both CCHFV N and HAZV N monomers (Figure 3.17). These vary between both Apo-HAZV N and RNA bound-HAZV N crystals, as well as between all three CCHFV N crystal forms.

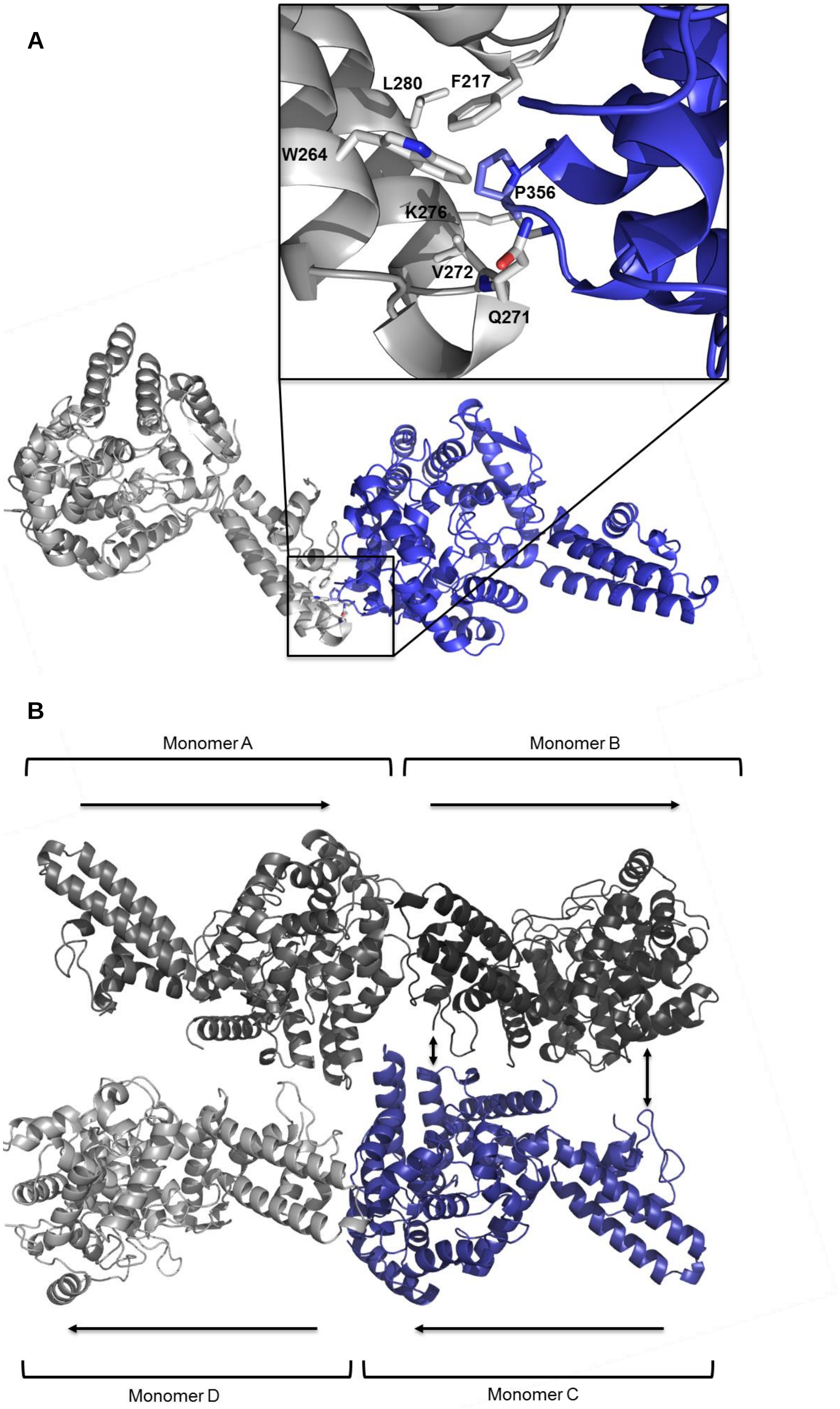

A
L280
F217
W264
P356
K276
V272
Q271
B
Monomer A
Monomer B
Monomer D
Monomer C

Figure 3.17 The 4 monomers in the Apo-HAZV N AU.

(A) A hydrophobic pocket is formed by 6 residues (Leu280, Trp264, Lys276, Val272, Glu271, and Phe217) of the arm domain, in which Pro356 of the adjacent globular domain is buried. (B) HAZV N monomers are arranged in a head-to-tail fashion with two pairs of monomers running antiparrallel to each other in the Apo-HAZV N AU. Electron density for the 4th monomer (black) was poor compared to the other monomers in the AU. Vertical arrows indicte lateral interactions betwen HAZV N monomers. Lateral contacts between monomers vary within the AUs of Apo-HAZV N and RNA bound-HAZV N.

3.4.2 The surface electrostatic potential of Apo-HAZV N

Given the RNA binding capability of HAZV N, the electrostatic surface potential of Apo-HAZV N was analysed. The electrostatic surface potential was calculated using a programme called Adaptive Poisson-Boltzmann Solver (APBS) and Figure 3.18 shows surface representations of the Apo-HAZV N protein colour-coded red to blue according to charge (negative to positive respectively). Areas of positive charge (blue) are linked to electrostatic interactions with the negative phosphate groups of the RNA (or other nucleic acid) backbone. This analysis revealed a positively charged ‘platform’ adjacent to the arm domain (Figure 3.18, left) in a similar position to a positively charged platform on the surface of the CCHFV N protein (Figure 1.11). This was suggested to be a potential RNA binding site, as it is one of the largest areas of continuous positive charge on the surface of both HAZV N and CCHFV N. Other areas of positive charge include an area on the base of the globular domain, and an area on the arm domain.

On the opposite face of the HAZV N monomer to the positively charged platform (rotate HAZV N monomer 180 °, Figure 3.18, right) there is a crevice in the surface of HAVZ N, which extends from the base of the globular domain to helix α13 that links the globular domain to the arm domain. This crevice could also be a potential RNA binding site. There are areas of negative charge on the surface of HAZV N (red), on the apex and side of the arm domain, and boardering the crevice in the globular domain (Figure 3.18, right). However, a large proportion of the surface of HAZV N is positively charged, which is consistent with the relatively high pI of HAZV N (8.73), and its ability to be purified by cation exchange chromatography.

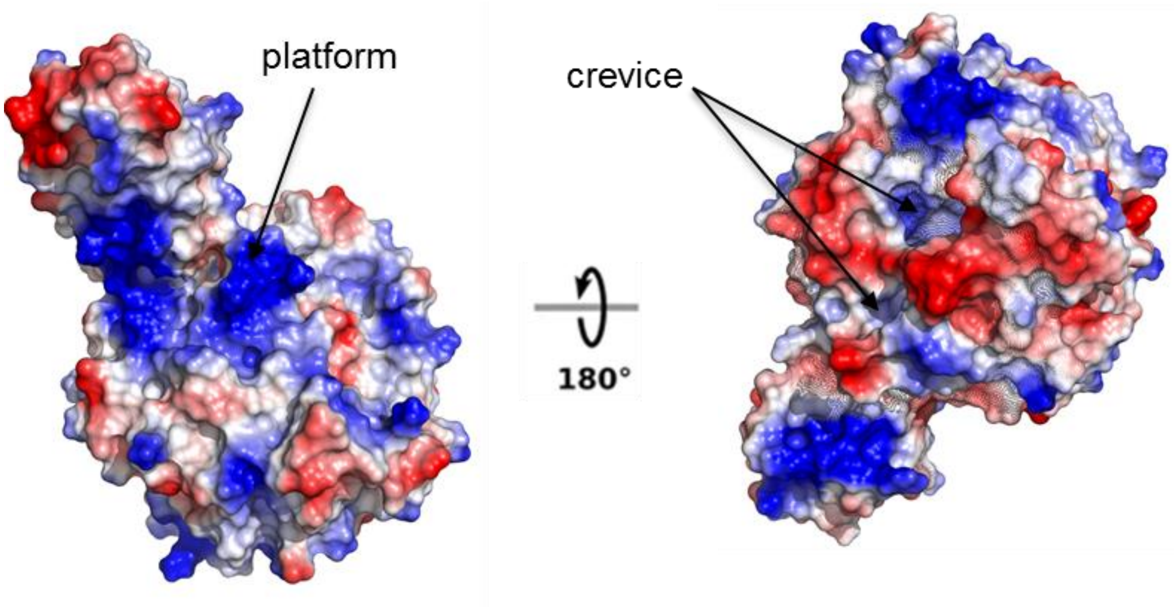


Figure 3.18 Surface representation of Apo-HAZV N electrostatic potential.

The surface charge of Apo-HAZV N. Electrostatic surface potential is shown colour coded from red (negative) to blue (positive) in dimensionless units of k_b T e_c^{-1}, where k_b is Boltzmann's constant, T is the temperature, and e_c is the charge of an electron.

3.4.3 The RNA-bound HAZV N model

In the previous section analysis of the electrostatic surface potential of Apo-HAZV N revealed several regions that could potentially bind RNA. In the RNA bound-HAZV N structure, the crevice in the globular domain was shown to be the RNA binding site, rather than the positively charged platform adjacent to the arm domain. RNA bound-HAZV N was solved by molecular replacement with the globular domain of CCHFV N to a final resolution of 3 Å, and contained 2 monomers in the AU (Figure 3.19 (A)). In one of the monomers, three nucleotides of ssRNA were bound within the crevice in the globular domain that extends from the base of the globular domain to helix α13 that links the globular domain to the arm domain (Figure 3.19 (B)).

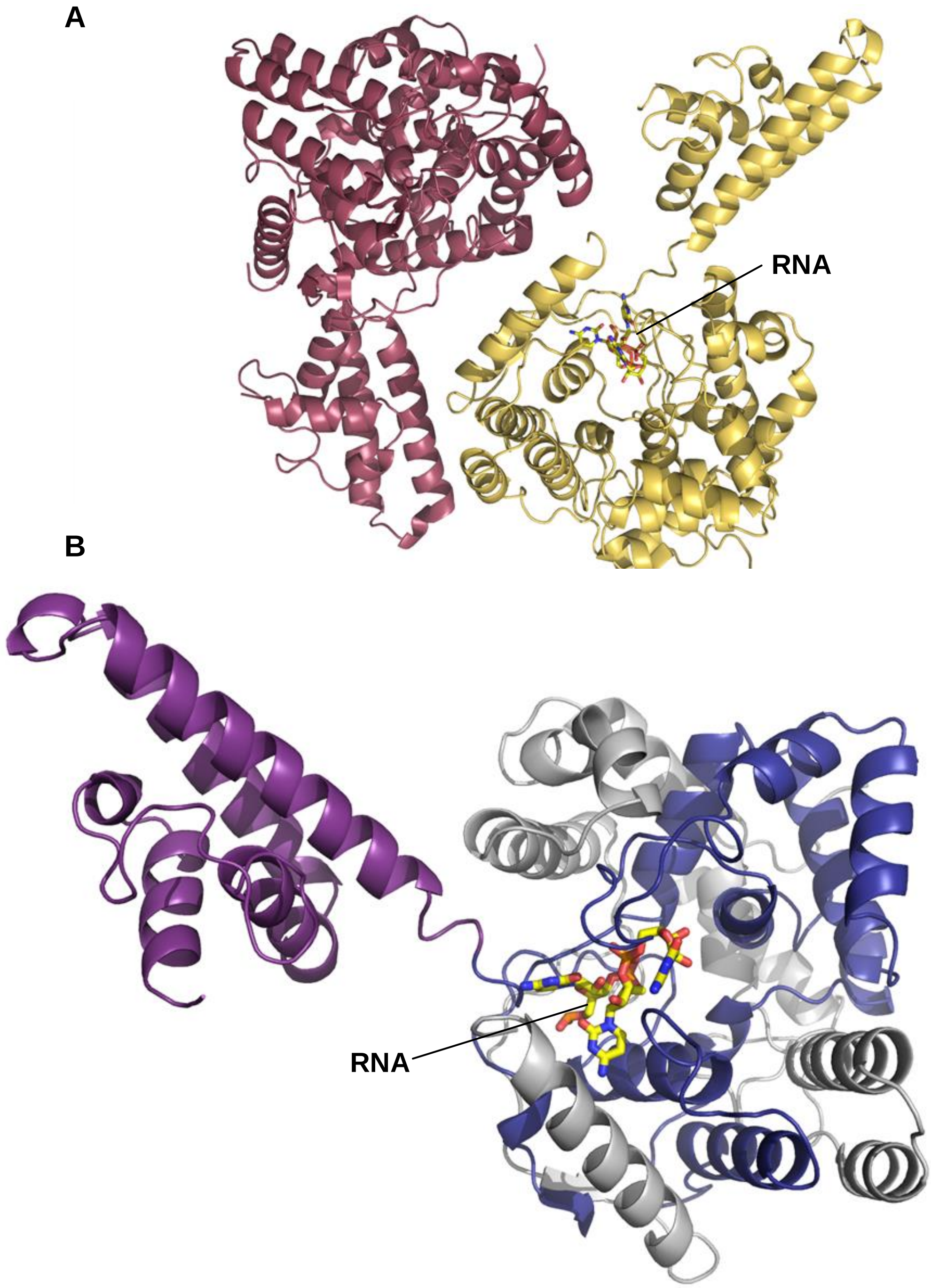


Figure 3.19. The two monomers in the RNA bound-HAZV N AU

RNA bound-HAZV N crystallised with 2 monomers in the AU (A), one monomer had RNA bound (yellow) and one did not (raspberry). (B) The RNA bound-HAZV N monomer is depicted in the same fashion as the Apo-HAZV N model for clarity; the N terminus is silver, the C terminus is blue, and the extended arm domain is purple. The RNA is depicted predominantly in yellow (carbon); phosphates are orange, nitrogen atoms are blue, and oxygen atoms are red.

The overall structure of RNA bound-HAZV N is the same as Apo-HAZV N; the description of the Apo-HAZV N monomer in section 3.4.1 can also be applied to the RNA-bound HAZV N monomer. As for Apo-HAZV N, RNA bound-HAZV N is mostly α helical, and consists of two domains; a globular domain composed of residues from both the N- and C- termini, and an arm domain that extends away from the globular domain. The major difference between Apo-HAZV N and RNA bound-HAZV N resides in the position of the arm domain relative to the globular domain. This is discussed further in section 3.5.1.

The three RNA nucleotides bound within the crevice may represent RNA that is inaccessible to the RNase enzyme used during purification of HAZV N, due to its buried position. This might explain why treatment of HAZV N with RNase was not sufficient to completely remove bacterially-derived RNA from HAZV N during purification, even after extensive RNase treatment (wash buffers containing high concentrations of NaCl were used to disrupt the interaction between the RNA and protein instead).

3.4.4 The electrostatic surface potential of RNA bound-HAZV N

The electrostatic surface potential of RNA bound-HAZV N was calculated as for Apo-HAZV N (section 3.4.2). The positively charged platform adjacent to the arm domain is highlighted in Figure 3.20, as is the crevice in the globular domain where RNA is bound. Comparison of the surface electrostatic potential of Apo-HAZV N (Figure 3.18) and RNA bound-HAZV N (Figure 3.20) reveals little difference in the overall surface electrostatic potential between the two models. The distribution of the areas of positive charge varies according to the change in position of the arm domain, which is discussed further in section 3.5.1. This arm domain movement also slightly alters the position of the negatively charged areas boardering the RNA binding crevice, however whether this has any bearing on the RNA binding ability of HAZV N is unknown.

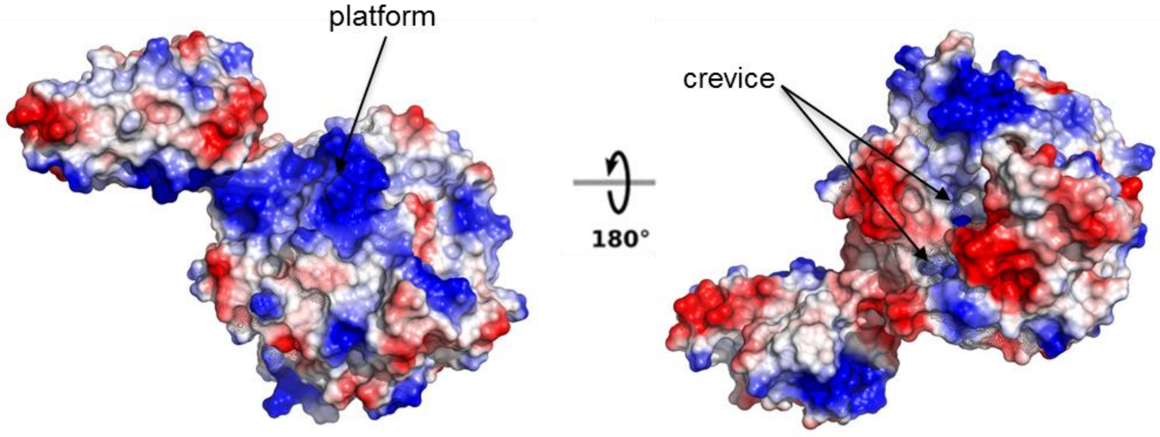


Figure 3.20 Surface representation of RNA bound-HAZV N electrostatic potential.

The surface charge of RNA-bound HAZV N. Electrostatic surface potential is shown colour coded from red (negative) to blue (positive) in dimensionless units of k_b T e_c^{-1}, where k_b is Boltzmann's constant, T is the temperature, and e_c is the charge of an electron.

3.4.5 Residues of HAZV N potentially involved in RNA binding

The 3 nucleotides of ssRNA bound to HAZV N are random RNA sequences derived from the expression host, and this structure of HAZV N is therefore unlikely to reveal any sequence specificity. Furthermore the electron density for the 3 RNA nucleotides is not good enough to determine whether these bases are purines or pyrimidines (for example if a purine base was prefered at a specific position, as has shown to be the case for LASV N protein). The 3 RNA nucleotides are oriented with their phosphate backbones contacting the foundation of the crevice, and the bases pointing outwards, towards the opening of the crevice. It is thought, that in a similar manner to LASV N (Hastie, Liu et al. 2011), an extensive network of interactions anchor the RNA phosphate backbone and RNA bases to HAZV N. Binding is thought to occur through hydrogen bond interactions between key amino acid residues, and both phosphates of the RNA backbone and the RNA bases.

Previously, to characterise the roles of residues involved in RNA binding, or the potential dimerisation site of CCHFV N, a panel of CCHFV N proteins with mutations in these residues were examined for their ability to promote luciferase reporter gene expression within a CCHFV mini-genome (Carter, Surtees et al. 2012). Mutations in 3 residues essentially abrogated CCHFV mini-genome transcriptional activity when compared to wild type N (K132, Q300 and K411). When the position of these mutated residues are mapped onto the RNA bound-HAZV N structure (residues N134, Q303 and K414 in HAZV N) 2 out of these 3 residues (K414 and Q303) are within bonding distance of the RNA backbone in the globular domain RNA binding crevice (Figure 3.21). Given the close structural similarity between the globular domains of HAZV N and CCHFV N (discussed in section 3.5.2), these data strongly suggest that these residues are involved in RNA binding in HAZV N, and by inference RNA binding in CCHFV N as well (rather than any other effect that could regulate mini-genome output).

K132/N134 is located on an N terminal alpha helix facing away from the crevice, and does not contact the RNA bound in the crevice. This residue could however be involved in dimerisation, or RNA binding of longer sequences of RNA, or L protein interaction. Mutation of residue H456 (residue H459 in HAZV N), which is positioned close to one of the RNA bases, has a less pronounced effect, resulting in a loss of ~50 % luciferase signal, and mutation of residue H453 (residue H456 in HAZV N) has no effect on minireplicon activity. This is similar to LASV N, where mutation of several residues contacting the bound RNA has no effect on mini-genome activity, whereas mutation of others abrogates mini-genome activity, or decreases the reporter gene signal by ~50 % (Hastie, Liu et al. 2011). CCHFV N residue K342 (residue K345 in HAZV N) was also mutated, as it formed part of the positively charged suspected RNA binding site adjacent to the crevice near the base of the globular domain. Mutation of K342 however, has no effect on CCHFV mini-genome function, and K345 is not located near to the bound RNA in HAZV N.

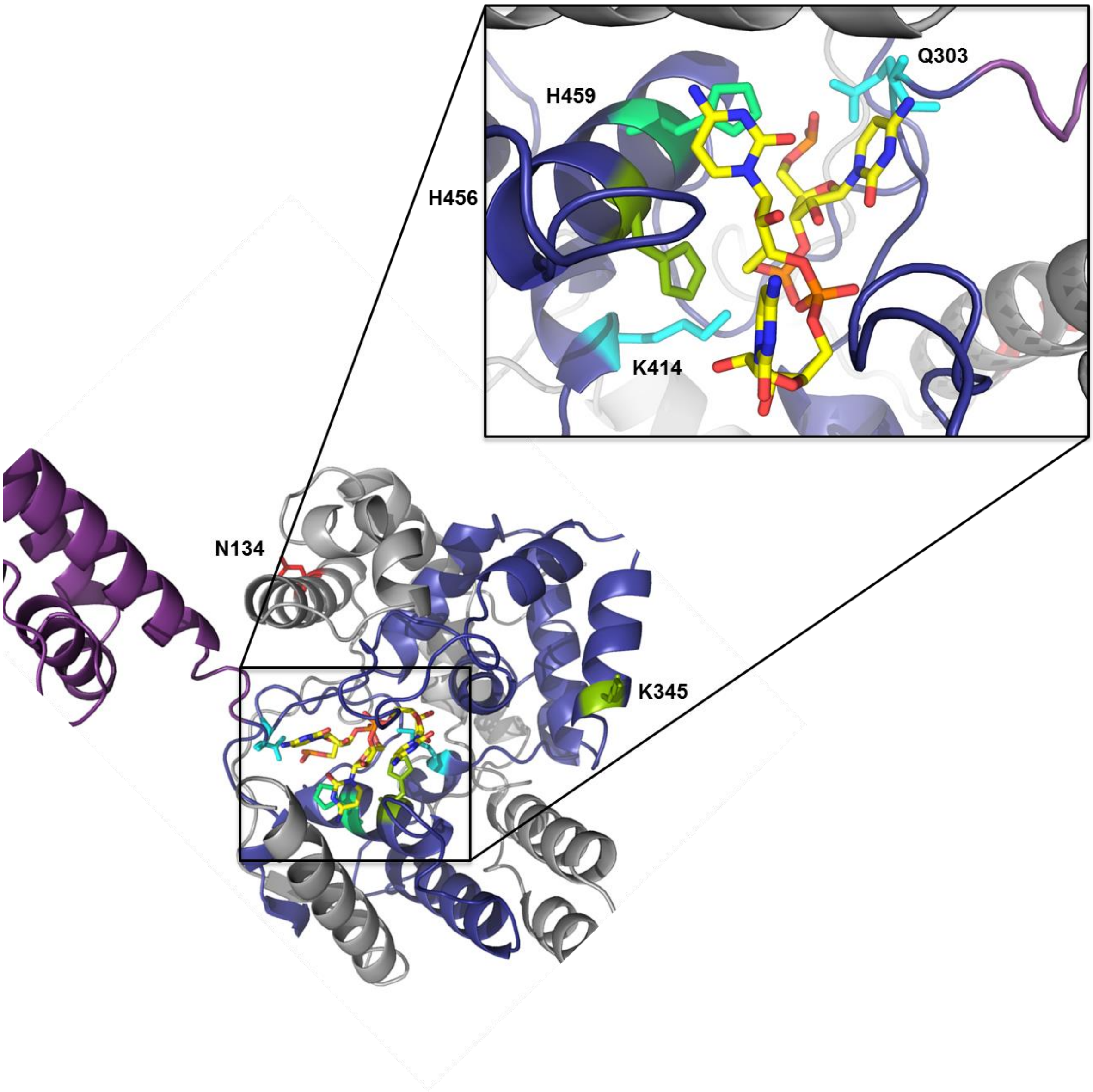


Figure 3.21 Potential RNA binding residues of RNA bound-HAZV N.

The RNA bound-HAZV N monomer with potential RNA binding residues Q303, K414, H456, H459, N134 and K345 highlighted. Mutation of Q300 and K411 in CCHFV N essentially abrogated CCHFV mini-genome transcriptional activity, and mutation of H456 reduced reporter gene expression by ~50 %. The corresponding positions of Q303, K414 and H459 in RNA bound-HAZV N, where they make close contact with both the RNA backbone and the RNA bases, strongly suggest these residues may be important for RNA binding.

3.5 COMPARISON OF HAZV N WITH OTHER STRUCTURES

3.5.1 Apo-HAZV N and RNA bound-HAZV N

RNA-bound HAZV N and Apo-HAZV N consist of two domains; a globular domain and an extended arm domain. Superposition of Apo-HAZV N and RNA-bound HAZV N reveals close structural similarity in the globular domain (RMSD = 1.1 for 368 residues) (Figure 3.22). In contrast, whilst the fold in the arm domain is the same (RMSD = 0.594 for 107 residues) the position of the arm domain is greatly altered with respect to the globular domain. The arm is rotated by about 50 °, and the apex of the arm domain is shifted by a distance of 29.19 Å (measured from the Cα position of Asp269, at the apex of the arm in each monomer). It is possible that the movement of the arm domain may be related to the mechanism of RNA binding, analogous to the gating mechanism proposed for LASV N. In the absence of RNA part of helix α5 of LASV N, and all of helix α6 lie across the RNA-binding crevice, occluding access to RNA. Upon RNA binding the loop linking helix α5 and helix α6 shifts outwards, helix α5 shortens, and helix α6 is rotated away from the crevice, enabling RNA to bind (Hastie, Liu et al. 2011). The electron density of RNA-bound HAZV N is too poor to determine whether a similar mechanism occurs upon HAZV N binding RNA; for example, what would be helix α13 in Apo-HAZV N (analogous to helix α5 of LASV, Figure 1.13) does not have secondary structural assignment in RNA bound-HAZV N. However, this could either be because it is more flexible and part of the RNA binding mechanism, or the electron density is not good enough to assign a secondary structure.

Figure 3.22 Superposition of Apo-HAZV N and RNA bound-HAZV N.

Apo-HAZV N (raspberry) and RNA bound HAZV N (yellow) structurally have very similar globular domains, however the relative position of the arm domain is greatly altered. Upon RNA binding, the arm domain is rotated by about 50 ° and the apex of the arm domain is shifted by 29.19 Å. The change in position of the arm domain may be related to the mechanism of RNA binding, however further work is needed to confirm the mechanism of HAZV N RNA binding.

The arm domain of CCHFV N is known to be flexible, and free to explore a range of conformations, indeed it is thought the flexibility of the arm domains of bunyavirus N proteins in general are essential to their function. The ability of the arm domain of HAZV N to adopt several different positions is therefore to be expected, however there appears to be 2 prefered general positions; one which is adopted when RNA is bound, and the other that is adopted when RNA is not bound (Figure 3.23). The superposition of the 4 Apo-N monomers in the AU as well as the RNA-free monomer from the AU of RNA bound-HAZV N shows that although the position of the arm domain does change slightly (similar to observations of CCHFV N by (Wang, Dutta et al. 2012), it is nowhere near the huge shift which is seen in the RNA bound structure. This suggests that although the arm domain is free to explore different conformations, it generally resides in the same position when RNA is not bound.

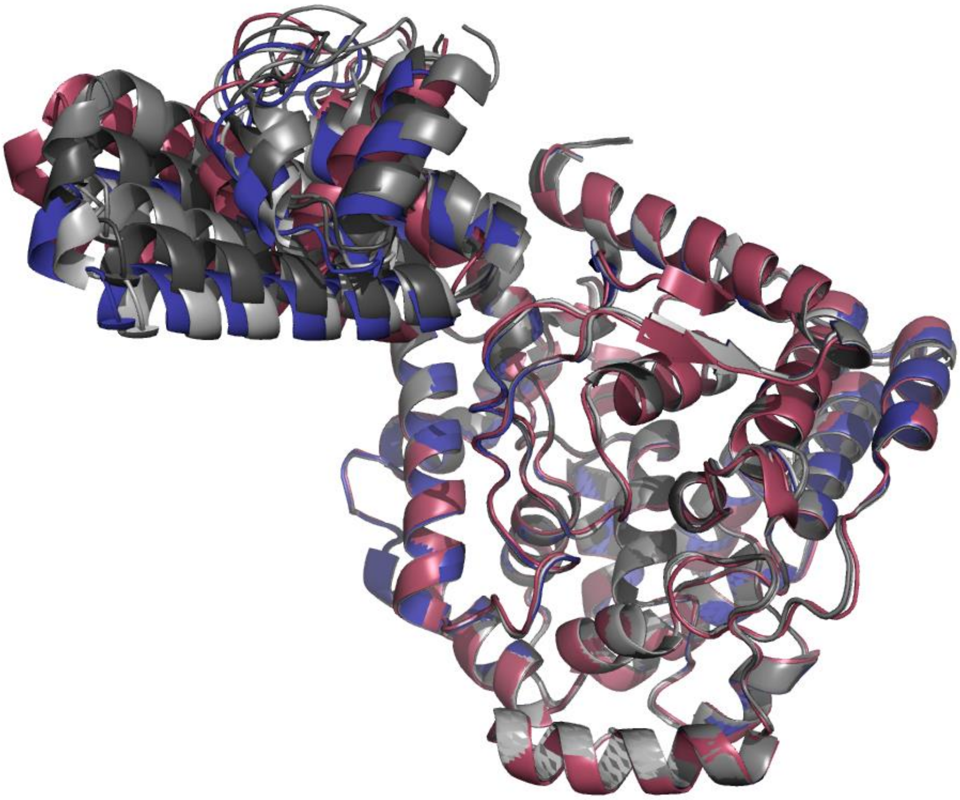


Figure 3.23 Superposition of RNA free HAZV N

Superposition of the 4 monomers from the Apo-HAZV N AU (greys and blue), and the RNA free monomer from the RNA bound HAZV N AU (raspberry). The globular domains superpose almost exactly, and the arm domains reside in a similar position, with some slight deviations, similar to CCHFV N (Wang, Dutta et al. 2012).

3.5.2 HAZV N and CCHFV N

HAZV N and CCHFV N likely perform the same function in their viral replication cycles, and share ~60 % sequence identity (Figure 1.12); this is reflected in the structural similarity between these proteins, which enabled the structure of HAZV N to be solved by molecular replacement with the globular domain of CCHFV N (residues 1-182, and 296-482). In good agreement with this, the crystal structures of HAZV N and CCHFV N are very similar (Figure 3.24). The globular domains of HAZV N and CCHFV N align extremely closely (RMSD = 0.698, over 368 residues that share 63 % sequence identity), as do the arm domains (RMSD = 0.886 over 107 residues that share 48 % sequence identity), however the arm domain adopts a different position relative to the globular domain (Figure 3.24). The arm domain of CCHFV N is rotated by 73.96 ° and translated by 27.05 Å (measured from the Cα of Asp 266/269), compared to the arm domain of HAZV N. Although neither HAZV N nor CCHFV N in Figure 3.24 have RNA bound, the difference in the position of the arm domain could be due to the different primary amino acid sequences of these proteins. Alternatively, the change in the position of the arm domain could also be related to differences in the crystal packing of these proteins.

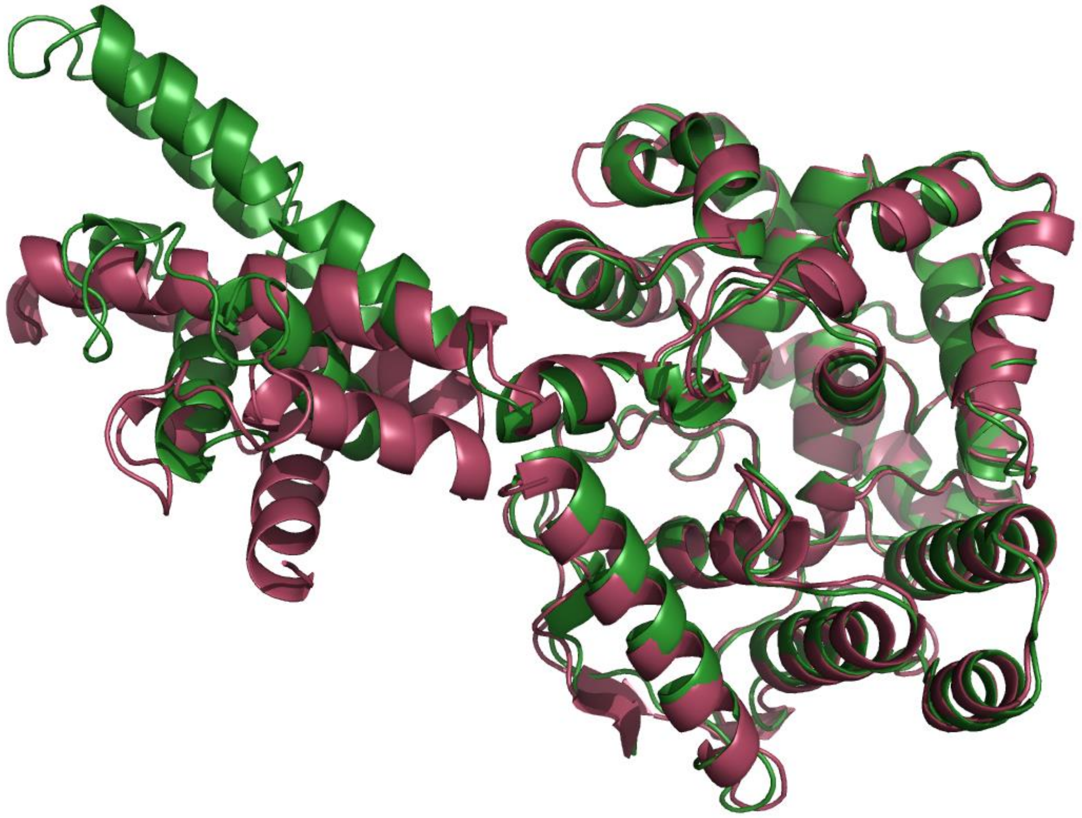


Figure 3.24 Superposition of CCHFV N and HAZV N

The globular domain of CCHFV N (green) and HAZV N (raspberry) superpose extremely closely. The relative position of the arm domain of CCHFV N is shifted compared to HAZV N through a rotation of 73.96 ° and translation of 27.05 Å.

3.5.3 HAZV N and LASV N

The globular domain of CCHFV N is structurally similar to the N-terminal RNA binding domain of LASV N, a member of the bisegemented *Arenaviridae* family (RMSD 3.38 Å for 237 residues with 8.01 % sequence identity) (Carter, Surtees et al. 2012). Unsurprisingly, the globular domain of HAZV N also aligns closely with the N terminal domain of LASV N; (Cα RMSD 3.426 for 240 residues) (Figure 3.25). Interestingly, superposition of RNA bound-HAZV N with the N-terminal domain of RNA bound LASV N (Figure 3.25) revealed not only the close structural similarity of the protein fold in the globular domain, but also a very close similarity in the position of the RNA in the RNA binding crevice. Of the three RNA bases that are present in the crevice in the globular domain of HAZV N, the position of base 2 and base 3 overlap between HAZV N RNA compared to LASV N RNA. The position of base 1 varies slightly between the RNA strands, as does the position of the ribose-phosphate backbone, however, overall the position of the RNA nucleotides is very similar in these two structurally related proteins.

The electron density is poor for regions of the RNA bound-HAZV N protein, and also poor for the RNA in the crevice; the RNA B factors are very high, and the electron density is insufficiently accurate to determine the precise position of each atom of the 3 RNA nucleotides. However, the position of the RNA in HAZV N was determined independently of the LASV N N-terminal domain structure, therefore the close overlap in the position of RNA in HAZV N and one of its closest structural relatives, LASV N, supports the placement of the RNA in the crevice of the globular domain of HAZV N as being correct.

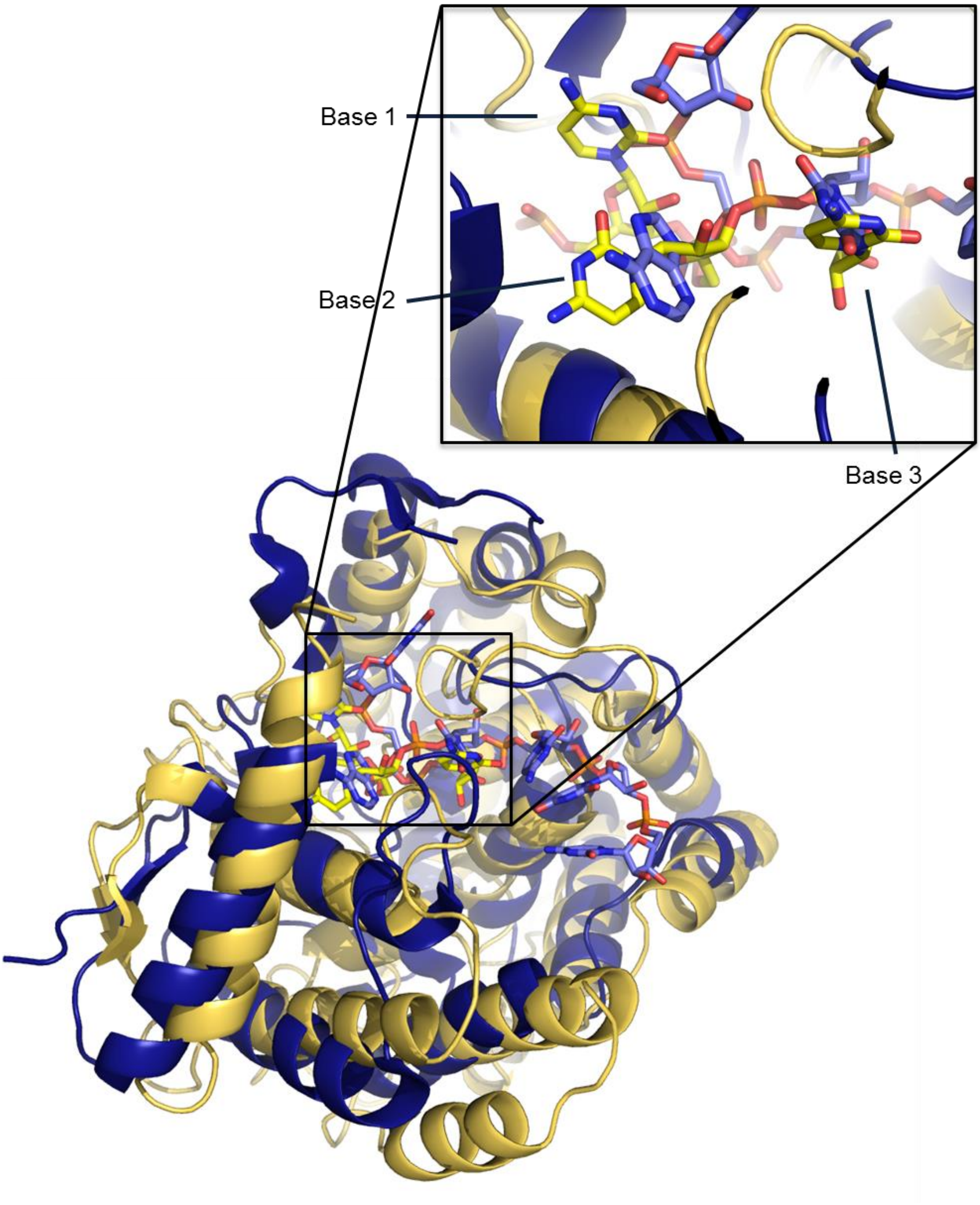


Figure 3.25 Superposition of the globular domain of RNA bound-HAZV N, and the RNA bound N terminal domain of LASV N.

The globular domain of RNA bound-HAZV N (yellow) and the N terminus of LASV N (blue) align quite closely (RMSD = 3.426). The RNA of HAZV N (yellow) and LASV N (light blue) also aligns well; the positions of base 2 and base 3 overlap between the two strands of ssRNA, with slight deviations in the position of base 1 and the phosphate backbone. Overall, the similarity of the RNA positions supports the correct placement of RNA in HAZV N, despite high B factors and poor electron density.

3.6 CHAPTER SUMMARY AND DISCUSSION

This chapter presents the optimisation of expression and purification of HAZV N protein, and determination of the X-ray structure of both Apo-HAZV N and RNA-bound HAZV N, following crystallisation, X-ray diffraction and data collection. HAZV-N was expressed from pGEX-GST-HAZV N, GST-HAZV N was purified using glutathione affinity chromatography, cleaved with HRV 3C protease, and HAZV N further isolated by cation exchange chromatography. This resulted in the purification of 4-7 mgs >95 % pure HAZV N, per 4 L bacterial culture. RNA bound-HAZV N was purified by omitting a high NaCl concentration binding buffer wash prior to elution of GST-HAZV N from the GST resin. This resulted in the purification of RNA bound-HAZV N with an A_{260}/A_{280} ratio of 1.04, and Apo-HAZV N with an A_{260}/A_{280} ratio of 0.84, indicating RNA-bound HAZV N co-purified with more nucleic acid. 1 mg of >95 % pure HAZV N was used to generate polyclonal antibodies targeting HAZV N, using an immunisation programme which is UK home office and USA National Institutes of Health compliant.

Crystals of Apo-HAZV N diffracted to 2.7 Å in the $P2_12_12_1$ space group, and crystals of RNA bound-HAZV N diffracted to 3 Å in the $P2_1$ space group. Both datasets were solved by molecular replacement with the globular domain of CCHFV N (residues 1-182 and 296-482). Apo-HAZV N and RNA bound-HAZV N both have very similar structures, and similar overall surface electrostatic potential. Apo-HAZV N and RNA bound HAZV N are mainly alpha-helical, containing a globular domain composed of 16 α helices from both the N and C termini of the protein, and an arm domain composed of 2 long alpha helices, that extends away from the globular domain. Contact within the pairs of monomers in the Apo-HAZV N AU occurs through a hydrophobic interaction, where six residues from the arm domain of one monomer forms a hydrophobic pocket in which Pro356 of the adjacent globular domain is buried. This interaction occurs in both Apo-HAZV N and RNA bound HAZV N crystals, as well as in all three CCHFV N crystal forms, resulting in the formation of long helical chains of N protein throughout these crystals. It could be this is a common feature of nairovirus N proteins, and may reflect the mechanism of multimerisation of these proteins during RNP formation, although mutational analysis would have to be performed to confirm this.

The globular domains and arm domains of Apo-HAZV N and RNA bound HAZV N are very similar, they superpose almost exactly, with a Cα RMSD of 1.1 over 368 amino acids, and 0.594 over 107 residues, respectively. The arm domain has shifted position dramatically in the RNA bound-HAZV N structure compared to the

Apo-HAZV N structure however, through a translation of 29.19 Å and a rotation of ~50 °, measured from the Cα of Asp 269. It is possible that this change in the position of the arm domain may be connected with the mechanism of RNA binding, analagous to the mechanism proposed for the nucleocapsid protein of LASV (Hastie, Liu et al. 2011). Superposition of RNA bound-HAZV N and the RNA bound N terminal domain of LASV N reveals that the RNA resides in a very similar position (base 2 and base 3 overlap), supporting the crevice in the globular domain as the binding site for RNA in HAZV N, even though the electron density is poor. LASV N helix α5 corresponds to the beginning of HAZV N helix α13, which links the arm domain to the globular domain, and has changed position in Apo-HAZV N compared to RNA bound-HAZV N. This supports the theory that the movement of the arm domain of HAZV N might possibly 'gate' RNA binding, in a similar manner to LASV N.

Additionally, superposition of the 4 monomers from the AU of Apo-HAZV N, and the RNA free monomer from the AU of RNA-bound HAZV N, revealed that both the globular domains and arm domains align almost exactly; there are slight shifts in the position of the arm domain, but these are nowhere near the movement that is observed between Apo-HAZV N and RNA-bound HAZV N arm domains. The similarity in the position of the HAZV N arm domains in five monomers, from two different unit cells and two different crystals, suggests that although the arm domain is free to explore different conformations, it generally resides in the same position when RNA is not bound. This supports the theory that a specific trigger (such as the presence of RNA, or another ligand or cofactor) is required before the arm domain will change position (possibly in order to enable HAZV N to bind RNA in some sort of gating mechanism). It is thought that in the absence of a P-protein (the phosphoprotein, a protein present in the non-segmented NSVs), sNSVs must have a mechanism to ensure monomeric N does not bind cellular RNA non-specifically, or multimerise inappropriately, during the virus replication cycle (Green, Cox et al. 2014). A lot more work needs to be carried out however, to determine whether the change in the position of the HAZV N arm domain is involved in the mechanism of HAZV N RNA binding, as the electron density both for regions of the protein domains, and the RNA, is too poor to confirm an RNA gating mechanism from this study.

This is the first time RNA has been co-crystallised with a nairovirus nucleocapsid protein. Three nucleotides of RNA bound in a crevice in the globular domain, that extends from the base of the globular domain to the helix α13, which links the globular domain to the arm domain. The orientation of the nucleotides is with the phosphate backbone towards the base of the crevice with the bases pointing outwards, towards the opening of the crevice. It is thought the buried position of the nucleotides in the crevice provides protection from RNase treatment during purification, and that the nucleotides are not available to base pair during RNA synthesis when bound to HAZV N in this crevice. Therefore, in a similar manner to LASV N, the N protein surely needs to dissociate from the RNA in the RNP during RNA synthesis, before the L protein can access viral RNA. In the case of LASV, it has been proposed that the L protein induces a conformational change in N in LASV RNPs, enabling N to transiently dissociate from the RNA to allow access by the L protein, before N re-encapsidates the RNA (Hastie, Liu et al. 2011). It is possible that an analagous mechanism also exists for HAZV N.

HAZV N is structurally extremely similar to CCHFV N. The globular domains align very closely (RMSD = 0.698, over 368 residues that share 63 % sequence identity), and the arm domains occupy different positions relative to the globular domains (the arm domain is rotated by 73.96 °, and translated by 27.05 Å, measured from the Asp 266/269 Cα in CCHFV N compared to HAZV N). The arm domains might adopt different positions due to primary sequence differences, or differences in crystal packing. The arm domain is flexible – the single alpha helix, and flexible loop (not visible in electron density for either structure) linking the arm domain and the globular domain is not structurally rigid, and is thought to allow the arm domains of both CCHFV N and HAZV N to adopt different conformations. This flexibility between domains is proposed to be a common feature of sNSV nucleocapsid proteins, that results in the greater flexibility of sNSV RNPs that enables the RNPs to carry out their specific functions (Reguera, Cusack et al. 2014). In good agreement with this, the arm domain of HAZV N can also adopt a number of different positions.

The extreme similarity in the globular domains of CCHFV N and HAZV N (the site of RNA binding), and the overall similarity in the organisation of the domains, suggests that CCHFV N and HAZV N adopt the same mechanisms to carry out their functions within their virus replication cycles. This goes someway towards validating the use of HAZV as a model for CCHFV infection. Furthermore, careful structure-based design of small-molecule inhibitors targeting the globular domain of CCHFV, should act as effective antivirals against both HAZV and CCHFV. The globular domain of CCHFV N and HAZV N is an excellent target for small-molecule inhibitors as it is the site of RNA binding; an essential function of these proteins. The RNA is sequestered in a crevice in the globular domain, allowing for the design of structure-based antiviral inhibitors aimed at preventing RNA binding. The fact that HAZV N and CCHFV N have such similar globular domains should enable HAZV to act as a model for CCHFV infection when examining the potential efficacy of these small molecule inhibitors under containment level 2 conditions, without the need to perform experiments under containment level 4 conditions.

The similarity in the globular domains of HAZV and CCHFV has also enabled the identification of residues potentially important in RNA binding. Mutation of residues in CCHFV N thought to be involved in RNA binding or dimerisation, and examination of the ability of mutant CCHFV N to promote reporter gene expression in a CCHFV mini-genome, identified 3 residues whose mutation abrogated reporter gene expression (K132, Q300 and K411) (Carter, Surtees et al. 2012). Examination of the location of these residues in RNA-bound HAZV N revealed that 2 of these residues (Q303 and K414) lie within bonding distance of the RNA backbone. This strongly suggests these residues are essential for or are involved in RNA binding in HAZV N and CCHFV N, as mutating these residues prevented CCHFV N from carrying out essential functions. Observations such as this support the proposal that CCHFV N and HAZV N have the same mechanisms of action, and that HAZV is a suitable model for CCHFV.

CHAPTER 4: IDENTIFYING HOST PROTEINS THAT INTERACT WITH CCHFV N AND HAZV N

4.1: CHAPTER INTRODUCTION

RNA viruses, including CCHFV and HAZV, have limited coding capacity and to compensate for this, their proteins exhibit remarkable structural economy, often possessing multiple functional domains, and the ability to interact with both viral proteins and components from the host cell. In addition to RNP formation, where the nairovirus N protein makes both homotypic N-N interactions and N-RNA interactions, it is thought the nairovirus N protein interacts with numerous host cell proteins. Interactions between host cell and viral proteins can have important consequences for virus replication; CCHFV N interacts with both actin, and Mx1, in the context of virus infection; the interaction of CCHFV N with actin aids viral replication, it is thought by trafficking CCHFV N to perinuclear regions of the cell where virus ‘factories’ form, whereas Mx1 is thought to inhibit replication (Andersson, Bladh et al. 2004; Andersson, Simon et al. 2004). This chapter aims to identify interactions with other cellular proteins, and examine whether CCHFV N and HAZV N therefore have other functions outside of RNP formation.

This chapter describes the identification of cellular proteins that interact with an (enhanced green fluorescent protein) EGFP-CCHFV N fusion protein by stable isotope labelling of amino acids in cell culture (SILAC) based mass spectrometry, coupled to EGFP immunoprecipitations (IPs). EGFP IPs were carried out with a GFP-Trap, that is composed of monoclonal anti-EGFP antibodies covalently coupled to agarose beads. This chapter then describes the validation of cellular heat shock protein 70 (HSP70) as an interacting partner of untagged CCHFV N, both when CCHFV N is expressed from a plasmid, and in the context of actively replicating CCHFV. Finally this chapter describes the interaction of HSP70 with HAZV N, both within cells, and within fractions containing purified HAZV, and preliminary experiments to understand the functional significance of the HSP70-HAZV N interaction during the replication cycle of HAZV.

4.1.1 Introduction to SILAC based mass spectrometry coupled to GFP-trap IPs

Immunoprecipitations (IPs) coupled to stable isotope labelling by amino acids in cell culture (SILAC) in combination with mass spectrometry (MS), provides a method for the unbiased identification of cellular proteins that co-IP with a ‘bait’ protein, within the context of the cellular environment. MS provides a reliable method of identifying peptides, however MS is not innately quantitative; approximate protein abundance can be estimated using the emPAI score (the exponentially modified protein abundance index; described in more detail in section 4.7.3), however this is not as precise as quantitation methods based around isotope labelling (Mann 2006). One of these isotope labelling techniques is SILAC, which is a high throughput, quantitative, proteomic technique that can be used to compare relative protein levels between two or more different cell populations. SILAC can be used to assess global proteome changes, or changes in the ‘interactome’; i.e. the total number of cellular proteins that interact with a specific ‘bait’ protein in comparison to a negative control.

Two or more cell populations are distinguished by growing cells in media containing isotopically labelled amino acids; these become incorporated into cellular proteins over time, as cell grow normally, and replace their equivalent unlabelled amino acids (5-7 cell divisions is generally sufficient to ensure > 95 % cellular proteins are labelled) (Ong, Blagoev et al. 2002). Several SILAC labelling media recipes exist, the most widely used include labelled arginine and lysine residues, due to the fact that trypsin cleaves polypeptide chains immediately after arginine and lysine residues. These recipes therefore ensure that every peptide generated by in-gel trypsin digestion, prior to MS analysis, contains at least one labelled amino acid. All tryptic peptides can therefore be distinguished from their functionally identical sister peptides, derived from a different cell population. Exceptions to trypsin cleaving polypeptide chains immediately after arginine or lysine residues include when either proline follows, or aspartic or glutamic acid residues neighbour, arginine or lysine, as this blocks or impedes trypsin cleavage respectively (Thiede, Lamer et al. 2000).

In this study the following SILAC media recipes were used: ‘medium’ (R6K4) media which contains ^{13}C-labelled arginine and 2D- (deuterium – a heavy form of hydrogen) labelled lysine, and ‘light’ (R0K0) media, which is label free. This enabled two different cell populations to be compared simultaneously.

When analysed by MS, pairs of functionally identical peptides, of different isotope composition, can be differentiated as the labelled amino acids will induce a shift in the mass/charge ratio (m/z) in comparison to the unlabelled peptides.

Comparison of the intensity of the labelled and unlabelled m/z peaks provides accurate quantitation of the relative abundance of labelled and unlabelled peptides present in the sample. This results in every identified protein being assigned a SILAC ratio, which reflects the relative abundance of that protein (derived from a labelled cell population) compared to the same protein derived from a control (unlabelled cell population) (Ong, Blagoev et al. 2002).

When applied to IP experiments in this study the SILAC ratio provides a guide to the specificity of the potential interaction; a low ratio indicates the identified protein immunoprecipitated in almost equal abundance with the control sample (EGFP in this study) and the protein of interest (EGFP-CCHFV N in this study), therefore this protein most likely bound non-specifically to EGFP, or components of the GFP-Trap. As the SILAC ratio of EGFP-CCHFV N/EGFP increases, it becomes increasingly likely that the cellular protein immunoprecipitated specifically with EGFP-CCHFV N.

In this study IPs were performed using a strategy known as mix after precipitation (MAP) (Figure 4.1). In MAP IPs, following cell lysis, separate IPs are carried out on each cell population, and equal volumes of resulting immunoprecipitates are mixed together. Proteins of interest and co-immunoprecipitating proteins present in the immunoprecipitate are then identified and quantified by MS (Figure 4.1). Using the MAP strategy, dynamic or weak interacting partners can be efficiently precipitated and quantified. However, as the two IPs are performed separately, experimental variability can be easily introduced (for example; insufficient/unequal cell lysis resulting in different amounts of lysate being incubated with GFP-Trap beads, different volumes of GFP-Trap beads being used, inaccurate pipetting when mixing the separate immunoprecipitates) which can be mistaken for biological variation. Therefore caution should be taken when preparing samples and performing IPs, and biological replicates and subsequent validation are essential.

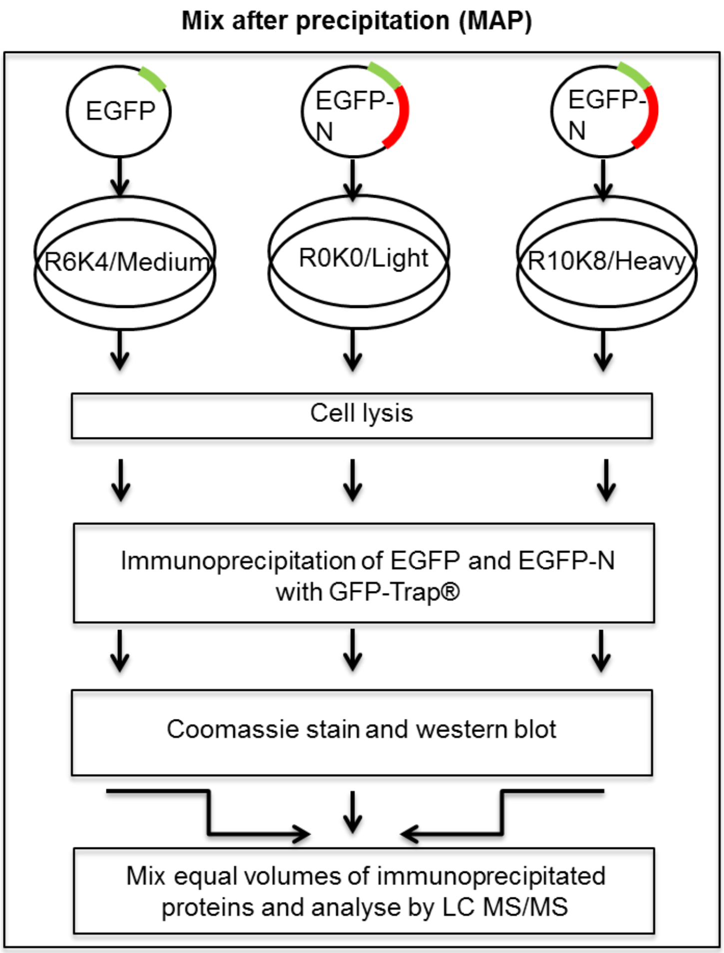


Figure 4.1 Schematic representation of SILAC labelled MAP IPs

A mix after immunoprecipitation (MAP) IP strategy was utilised in this study. An example involving 3 labelled cell populations is depicted here, however only 'Medium' and 'Light' populations were utilised in this study.

4.2 RESULTS: CELLULAR PROTEINS THAT CO-IP WITH EGFP-CCHFV N

4.2.1 Cloning of CCHFV N ORF into pEGFP-C2 expression vector

The CCHFV N ORF was originally obtained in a pCAGGs expression vector (Figure 4.8). The CCHFV N fragment was subcloned by PCR amplification of the CCHFV N ORF using primers to incorporate a 5' XhoI site and 3' BamHI restriction endonuclease site. Subcloning was completed by direct digestion of this PCR product, and ligation into the empty double digested pEGFP-C2 vector using XhoI and BamHI restriction enzymes; the correct orientation and sequence of the insert was then confirmed by sequence analysis. This construct will be referred to as pEGFP-C2-CCHFV N and resulted in the expression of an N terminally EGFP tagged CCHFV N fusion protein; EGFP-CCHFV N. A schematic representation of pEGFP-C2-CCHFV N is shown in Figure 4.2 (B) and a schematic representation of the EGFP-CCHFV N fusion protein in Figure 4.2 (A).

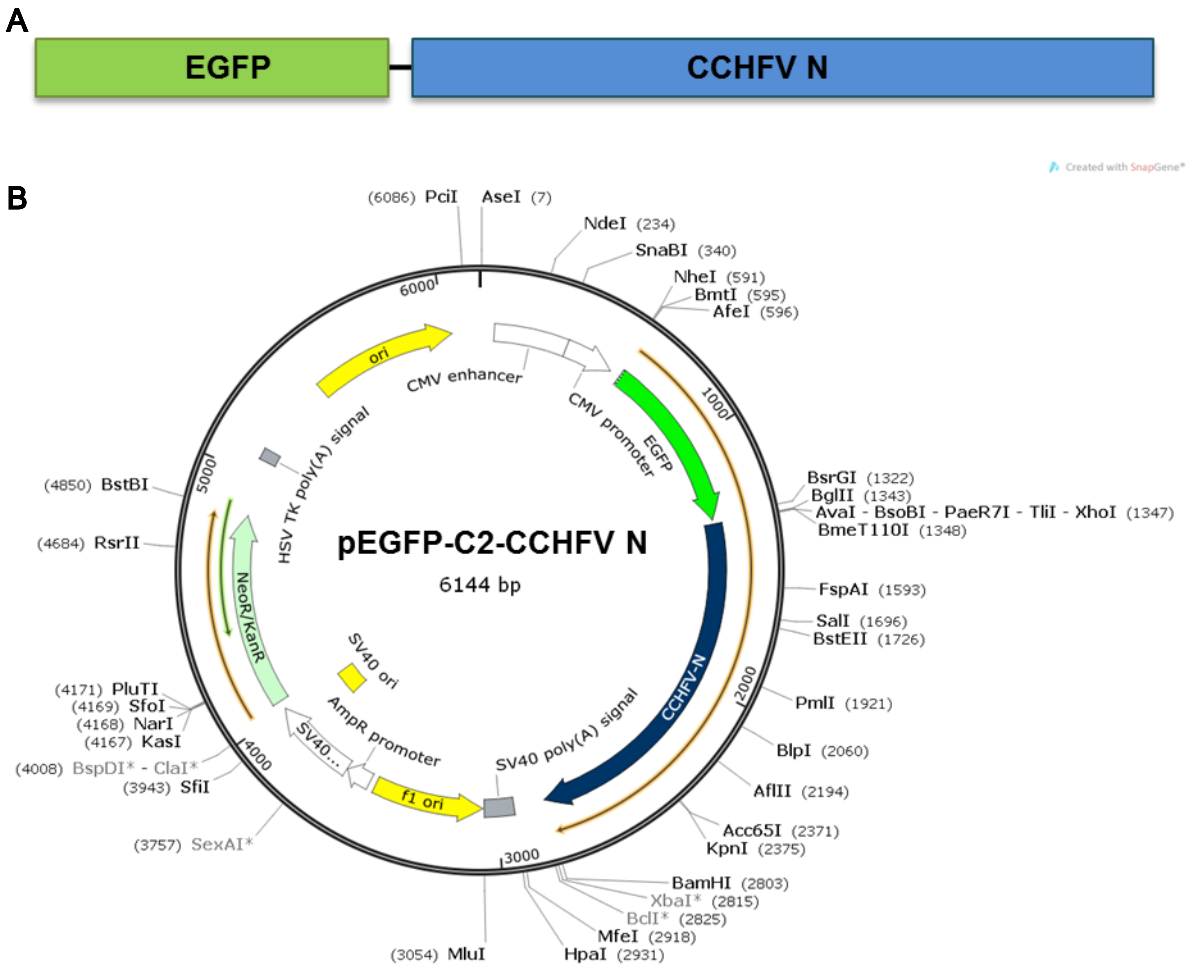


Figure 4.2 Schematic representations of EGFP-CCHFV N and pEGFP-C2-CCHFV N

CCHFV N was fused at the N terminus to EGFP (A), which enabled direct visualisation of EGFP-CCHFV N by fluorescence microscopy, and detection of EGFP-CCHFV N by western blot analysis with anti-EGFP antibodies. The map was generated using SnapGene.

The CMV promoter in pEGFP-C2-CCHFV N enables strong, constitutive expression of EGFP-CCHFV N in many mammalian cell types, and the SV40 poly(A) signal increases mRNA stability, resulting in high level protein expression. The N-terminal EGFP tag allowed direct visualisation of EGFP-CCHFV N in transfected cells by fluorescence microscopy. EGFP-CCHFV N could also be detected by commercially available antibodies that bind EGFP, when EGFP-CCHFV N expression was analysed by western blotting.

4.2.2 EGFP-CCHFV N expression

EGFP-CCHFV N was expressed in HEK293T (293T) cells following calcium phosphate mediated transfection. 293T cells are a human kidney epithelial cell line that constitutively expresses the SV40 large T antigen, and were chosen due to their human origin, which facilitates downstream MS analysis, and their high transfectability. Calcium phosphate-mediated transfection is based on the formation of a calcium phosphate-DNA precipitate, as a solution of calcium and DNA is slowly added drop-wise to a solution containing phosphate; calcium phosphate facilitates binding of DNA to the cell surface, where it then enters the cell by endocytosis. Following cell entry, EGFP-CCHFV N was expressed as a fusion protein of the correct predicted molecular weight, 87 kDa, and localised in the cytoplasm of transfected cells (Figure 4.3).

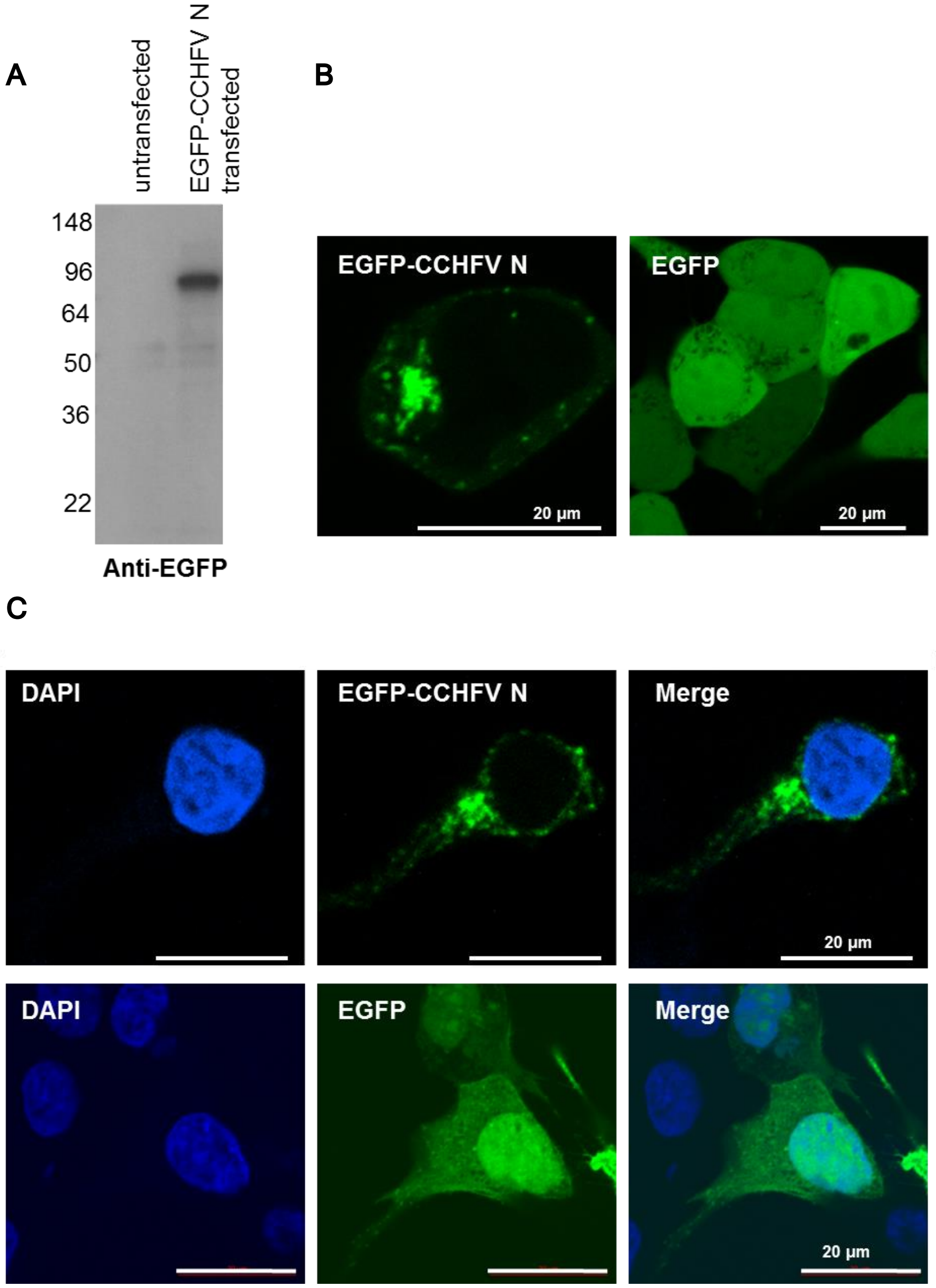


Figure 4.3 Expression of EGFP-CCHFV N in 293T cells.

pEGFP-C2-CCHFV N transfection of 293T cells resulted in the expression of an EGFP-CCHFV N fusion protein of the correct molecular weight, 87 kDa, when analysed by western blot (A). Live cell imaging confirmed the expression of EGFP-CCHFV N (B), as did confocal microscopy of formaldehyde-fixed cells (C). In all images EGFP-CCHFV N or EGFP is green, and in (C) the nucleus is stained blue with DAPI.

293T cells transfected with pEGFP-C2-CCHFV N were analysed by western blot 24 hrs after transfection; cells were lysed for 30 minutes on ice, then lysates were clarified by centrifugation, and protein content quantified by BCA assay. Protein samples were denatured, then 2 µg protein lysate was loaded per well, and cellular proteins were resolved by SDS PAGE. Proteins were transferred to polyvinylidene difluoride (PVDF) membranes, which were blocked and probed with primary rabbit polyclonal anti-EGFP antibodies. Secondary anti-rabbit HRP-conjugated antibodies were used to detect primary anti-EGFP, and were then visualised after the addition of a chemiluminescent substrate (ECL) and exposure to light-sensitive film. EGFP-CCHFV N could be detected as an 87 kDa fusion protein of the correct predicted molecular weight (Figure 4.3 (A)).

The subcellular distribution of EGFP-CCHFV N was observed by both confocal microscopy analysis of 293T cells that were fixed 24 hrs after transfection, and by live cell imaging. The cell membranes of formaldehyde fixed cells (C) were permeabilised, and the nucleus stained with 4',6-diamidino-2-phenylindole (DAPI), a blue fluorescent probe which, upon selective binding to the minor groove of dsDNA, fluoresces brightly when excited with light of wavelength ~360 nm. 293T cells were grown on glass coverslips prior to transfection and fixation, and following permeabilisation, were mounted onto glass slides using a mounting fluid (Vectorshield) containing DAPI. Confocal images were captured on a Zeiss 510 META laser scanning microscope (LSM).

Transfection of 293T cells with pEGFP-C2 was used as a control to visualise the subcellular distribution of native EGFP, in the absence of the CCHFV N moiety. The subcellular distribution EGFP was diffuse within both the cytoplasm and nucleus of 293T cells (Figure 4.3 (B) and (C)). In contrast, EGFP-CCHFV N localised in the cytoplasm of transfected cells, and was often found in perinuclear regions, but could also be visualised in other areas of the cytoplasm, often in discrete puncta. This was similar to the previously described subcellular distribution of CCHFV N both during infection, and transfection of CCHFV N in the absence of other viral proteins or RNA (Andersson, Simon et al. 2004), where CCHFV N was observed in the cytoplasm, often in perinuclear regions. It is thought trafficking of CCHFV N to these perinuclear regions is dependent on an intact actin cytoskeleton, and occurs over time (up to 8 hrs following infection); the initial cytoplasmic distribution of CCHFV N during infection is more disperse (Andersson, Simon et al. 2004).

4.2.3 IP of EGFP-CCHFV N

After transfection of pEGFP-C2-CCHFV N or pEGFP-C2, and expression of EGFP-CCHFV N or EGFP, EGFP-CCHFV N and EGFP were immunoprecipitated (IP'd) from 293T cells using a GFP-Trap. GFP-Traps are composed of agarose beads covalently conjugated to camelid antibodies (only contain the antibody 'heavy' chain; GFP-Traps only contain the variable domains of these antibodies to reduce non-specific binding) that specifically recognise and bind to EGFP. The variable domain of anti-EGFP antibodies within GFP traps bind strongly to EGFP with high specificity, enabling the IP of sufficient quantities of EGFP-CCHFV N to allow identification and quantification of cellular proteins that co-IP with EGFP-CCHFV N by MS.

293T cells were grown in 'light' (R0K0), or 'medium' (R6K4) SILAC media for two weeks prior to transfection to allow sufficient SILAC labelling, as described in section 2.2.4.3. 'Light' (R0K0) cells were transfected with pEGFP-C2-CCHFV N, and 'medium' (R6K4) cells were transfected with pEGFP-C2. 24 hrs later, two separate GFP-Trap based IPs were carried out, and EGFP or EGFP-CCHFV N was IP'd from labelled cells (Figure 4.4). Briefly; cells were lysed for 30 minutes on ice, then protein lysates were clarified by centrifugation. Clarified lysates were incubated with GFP-Trap beads for 2 hrs to allow EGFP or EGFP-CCHFV N to bind, then, after several wash steps, proteins were eluted from the GFP-Trap beads by heating at 95 °C for 10 minutes in the presence of 2x LDS-sample buffer. The resulting immunoprecipitates contained EGFP or EGFP-CCHFV N, as well as cellular proteins that immunoprecipitated with them. Equal volumes of these immunoprecipitates were mixed together and the proteins identified and quantified by MS.

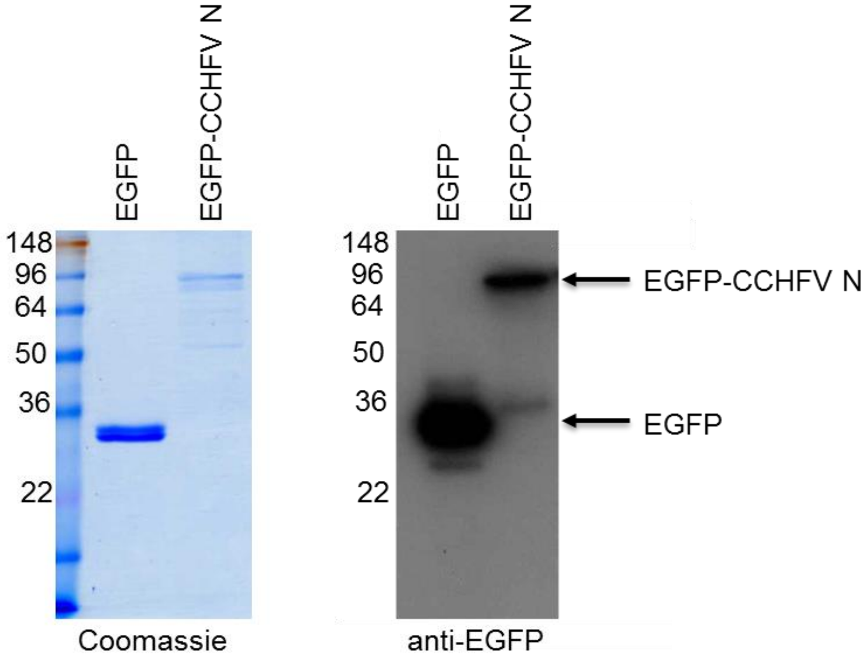


Figure 4.4 IP of EGFP and EGFP-CCHFV N from 293T cells

SDS PAGE (left) and western blot analysis (right) of GFP-Trap based IPs confirmed the immunoprecipitation of EGFP or EGFP-CCHFV N from SILAC labelled cells.

Western blot analysis with anti-EGFP confirmed the IP of EGFP or EGFP-CCHFV N from transfected cells (Figure 4.4, right). SDS PAGE analysis followed by Coomassie staining confirmed the co-IP of cellular proteins; faint bands are present in the EGFP-CCHFV N IP lanes in Figure 4.4 (left), which represented cellular protein that immunoprecipitated with EGFP-CCHFV N. Generally, the ability to see proteins that co-IP with EGFP-CCHFV N by SDS PAGE analysis corresponds well with the ability to identify and quantify these cellular proteins by MS.

IP of EGFP from pEGFP-C2 transfected cells was used as a control to identify and quantify cellular proteins that bound non-specifically to EGFP or to components of the GFP-Trap (the camelid antibodies, or the agarose beads) during the IP. This enabled differentiation between proteins that bound non-specifically to EGFP, or components of the GFP-Trap, and those that bound specifically to EGFP-CCHFV N, when immunoprecipitates were combined and analysed by SILAC-based MS.

4.2.4 Identification and quantification of immunoprecipitated cellular proteins by MS analysis

Cellular proteins that immunoprecipitated with EGFP or EGFP-CCHFV N were identified and quantified by SILAC based MS analysis. MS analysis was performed by Dundee Cell Products, and was carried out twice on the same sample; these methods are described in section 2.2.8.

MS identification and quantification of immunoprecipitated proteins resulted in the generation of a unique SILAC ratio for each identified protein (EGFP-CCHFV N/EGFP); this was a measure of the relative abundance of a protein that immunoprecipitated with EGFP-CCHFV N, compared to EGFP. For all identified proteins, the ratio of its abundance that co-immunoprecipitated with EGFP-CCHFV N compared to EGFP could be used as a guide to the specificity of its interaction with EGFP-CCHFV N. Although the same MAP IP sample was analysed twice, the results of each analysis varied slightly; between the two analyses a total of 192 proteins were identified, 112 proteins were common to both analyses, 46 proteins unique to the first analysis and 34 proteins unique to the second. This variability between MS runs emphasised the need to validate potential interacting partners by independent methods, such as western blot analysis. The results of the two MAP analyses in this study are referred to as MAP^1 and MAP^2.

A protein ratio frequency graph was generated using the SILAC ratios in order to differentiate between proteins that bound non-specifically to components of the GFP-Trap, or EGFP, and those that bound specifically to EGFP-CCHFV N (Figure 4.5). Previous studies have suggested that the majority of proteins identified by mass spectrometry in SILAC-based IP experiments represent proteins that bind non-specifically to EGFP, or components of the GFP-Trap (Trinkle-Mulcahy, Boulon et al. 2008). Examination of the EGFP-CCHFV N/EGFP ratio for the two MAP datasets (Figure 4.5) indicated that the majority of proteins exhibited fewer than a 6-fold difference, therefore a 6-fold ratio or below was chosen as a cut-off threshold for non-specific interacting proteins in this study. Cellular proteins exhibiting ratios of 6 or greater were considered to potentially represent specific interacting partners of EGFP-CCHFV N. This 6-fold cut off value was arbitrary, and served to reduce the number of candidate interacting proteins that would be validated by western blot analysis.

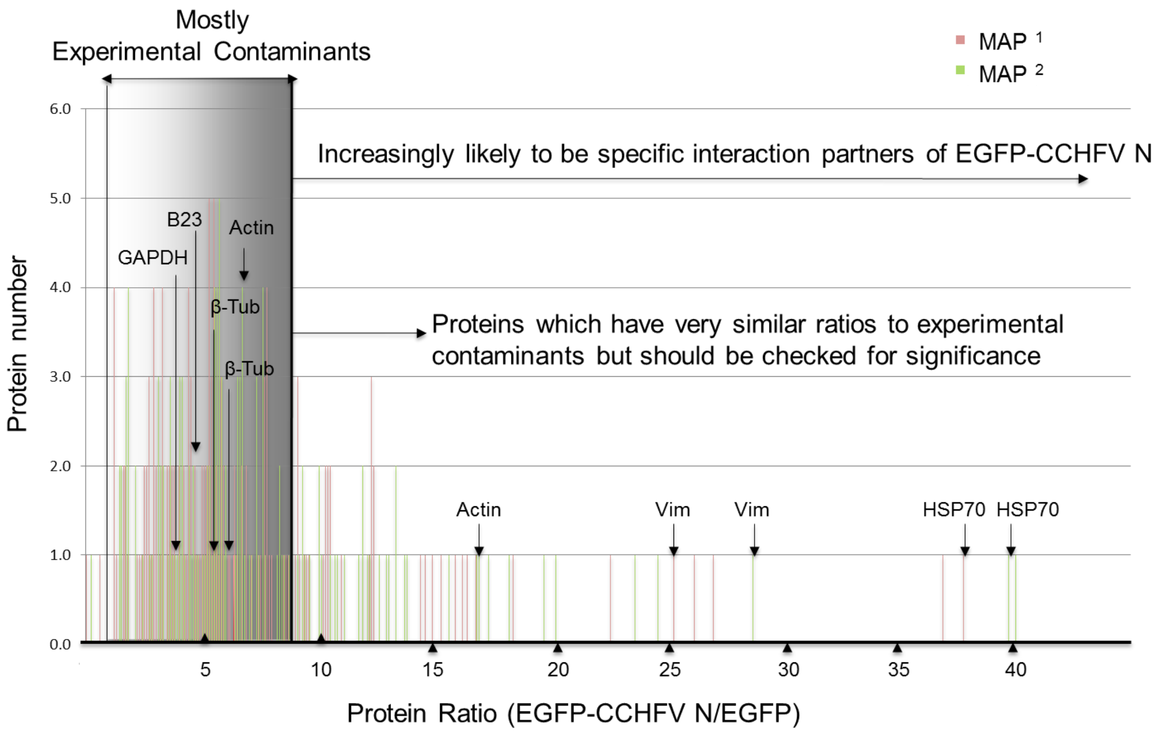


Figure 4.5 Protein ratio frequency graphs of EGFP-CCHFV N IP analysis

The protein ratio frequency graph of the MAP IPs shows the numbers of identified cellular proteins with each SILAC ratio (EGFP-CCHFV N/EGFP). In IP experiments of this type, many of the identified IP'd proteins represent experimental contaminants, which bind non-specifically to components of the GFP-Trap. An increased abundance ratio indicated an increased likelihood the protein is a specific interacting partner of EGFP-CCHFV N. In this analysis the majority of proteins were described by a normal distribution centered around a ratio of 5, and a 6-fold cut off value was chosen to distinguish between specific and nonspecific EGFP-CCHFV N interacting partners. The MS analysis was performed twice on the same sample, and the ratios derived from each analysis are represented as green and red bars. Highlighted are some proteins that were validated by western blot analysis for their ability to co-IP with EGFP-CCHFV N (Figure 4.6).

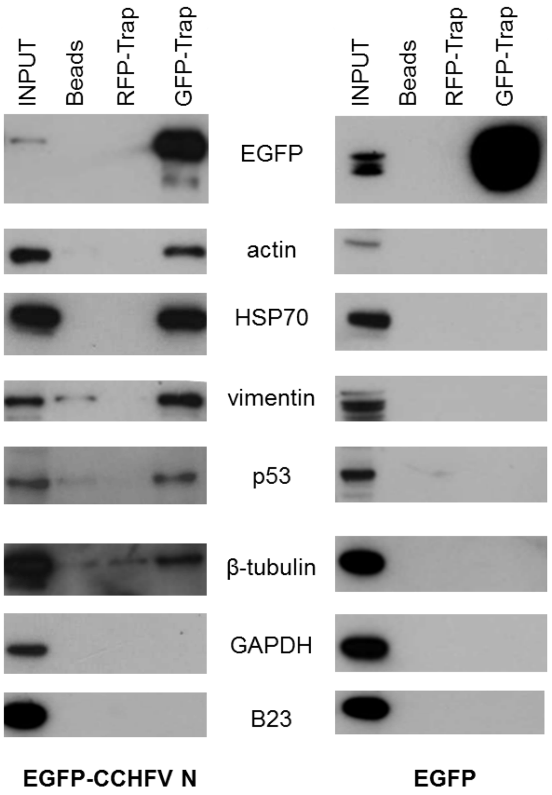
A table summarising the data from these two mass spectrometry analyses is included in Appendix II.

4.2.5 Independent validation of potential interacting partners of EGFP-CCHFV N by western blot analysis

Putative cellular interacting partners of EGFP-CCHFV N were initially identified by SILAC based MS analysis using immunoprecipitates from labelled 293T cells. IPs were repeated using lysates from unlabelled 293T cells expressing either EGFP-CCHFV N or native EGFP, in order to validate selected co-immunoprecipitated proteins (highlighted in Figure 4.5) by western blot analysis (Figure 4.6). Selected cellular proteins were chosen due to their identification in the MS analysis by multiple peptides, and their significantly greater abundance in immunoprecipitates from EGFP-CCHFV N expressing cells, compared to EGFP expressing cells.

Control IPs were performed using both uncoupled agarose beads, and an RFP-trap (single domain anti-RFP antibodies conjugated to the same agarose bead matrix as used in the GFP-trap), to verify the selected cellular proteins did not IP non-specifically with components of the GFP-trap. Western blot analysis with anti-EGFP confirmed the GFP-trap specifically IP'd EGFP-CCHFV N and EGFP, whereas the RFP-trap and the agarose bead controls did not (Figure 4.6). This confirmed the effectiveness of the GFP-trap methodology for the specific IP of EGFP and EGFP-CCHFV N.

CCHFV N has previously been shown to bind to actin (Andersson, Simon et al. 2004); this known interaction was used to assess the success of the GFP-trap and MS identification procedure, and as a positive control in western blot validation experiments. In good agreement with previous work, western blot analysis of IPs from EGFP-CCHFV N expressing cell lysates revealed actin immunoprecipitated abundantly with EGFP-CCHFV N, but was not immunoprecipitated, or was immunoprecipitated to a much lower extent, with the control RFP-trap or unbound agarose beads (Figure 4.6).



Proteins common to both analyses	UniProt accession number	SILAC ratios	
		MAP¹	MAP²
actin cytoplasmic 2	P63261	6.8	16.46
heat shock 70 kDa protein 1	P08107	37.85	40.11
heat shock cognate 71 kDa	P11142	27.11	24.72
vimentin	P08670	25.44	28.83
cellular tumor antigen p53	P04637	26.26	
tubulin beta chain	P07437	5.12	5.51
GAPDH	E7EUT4	3.95	
B23	P06748	4.53	

Figure 4.6 Independent validation of potential interacting partners of EGFP-CCHFV N by western blot analysis.

Western blot analysis (left) confirmed the specific co-IP of actin, HSP70/HSC70 (heat shock 70 kDa protein 1 (HSP70) and heat shock cognate 71 kDa (HSC70); the antibody used does not distinguish between these two proteins), vimentin, p53, and β-tubulin. GAPDH and B23 did not co-IP with EGFP-CCHFV N, supporting the 6-fold cut off value chosen for specific interacting partners. The table presenting the SILAC ratios for these proteins (right) includes the UniProt accession number, which was used to access information about the proteins (gene name, primary sequence, function) from the online UniProt database.

Cytoplasmic actin was identified in both MAP^1 and MAP^2 MS analyses by 19 and 17 peptides with increased ratios of 6.8 and 16.5 -fold, respectively. Other cellular proteins that exhibited high SILAC ratios included heat shock protein 70 (HSP70), heat shock cognate 71 kDa (HSC70), vimentin, and the tumor suppressor protein p53.

For each of these cellular proteins, western blot analysis of immunoprecipitates from EGFP-CCHFV N expressing cells detected a corresponding protein of appropriate molecular weight in both the input material and following IP with the GFP-trap, with no or much reduced binding to the unbound agarose beads, or the RFP-trap. These cellular proteins were not identified in GFP-trap immunoprecipitates from cells expressing EGFP, which strongly suggested that the interaction between each of the listed cellular proteins was specific to EGFP-CCHFV N, rather than EGFP.

In the case of HSP70, the anti-HSP70 antibody used to probe western blots also detected HSC70 (HSC70 was also assigned high SILAC ratios in both MS analyses (Figure 4.6, right)). HSP70 and HSC70 are both members of the HSP70 family of molecular chaperones; they are structurally very similar, share > 85 % sequence identity, and perform many identical functions within the cell (Liu, Daniels et al. 2012). In sections 4.4 – 4.8, this anti-HSP70 antibody was used to analyse IPs, and the identified protein is referred to as HSP70 for clarity.

GAPDH and nucleophosmin (B23.1/NPM1) were identified in the MAP MS analyses with low EGFP-CCHFV N/EGFP SILAC ratios of 3.95 and 4.5, respectively (Figure 4.6). As described above, previous work by others has indicated these ratios likely represent proteins that interacted non-specifically with EGFP, or components of the GFP-Trap (Boulon, Ahmad et al. 2010). In agreement with this, western blot analysis failed to detect the presence of either of these proteins in the GFP trap immunoprecipitate from EGFP-CCHFV N expressing cells (Figure 4.6), indicating these proteins likely bound non-specifically to EGFP or components of the GFP-Trap.

β-tubulin also exhibited relatively low ratios of 5.1 and 5.5 in the two MAP MS analyses. However, β-tubulin was detected in GFP trap immunoprecipitates from EGFP-CCHFV N transfected cells, although with much reduced abundance compared to the amount of tubulin present in the input cell lysate. This was in contrast to HSP70, p53, and vimentin, which exhibited relatively high abundances in EGFP-CCHFV N IPs; this suggested β-tubulin may represent a weak or transient interacting protein.

4.3 SECTION SUMMARY

This section has presented the identification of cellular proteins that interact with EGFP-CCHFV N, through the use of SILAC-based MS analysis of GFP-Trap IPs. IP of EGFP-CCHFV N and EGFP from different SILAC labelled cell populations allowed the discrimination between cellular proteins that interacted specifically with EGFP-CCHFV N, and those that bound non-specifically to EGFP or components of the GFP-Trap. The MS analysis was performed twice on the same sample, and a total of 192 proteins were identified, 112 proteins were common to both analyses, 46 proteins unique to the first analysis and 34 proteins unique to the second.

Each identified protein was assigned a SILAC ratio, which acted as a guide to the specificity of its interaction with EGFP-CCHFV N; the higher the ratio, the more likely the identified protein was a specific interacting partner of EGFP-CCHFV N. Due to the majority of proteins having a ratio of 6 or below, a ratio of 6 was arbitrarily chosen as the cut-off for specific interacting partners; proteins with a ratio above 6 were considered to potentially be interacting partners of EGFP-CCHFV N, and those with a ratio below 6 were considered to be non-specific. The cellular protein actin, a known interacting partner of CCHFV N (Andersson, Simon et al. 2004) was used as a positive control to assess the success of the GFP-Trap and identification in the MS procedure. In good agreement with previous work, actin bound specifically to EGFP-CCHFV N (Figure 4.6), and was identified in the MS analysis with a ratio of 16.46 and 6.8. Independent validation of potential interacting partners of EGFP-CCHFV N was undertaken by western blot analysis, which confirmed the specific IP of several proteins (including actin) with high ratios; HSP70 (37.85, 40.11), vimentin (25.44, 28.33), and cellular tumor antigen p53 (26.26). β-tubulin had a lower ratio (5.12, 5.51), but still immunoprecipitated with EGFP-CCHFV N. GAPDH and B23 had low SILAC ratios, (3.95, 4.53 respectively) and did not co-IP with EGFP-CCHFV N.

HSP70 was chosen to be further validated as an interacting partner of untagged CCHFV N and HAZV N due to the high SILAC ratio of HSP70 (and HSC70) in both MS analyses, the abundance of HSP70 that immunoprecipitated with EGFP-CCHFV N when analysed by western blot (Figure 4.6), and evidence from the literature that HSP70 interacts with viral proteins from several different viruses and is recruited during their replication cycles (Mayer 2005). In some instances the recruitment of HSP70 is beneficial to the virus replication cycle, in others viral protein interaction with HSP70 appeared to be detrimental to virus replication (Yu, Ye et al. 2009). In either case the interaction of HSP70 with viral proteins impacted virus replication, which suggested the interaction of HSP70 with EGFP-CCHFV N could also potentially impact CCHFV replication. Experiments examining the interaction of HSP70 with un-tagged CCHFV N and HAZV N are summarised in the following sections.

4.4 RESULTS: IMMUNOPRECIPITATION OF HSP70 WITH CCHFV N

IPs of un-tagged CCHFV N were carried out to confirm that the co-IP of HSP70 with EGFP-CCHFV N was not an artefact of overexpressing the EGFP-CCHFV N fusion protein.

4.4.1 Validation of anti-CCHFV N antibodies derived from sheep serum

Antibodies targeting CCHFV N were generated using the same method described for the anti-HAZV N antibodies (section 3.2.4). Briefly, recombinant CCHFV N was purified from *E. coli* (Carter, Barr et al. 2012) and 1 mg >95 % pure CCHFV N was sent to Altabioscience. A sheep immunisation programme compliant with UK home office regulations was undertaken, which resulted in 3 succesive serum bleeds containing antibodies recognising CCHFV N. The ability of anti-CCHFV N antibodies present in the 3rd bleed serum to recognise CCHFV N expressed from several different sources was assayed by western blot analysis (Figure 4.7) using the same methodology described in section 4.2.2. The serum will be refered to as anti-CCHFV N.

Anti-CCHFV N specifically recognised CCHFV N from several different sources; purifed from *E. coli*, 293T cell expressed EGFP-CCHFV N, and untagged CCHFV N (expressed from a pCAGGS plasmid, described in section 4.4.2), whilst not cross reacting with cellular proteins in 293T whole cell lysate (Figure 4.7). This confirmed the success of the immunisation programme, and anti-CCHFV N was subsequently used to detect CCHFV N during western blot analysis, IP experiments and immunofluorescence (IF) experiments.

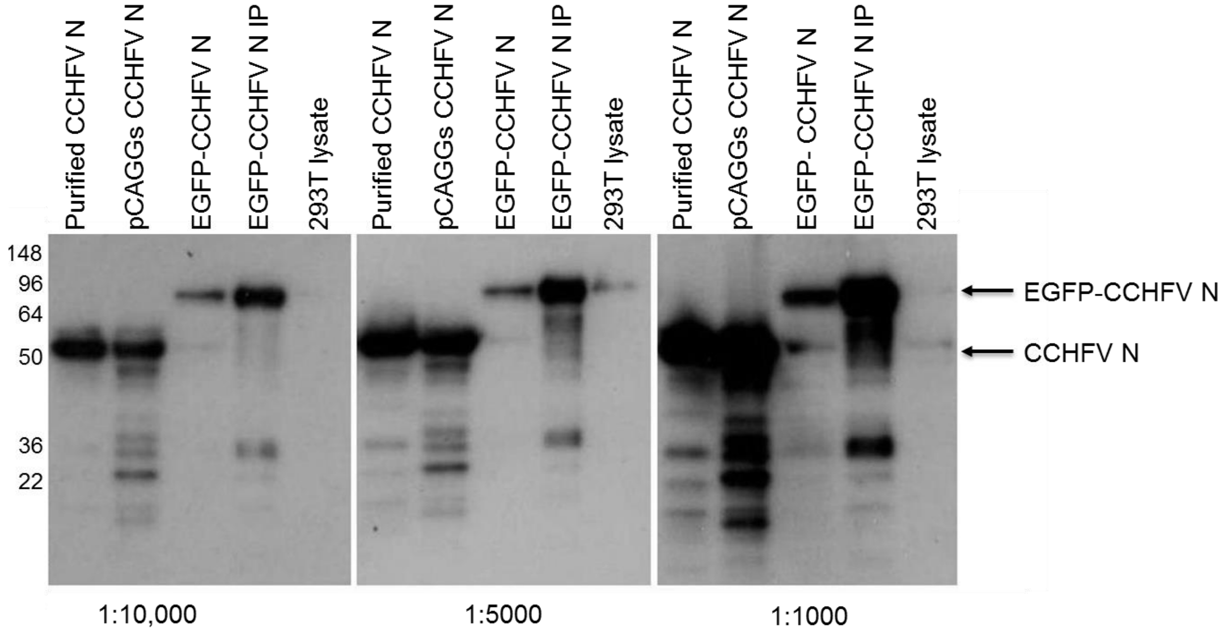


Figure 4.7 Western blot analysis of CCHFV N from diffferent sources using anti-CCHFV N.

Anti-CCHFV N specifically recognised CCHFV N purified from E. coli, expressed from a pCAGGs-CCHFV N plasmid (described section 4.4.2), EGFP-CCHFV N, and IP'd EGFP-CCHFV N. The absence of bands of corresponding molecular weight in whole cell lysate from 293T cells indicates anti-CCHFV N did not cross-react with cellular proteins when diluted 1:10,000 (v/v) in 5 % (w/v) BSA in TBST. At higher concentrations of anti-CCHFV N (1:1000-, 1:5000 (v/v) in 5 % (w/v) BSA in TBST), there were faint bands in the 293T whole cell lysate lane; these may either represent bleed-over from adjacent lanes, or cellular proteins cross-reacting. Western blot courtesy of Alex Brown.

4.4.2 CCHFV N expression

CCHFV N was expressed from the pCAGGS-CCHFV N plasmid (Figure 4.8). High level gene expression in mammalian cells is driven by the synthetic CAG promoter, and resulted in the expression of un-tagged CCHFV N.

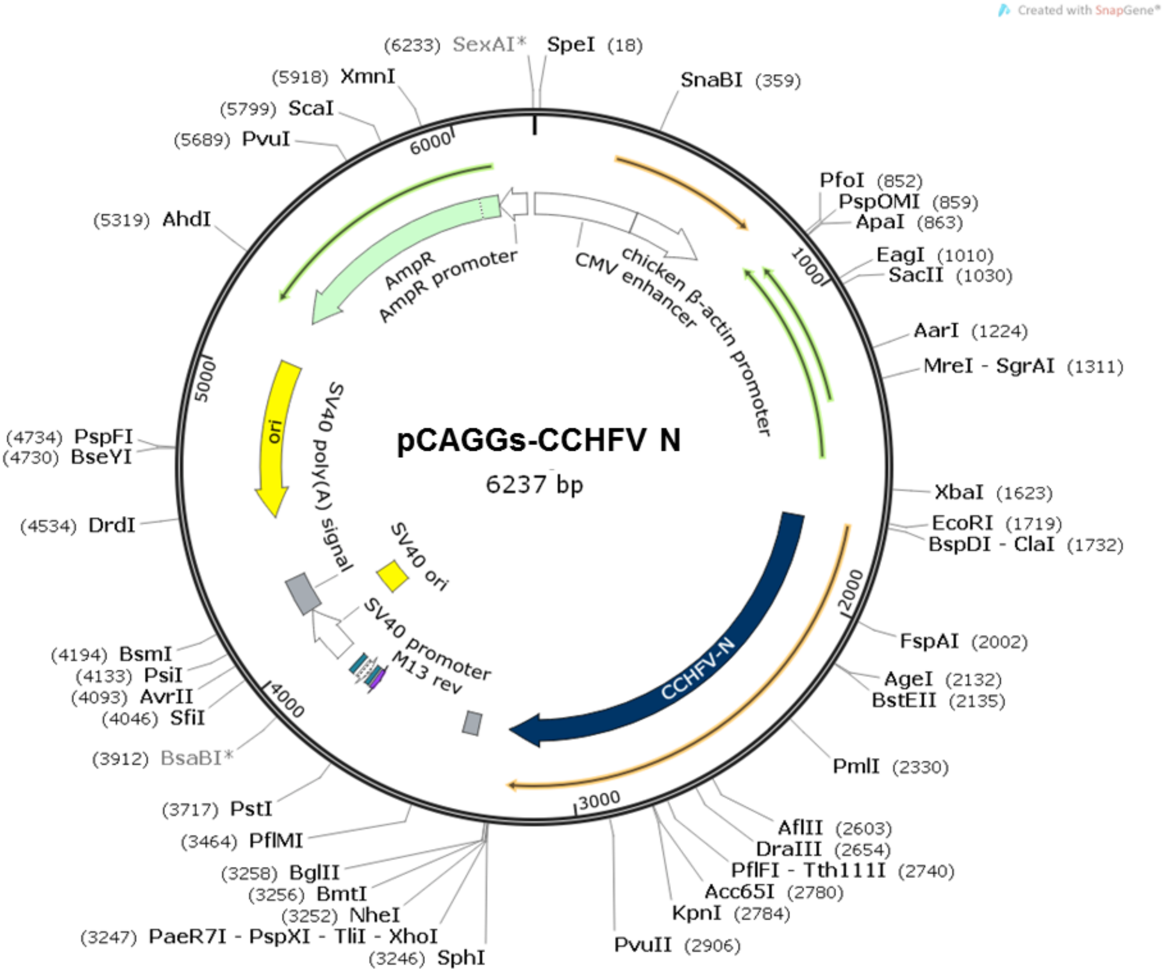


Figure 4.8 Plasmid map of the CCHFV N ORF in a pCAGGs vector

Untagged CCHFV N was expressed from pCAGGs-CCHFV N in mammalian cells. Image generated with SnapGene.

CCHFV N was expressed in 293T cells using calcium phosphate mediated transfection, and in HUH-7 cells (a hepatocyte derived carcinoma cell line, originally isolated from a liver tumor) using lipofectamine-2000 based transfection. Lipofectamine 2000, and RNAiMAX (utilised in section 4.9.1) are specially designed cationic lipids that facility entry of DNA and siRNA into cells. CCHFV N was expressed as a 53 kDa protein (Figure 4.9 (A)), which was detected by western blot analysis (described in section 4.2.2) and indirect IF staining (Figure 4.9 (B)) using anti-CCHFV N.

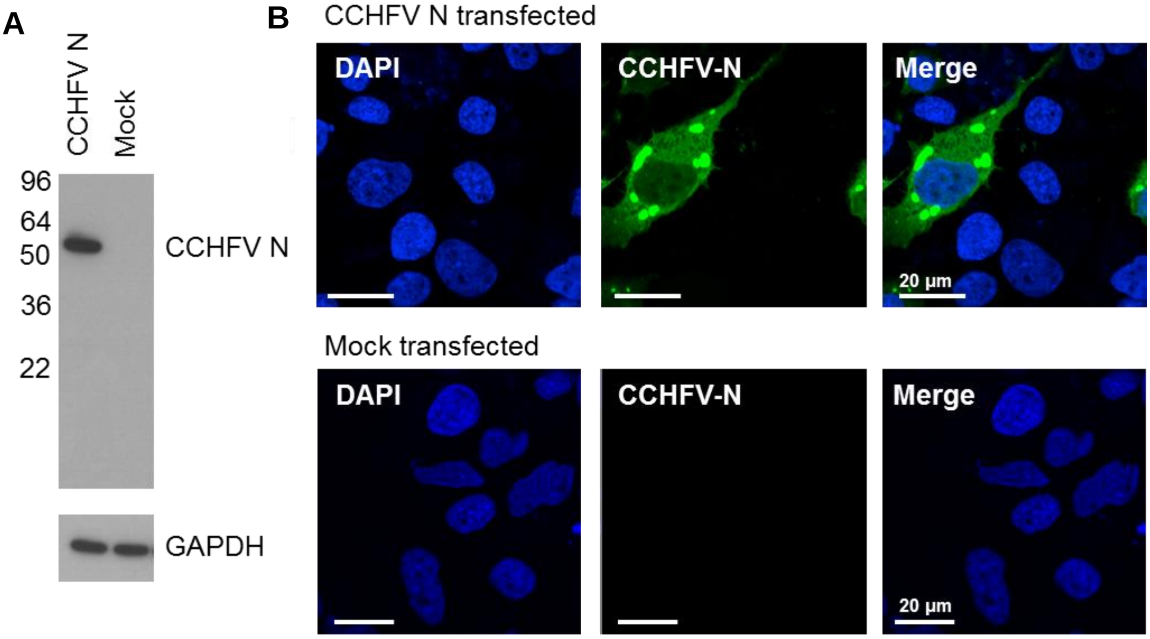


Figure 4.9 Expression of CCHFV N in 293T and HUH-7 cells

pCAGGs-CCHFV N transfection of 293T cells followed by western blot analysis resulted in detection of expressed CCHFV N of the predicted molecular weight with anti-CCHFV N (A). Indirect IF analysis of pCAGGs-CCHFV N transfected HUH-7 cells using secondary anti-sheep alexaFluor 488 conjugated antibodies to detect anti-CCHFV N (B). Anti-CCHFV N specifically detected CCHFV N (green), which was localised predominantly in the cytoplasm of transfected cells, often in perinuclear regions. In all images the nucleus is stained blue with DAPI.

HUH-7 cells were chosen to perform indirect IF analysis due to the larger diameter of their cytoplasm compared to 293T cells. Briefly; HUH-7 cells were grown on coverslips, and 24 hrs after transfection with pCAGGs-CCHFV N, cells were fixed, cell membranes permeabilised, and CCHFV N detected with anti-CCHFV N. Anti-CCHFV N bound to CCHFV N was then detected with alexaFluor 488 conjugated secondary anti-sheep antibodies and coverslips were mounted onto glass slides in ProLong Gold containing DAPI, to allow visualisation of cell nuclei. Confocal images were captured on a Zeiss 510 META laser scanning microscope (LSM). In good agreement with previous work the subcellular distribution of CCHFV N was cytoplasmic, and CCHFV N often localised in discrete perinuclear regions (Figure 4.9 (B)) (Andersson, Simon et al. 2004; Bergeron, Albarino et al. 2010).

Expression of CCHFV N in the context of actively replicating virus was also analysed by western blot and indirect IF using anti-CCHFV N (Figure 4.10). Infection of SW13 cells with actively replicating CCHFV was carried out at Porton Down by Dr Stuart Dowall of Public Health England. SW13 cells are a human adrenal cortex carcinoma cell line, originally derived from a small cell carcinoma tissue biopsy. 48 hrs after CCHFV infection, cells were fixed overnight, then indirect IF analysis was carried out as previously described. The sub-cellular distribution of CCHFV N in the context of actively replicating CCHFV was similar to that of CCHFV N expressed from pCAGGs-CCHFV N; CCHFV N was cytoplasmic, and was often observed in discrete perinuclear regions (Figure 4.10 (B)). Lysis of CCHFV infected SW13 cells, and western blot analysis confirmed the specific recognition of CCHFV N (53 kDa) by anti-CCHFV N (Figure 4.10 (A)).

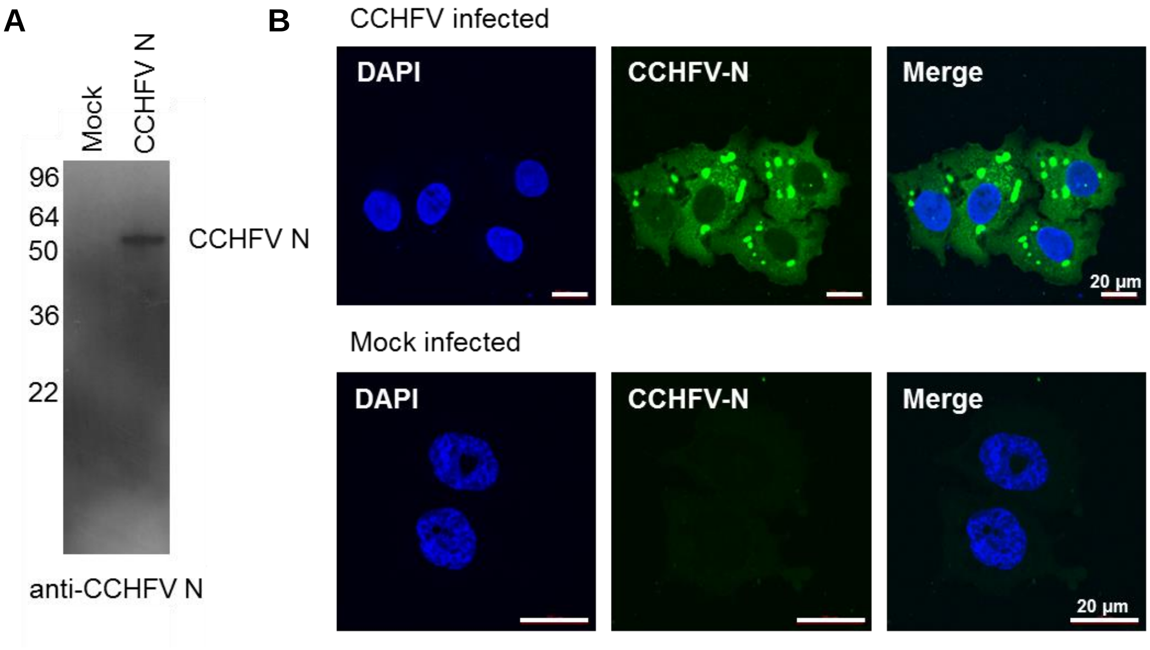


Figure 4.10. Western blot and IF analysis of CCHFV N in the context of actively replicating CCHFV.

During CCHFV infection, CCHFV N was expressed as a 53 kDa protein that was detectable by western blot analysis (A) using anti-CCHFV N. (B) The subcellular localisation of CCHFV N during infection was cytoplasmic; CCHFV N was often observed in discrete perinuclear regions, similar to the subcellular distribution of CCHFV N expressed from pCAGGs-CCHFV N.

The similarity between the subcellular distribution of CCHFV N expressed in the context of viral infection, and expressed from pCAGGs-CCHFV N, suggested that CCHFV N might be trafficked in a similar way in transfected- and infected-cells. This subcellular distribution of CCHFV N varied slightly from the subcellular distribution of EGFP-CCHFV N. Although both EGFP-CCHFV N and CCHFV N were cytoplasmic, EGFP-CCHFV N often localised in a single discrete region adjacent to the nucleus, whereas CCHFV N localised in several areas in the cytoplasm both adjacent to the nucleus, and in other regions. In both cases several smaller, discrete puncta in other areas of the cytoplasm were also observed, in addition to the larger accumulations of EGFP-CCHFV N or CCHFV N.

4.4.3 Immunoprecipitation of CCHFV N from transfected and infected cells

IPs were repeated using CCHFV N instead of EGFP-CCHFV N, to confirm that co-IP of HSP70 was not an artefact of the EGFP-CCHFV N fusion protein. CCHFV N was IP'd from both pCAGGs-CCHFV N transfected cells, and from CCHFV infected cells, using anti-CCHFV N conjugated to Protein G-coupled Dynabeads (Figure 4.11). Dynabeads are uniform, 2.8 µm superparamagnetic beads with recombinant protein G (a bacterial protein that binds antibodies (in this case anti-CCHFV N) via their Fc region) covalently conjugated to their surface. The use of Dynabeads instead of agarose/sepharose slurry enabled fast, gentle, isolation of protein complexes without time-consuming and potentially physically damaging centrifugation steps. Briefly; Dynabeads were incubated with anti-CCHFV N to allow antibodies to bind, then anti-CCHFV N bound Dynabeads were incubated with cell lysates either from pCAGGs-CCHFV N transfected cells, or CCHFV infected cells, to enable anti-CCHFV N to bind CCHFV N. After washing Dynabeads with antigen-antibody complexes bound, proteins were eluted by heating at 95 °C in 2 x LDS buffer. Anti-CCHFV N bound Dynabeads were incubated with mock transfected/mock infected cell lysates, and un-bound Dynabeads were incubated with pCAGGs-CCHFV N transfected, or CCHFV infected cell lysates, to control for cellular proteins binding non-specifically to anti-CCHFV N or the Dynabeads.

CCHFV N was IP'd when anti-CCHFV N bound Dynabeads were incubated with lysates from pCAGGs-CCHFV N expressing, and CCHFV infected, cells (Figure 4.11, boxed lanes). HSP70 co-immunoprecipitated with CCHFV N, both in the context of virus infection (Figure 4.11 (B)), and when CCHFV N was expressed from pCAGGs-CCHFV N (Figure 4.11 (A)). HSP70 was absent from, or present to a much lesser extent in, control lanes, suggesting HSP70 immunoprecipitated specifically with

CCHFV-N from both CCHFV infected- and pCAGGs-CCHFV N transfected-cells. The presence of small amounts of HSP70 in the mock anti-CCHFV N lanes could be due to the fact that anti-CCHFV N does not represent purified antibodies; the anti-CCHFV N antibodies are utilised within a mixture of other proteins in sheep serum; some of which could bind non-specifically to Dynabeads and cellular proteins (such as HSP70) in cell lysates.

B23 and GAPDH did not co-IP with EGFP-CCHFV N, when results of the MS analysis were validated by western blot (Figure 4.6). B23 and GAPDH had low SILAC ratios (3.95 and 4.24) and their presence in the MS analysis was thought to be due to non-specific binding to EGFP or components of the GFP-Trap during IP of EGFP-CCHFV N. In good agreement with these observations, neither B23 nor GAPDH immunoprecipitated with CCHFV N, but HSP70 did co-IP with CCHFV N (Figure 4.11 (A), boxed lane), further validating the MS analysis in section 4.2, and the choice of a 6-fold cut off value for cellular proteins considered to be specific interacting partners of EGFP-CCHFV N.

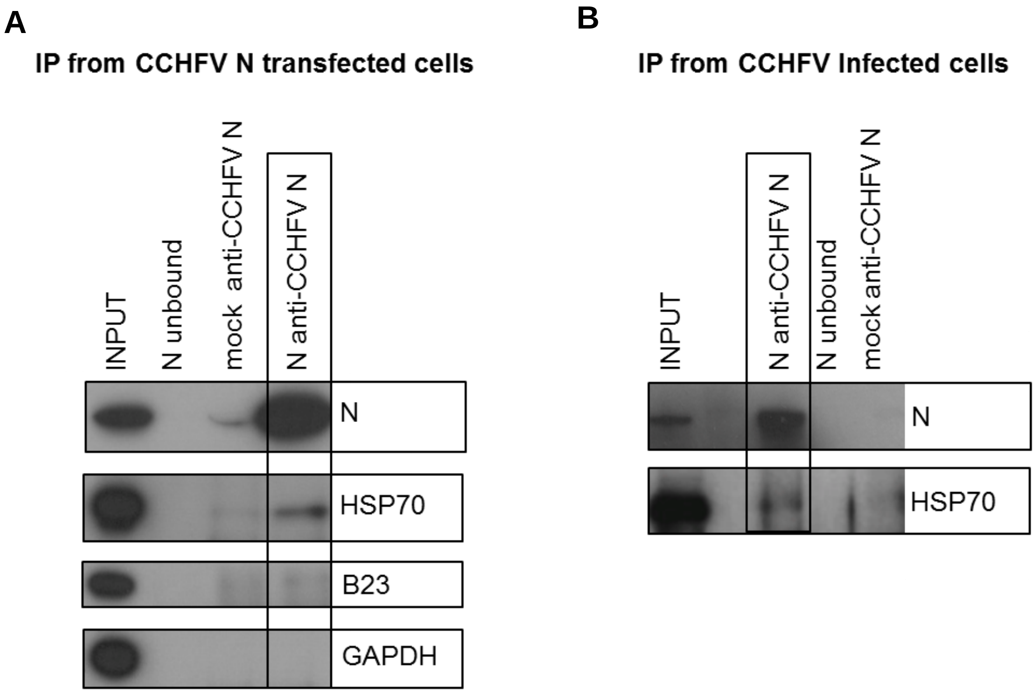


Figure 4.11. Co-immunoprecipitation of HSP70 with CCHFV N from pCAGGs transfected and CCHFV infected cells.

Anti-CCHFV N was used to IP CCHFV N from 293T cells expressing pCAGGs-CCHFV N (A), and from CCHFV infected SW13 cells (B) using Protein G coupled Dynabeads. In both cases CCHFV N was IP'd when anti-CCHFV N was incubated with lysates from pCAGGs-CCHFV N transfected or CCHFV infected cells (lane: N anti-CCHFV N, highlighted with a box), but is absent from, or present to a much lesser extent in, control lanes (lanes: N unbound, and mock anti-CCHFV N). HSP70 immunoprecipitated with CCHFV N from both pCAGGs-CCHFV N transfected and CCHFV infected cells, and is absent from, or present to a much lesser extent in, control lanes (A and B). B23 and GAPDH do not co-IP with CCHFV N (A), further validating MS results from section 4.2.

4.5 SECTION SUMMARY

This section presents the co-IP of HSP70 with CCHFV N, both in the context of actively replicating CCHFV, and when CCHFV N is expressed from pCAGGs-CCHFV N. Expression of CCHFV N in mammalian cells from pCAGGs-CCHFV N and during CCHFV infection was detected by western blot and indirect IF analysis using anti-

CCHFV N; transfection of pCAGGs-CCHFV N resulted in the expression of CCHFV N of the correct molecular weight (53 kDa), which localised in the cytoplasm of cells, often in discrete perinuclear regions. The subcellular localisation of pCAGGs-CCHFV N expressed CCHFV-N was shown to be similar to that of CCHFV N during infection, which suggested pCAGGs-CCHFV N expressed CCHFV N was trafficked in a similar manner to CCHFV N during infection.

The specific recognition of CCHFV N by anti-CCHFV N enabled the IP of CCHFV N from both pCAGGs-CCHFV N transfected, and CCHFV infected cell lysates, using anti-CCHFV N coupled Dynabeads. HSP70 was shown to co-IP with CCHFV N from both pCAGGs-CCHFV N transfected, and CCHFV infected cell lysates. This suggested that the co-IP of HSP70 with EGFP-CCHFV N was not due to expression of the EGFP-CCHFV N fusion protein; instead HSP70 appeared to co-IP specifically with the CCHFV N moiety of EGFP-CCHFV N.

HSP70 is a member of the HSP70 family of cellular chaperones. The HSP70 family of proteins are involved in a wide variety of cellular processes, including the folding of newly synthesised polypeptides, the assembly and disassembly of oligomeric structures, the disaggregation of protein aggregates, and the refolding of stress denatured proteins, amongst other processes. During infection, viruses need to synthesise a large number of viral proteins in a relatively short space of time, and it is thought that protein folding can become a rate-limiting step (Mayer 2005). It has been proposed that most viruses therefore require cellular chaperones during their replication cycles; many viruses recruit members of the HSP70 family as they are central components of the cellular chaperone network (Mayer 2005). In addition, HSP70 family members are involved in the regulation of several cellular pathways, including the ER stress response, the cell cycle, and apoptosis. Viruses are also thought to interact with HSP70 family members in order to influence the regulation of these cellular pathways, to create a cellular environment more favourable for virus replication (Mayer 2005).

CCHFV N interacted with HSP70, however the consequences of this interaction for the host cellular environment, and for the replication cycle of CCHFV remained unknown. To further investigate this HAZV was used as a model for CCHFV infection, as it can be more easily manipulated under containment level 2 conditions. The manipulation of HAZV, and the interaction of HAZV N with HSP70, is described in the following sections.

4.6 RESULTS: IMMUNOPRECIPITATION OF HSP70 WITH HAZV N

4.6.1 HAZV growth curves

Actively replicating HAZV was a kind gift from Dr Roger Hewson and Dr Stuart Dowall of Public Health England. HAZV was obtained as an SW13 cell culture supernatant at an infectious titre of 1.7×10^6 pfu/mL (plaque forming units per mL). All assays using HAZV were carried out in SW13 cells (defined in section 2.1.3). Conditions were optimised for HAZV infection so that a stock of HAZV could be grown for use in subsequent experiments (Figure 4.12). Briefly; SW13 cells were infected at 3 different MOIs (multiplicity of infection; an estimation of the number of virus particles available to infect each cell), MOI 0.001, MOI 0.01, and MOI 0.1. Cytopathic effects were not observed 24 hours after infection when cells were infected at MOI 0.001, MOI 0.01 or MOI 0.1, therefore virus-containing cell culture supernatants were harvested day 2, day 3, day 4 and day 5 post-infection, to determine the time point at which the most virus could be harvested (Figure 4.12 (A)).

Infectious HAZV was diluted by the appropriate factor in serum-free media (SFM) to infect SW13 cells in a 96-well plate format at MOI 0.001, MOI 0.01 and MOI 0.1. Complete media was removed from SW13 cells, which were washed once in PBS, prior to the application of HAZV-containing SFM, diluted to infect cells at the appropriate MOI. HAZV was incubated with SW13 cells for 1 hr at 37 °C with gentle rocking, to allow HAZV to adsorb to cell surfaces; HAZV containing SFM was then removed and replaced with DMEM containing 2 % FBS. HAZV was allowed to replicate in SW13 cells for 2-, 3-, 4- or 5- days; at each time point the total volume of cellular supernatant was removed and the infectious HAZV titre estimated by plaque assay, and plotted against the number of days post-infection (Figure 4.12 (A)). Figure 4.12 (B) shows a representative HAZV plaque assay.

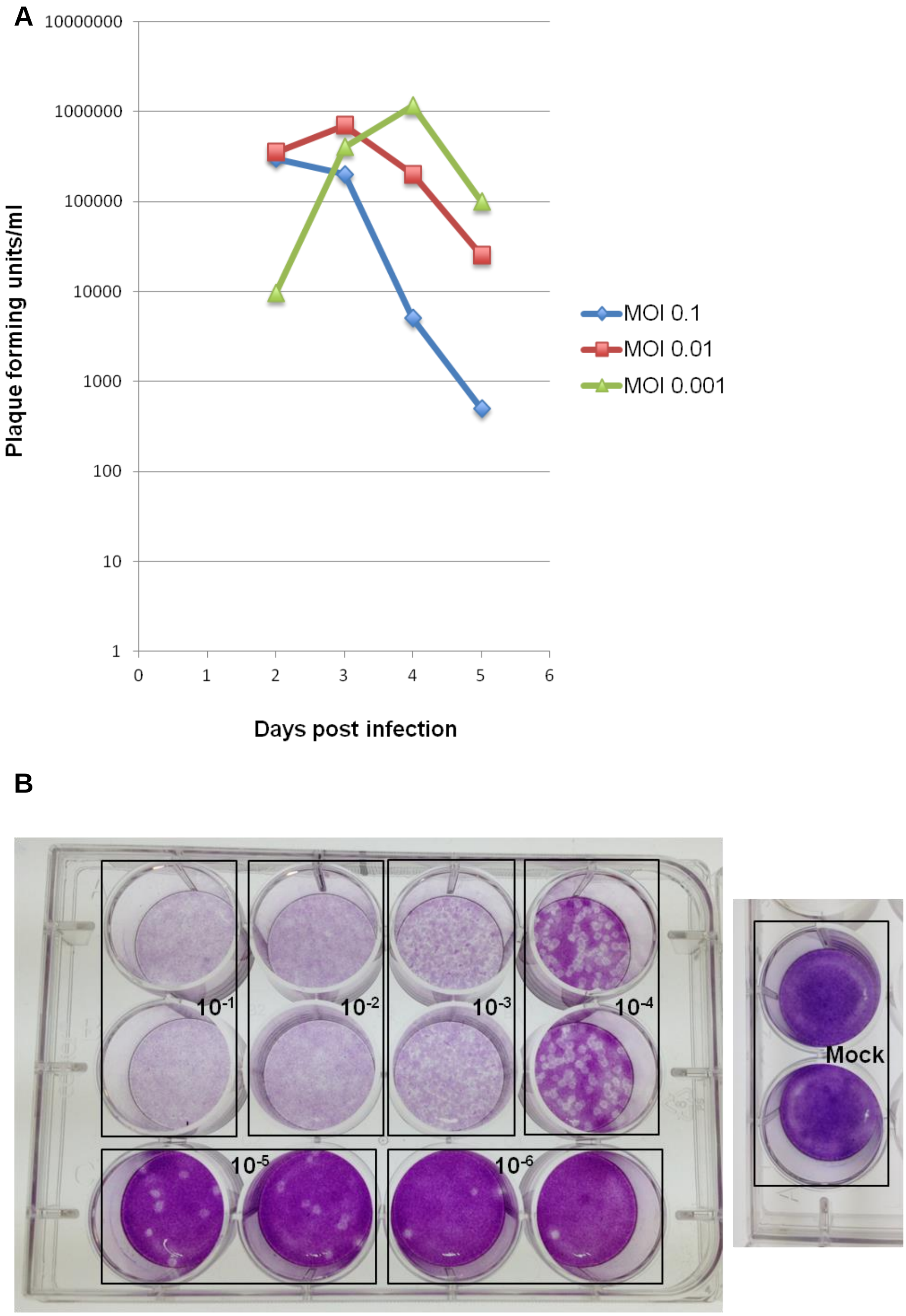


Figure 4.12 HAZV growth curve and representative HAZV plaque assay

Conditions for growing viral stocks of HAZV were determined by infecting SW13 cells at MOI 0.001, MOI 0.01, and MOI 0.1, and harvesting infectious supernatant 2-, 3-, 4- and 5- days after infection. Viral titre was estimated by plaque assay (B), and plotted against the number days of HAZV replication prior to harvesting (A). Cells that were infected at an MOI of 0.001, and the supernatant harvested 4 days later, secreted the most infectious HAZV into the supernatant.

Plaque assays were performed in duplicate in 12-well plates; 25 µL HAZV was diluted by a factor of 10 in 225 µL SFM (10^{-1} dilution), which was further serially diluted by a factor of 10 to generate 10^{-2} – 10^{-6} dilutions. These were used to infect cells as described above; after HAZV adsorption, virus was removed from cells and replaced with complete media diluted 1:1 with 1.6 % carboxy-methyl cellulose (CMC). HAZV is cytopathic, therefore the CMC/complete media overlay ensured that over the course of the assay (6 days), plaques (areas where cells were absent) were formed as HAZV infection resulted in cell death in very defined regions. After fixing cells, plaques were visualised by staining cells with crystal violet solution. The virus stock plaque assayed in Figure 4.12 (B) had an estimated titre of ~6 x 10^6 pfu/mL.

SW13 cells infected with HAZV at an MOI 0.001 secreted the most infectious virus 4 days after infection; therefore these conditions were used to grow all subsequent stocks of HAZV. After the peak titre of infectious HAZV was reached (day 4 if infected at MOI 0.001 and day 3 if infected at MOI 0.01) the titre of infectious HAZV in the supernatant declined sharply. This has previously been reported during HAZV infection (Clerx, Casals et al. 1981), however the reason why this occurs is unknown, but could be related to the stability of HAZV particles in the cell culture supernatant at 37 °C.

4.6.2 Detection of HAZV N

HAZV N was detected in HAZV infected cells with anti-HAZV N antibodies that were generated using the same sheep immunisation programme used to generate anti-CCHFV N. Section 3.2 describes the purification of >95 % pure HAZV N, which was used to inoculate sheep; antibodies targeting HAZV N are contained within the sheep serum that was harvested in 3 successive bleeds. Sheep serum containing anti-HAZV N antibodies will be referred to as anti-HAZV N.

The ability of anti-HAZV N to specifically recognise HAZV N was assayed by western blot and indirect IF analysis (Figure 4.13). Western blot analysis was carried out as follows; SW13 cells were infected with HAZV at an MOI 1 for 24 hrs then lysed for 30 minutes on ice. Following clarification of cell lysates by centrifugation, total protein was quantified by BCA assay, and 5 µg protein loaded per lane, after denaturing protein samples. Cell lysates were resolved by SDS PAGE, then proteins were transferred to FL-PVDF (fluorescent polyvinylidene difluoride) membranes. Membranes were blocked, then incubated with anti-HAZV N, which bound HAZV N. Anti-HAZV N was detected by secondary anti-sheep antibodies covalently conjugated to IRDye 800CW, a fluorescent dye that was visible in the 800 channel when analysed with the LiCor Odyssey Sa Infrared imaging system.

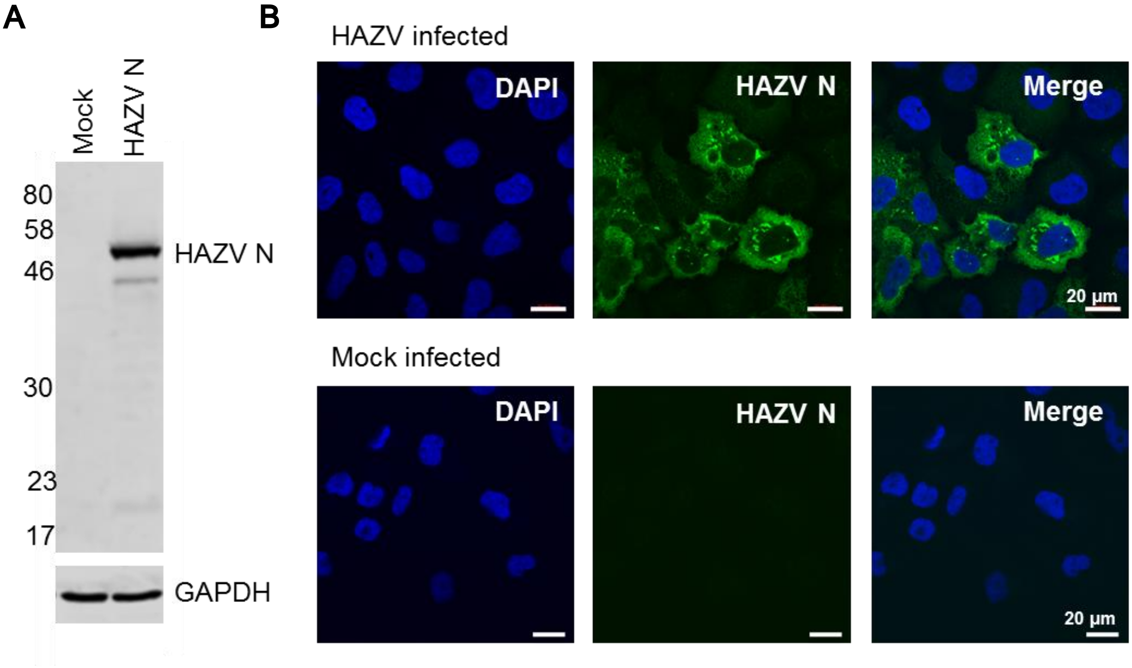


Figure 4.13 Detection of HAZV N by western blot and indirect IF analysis.

Anti-HAZV N recognised HAZV N from HAZV infected cells when analysed by western blot (A) and indirect IF (B). In (A) anti-HAZV N detected a protein of the correct predicted molecular weight (54 kDa) from HAZV infected cells when analysed by western blot. In (B) HAZV N was detected in the cytoplasm of infected cells (green); the subcellular distribution of HAZV N was similar to that of CCHFV N. The nucleus is stained blue with DAPI in (B). The absence of bands in the mock infected lanes in (A) and the absence of immunofluorescent staining in mock infected cells in (B) indicated that anti-HAZV specifically recognised HAZV N and did not cross react with cellular proteins.

Anti-HAZV N detected a protein of the predicted molecular weight of HAZV N (54 kDa) in lysates from HAZV infected cells, which was absent from lysates of mock infected cells (Figure 4.13 (A)). This indicated that anti-HAZV N specifically recognised HAZV N, and did not cross react with cellular proteins. Indirect IF analysis of HAZV infected cells was carried out 24 hrs after HAZV infection using anti-HAZV N as described in section 2.2.6.6. The subcellular distribution of HAZV N was cytoplasmic, and was similar to the subcellular distribution of CCHFV N (Figure 4.13 (B)).

4.6.3 The interaction of HSP70 with HAZV N

HAZV N and CCHFV N are structurally very similar (section 3.5.2); CCHFV and HAZV belong in the same serogroup within the same genus, and therefore share similar replication strategies. CCHFV N immunoprecipitated with HSP70, and given the structural similarity in these N proteins, and the replication strategy shared by these viruses, the ability of HAZV N to co-IP with HSP70 was also examined. Anti-HAZV N specifically recognised HAZV N synthesised during infection, and was therefore suitable to use to IP HAZV N from infected cells.

Briefly; anti-HAZV N was bound to Protein G conjugated Dynabeads, as described in section 2.2.7.2. SW13 cells were infected with HAZV for 24 hrs, then lysed for 30 minutes on ice. IP of HAZV N from infected cell lysates was carried out as described in section 2.2.7.2; lysates from mock infected cells were incubated with anti-HAZV N bound Dynabeads to control for cellular proteins binding non-specifically to anti-HAZV N, and HAZV infected lysates were incubated with unbound Dynabeads, to control for HAZV N and cellular proteins binding non-specifically to Dynabeads.

HAZV N was efficiently IP’d from infected cell lysates using this protocol (Figure 4.14 (A)). In good agreement with the CCHFV N IPs, HSP70 immunoprecipitated with HAZV N from HAZV infected cells (Figure 4.14 (A) boxed lane). HSP70 was absent from, or present to a much lesser extent in control lanes, suggesting HSP70 immunoprecipitated specifically with HAZV-N. The presence of small amounts of HSP70 in the mock anti-HAZV N lanes could be due to the fact that anti-HAZV N does not represent purified antibodies; the anti-HAZV N antibodies are utilised within a mixture of other proteins in sheep serum; some of which could bind non-specifically to Dynabeads and cellular proteins (such as HSP70) in cell lysates.

Indirect IF analysis was used to determine the subcellular localisation of HSP70 during HAZV infection (Figure 4.14 (B)). Experiments were carried out using a similar protocol to that described in section 2.2.6.6, however antibody incubations were performed sequentially; following cell permeabilisation and blocking, primary mouse anti-HSP70 antibody was applied, which bound HSP70. After several PBS washes, secondary alexaFluor 594-conjugated anti-mouse antibody was applied, which bound the anti-HSP70 mouse primary antibody, and resulted in visualisation of HSP70 when excited with light of 594 nm. Anti-HAZV N was then applied, which bound HAZV N, and alexaFluor488 conjugated anti-sheep secondary antibody was used to bind to and visualise HAZV N, when excited with light of wavelength 488 nm. Coverslips were mounted in ProLong Gold containing DAPI to allow visualisation of cell nuclei, and images were captured on a Zeiss LSM510 confocal microscope. HSP70 (red, Figure 4.14 (B)) was detected within both the cytoplasm and nucleus of mock- infected and HAZV-infected cells. In HAZV infected cells (green, Figure 4.14 (B)), HSP70 occasionally co-localised with HAZV N in distinct puncta in the cytoplasm (yellow, Figure 4.14 (B)). The partial co-localisation of HSP70 with HAZV N within the cytoplasm of infected cells supported western blot analysis of the co-IP of HSP70 and HAZV N, and suggested a possible role for HSP70 during the replication cycle of HAZV.

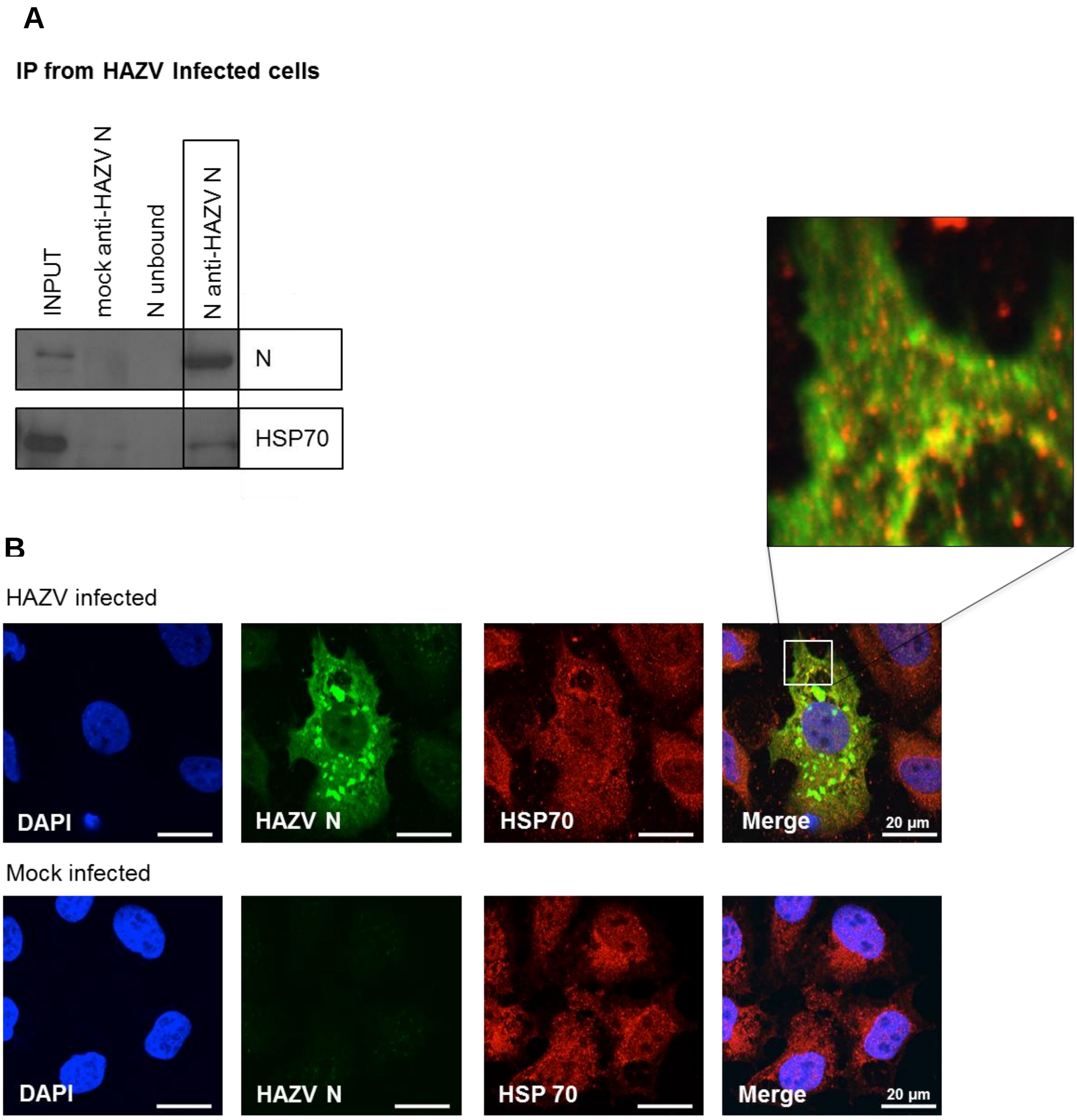


Figure 4.14 Co-IP and co-localisation of HSP70 and HAZV N

HSP70 immunoprecipitated with HAZV N from HAZV infected cells (A, boxed lane), and HSP70 is absent from, or present to a much lesser extent in control lanes. Indirect IF analysis indicated that HSP70 (red) was present in both the cell cytoplasm and nucleus, and localised in discrete cytoplasmic puncta. Occasionally HAZV N (green) co-localised with HSP70 in the cytoplasm (yellow), which supported the IP data (A), and suggested that HSP70 may have a function within the replication cycle of HAZV. The nucleus is stained blue with DAPI.

4.7 RESULTS: PURIFICATION OF HAZV PARTICLES, AND THE INTERACTION BETWEEN HSP70 AND HAZV N

4.7.1 Purification of HAZV particles

Members of the HSP70 family are detected in purified virus particles of several viruses whose proteins interact with HSP70, including rabies virus, hepatitis C virus (HCV) and respiratory syncytial virus (RSV) (Sagara and Kawai 1992; Parent, Qu et al. 2009; Radhakrishnan, Yeo et al. 2010). To determine whether HAZV particles also co-purified with HSP70, HAZV was purified from the supernatant of infected cells, and the virus containing fraction analysed by western blot using anti-HSP70 antibodies. HAZV was purified from the supernatant of infected cells by overnight precipitation at 4 °C with PEG 6000; the precipitate was centrifuged at 4 °C and the resulting pellet resuspended in TNE buffer. Precipitated HAZV diluted in TNE buffer was ultracentrifuged through a 5 % - 25 % continuous iodixanol gradient to further purify HAZV particles by density gradient centrifugation. After ultracentrifugation HAZV particles were isolated as a white band, with an approximate density between 0.33 g/mL and 0.55 g/mL; the fraction containing purified HAZV was aspirated from the gradient, then plaque assayed to estimate infectious virus titre (Figure 4.15).

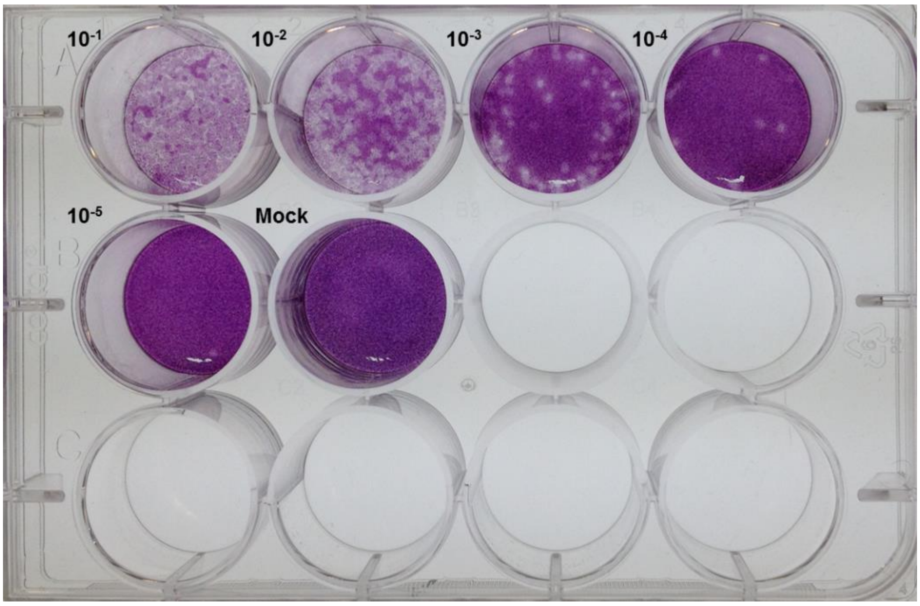


Figure 4.15 Plaque assay of the fraction containing infectious HAZV

After aspiration from the 5 %-25 % iodixanol gradient, the fraction containing purified HAZV particles was plaque assayed as described in section 2.2.9.2. The fraction containing purified HAZV had an estimated titre of ~2 x 10^5 *pfu/mL.*

Iodixanol supplied as a 60 % solution in water (‘OptiPrep’; density 1.32 g/mL) was used as the density gradient medium in the purification of HAZV particles, as iodixanol is non-ionic, non-toxic to cells, and is metabolically inert. In contrast to sucrose, iodixanol solutions have low viscosity and low osmolarity therefore HAZV particles purified by density gradient centrifugation through a 5 % - 25 % iodixanol gradient could be directly applied to cells, without dialysis, or affecting the osmotic properties of DMEM.

The fraction containing purified HAZV particles had an estimated titre of ~2 x 10^5 pfu/mL. The titre of HAZV decreased during purification, (the supernatant from which HAZV particles were purified had an estimated titre of ~5 x 10^6 pfu/mL) however the titre remained high enough to use the fraction containing infectious HAZV in further experiments. In addition to the fraction containing infectious HAZV particles, 4 more fractions were collected from the iodixanol gradient, (numbered 2-5 in Figure 4.16, representing fractions collected from the bottom of the iodixanol gradient (5) to the top of the gradient (2)). Fractions from the iodixanol gradient were mixed 1:1 in 2 x LDS buffer, heated at 95 °C for 10 minutes, then analysed by SDS PAGE; resolved gels were stained with Coomassie solution or silver stained (Figure 4.16).

Both Coomassie and silver stain analysis suggested the fraction containing infectious HAZV had the highest abundance of total protein. This was not unexpected, as cellular supernatant (from which HAZV was purified) should not contain a very high concentration of protein, with the exception of BSA (bovine serum albumin), the dominant protein species in FBS (foetal bovine serum) used to supplement DMEM. A band corresponding to the predicted molecular weight of BSA (66.4 kDa) was detected in fraction 2, fraction 3, and fraction 4 when fractions were silver stained.

When fractions were Coomassie stained (Figure 4.16 (A)), a band corresponding to the predicted molecular weight of HAZV N (54 kDa) could clearly be seen in the infectious HAZV containing fraction (lane; aspirated HAZV), and appeared to be the most abundant protein in this fraction.

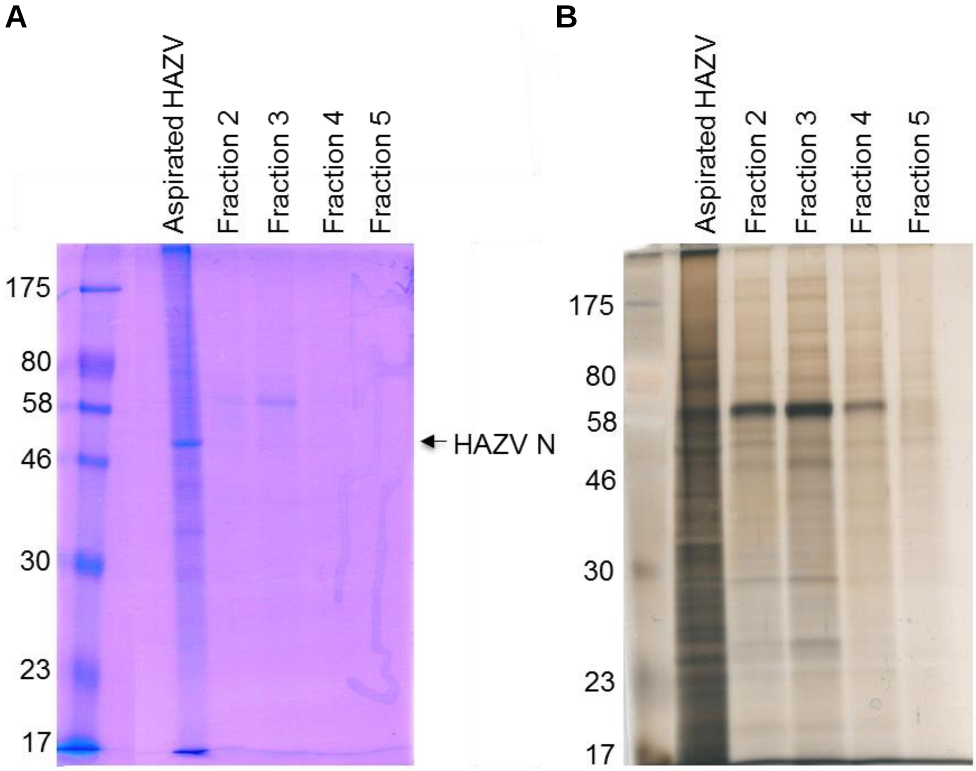


Figure 4.16 SDS PAGE analysis of density gradient centrifugation fractions from HAZV purification.

The final stage in the purification of HAZV particles from infectious SW13 cell supernatants was density gradient centrifugation through a 5 % - 25 % iodixanol gradient. HAZV particles were isolated as a white band 2/3 way down the gradient; this fraction was aspirated (lane; aspirated HAZV), and the remaining 5 fractions collected from the bottom of the gradient (5) to the top of the gradient (2). Fractions were analysed by SDS PAGE, and either Coomassie stained (A) or silver stained (B).

In addition to this proposed band of HAZV N, several fainter bands could be detected by Coomassie stain, and many more bands could be detected by silver stain, in the fraction containing infectious HAZV. This suggested the fraction containing infectious HAZV particles also contained multiple cellular proteins, in addition to HAZV encoded proteins.

4.7.2 Analysis of fractions containing purified HAZV particles using western blotting and IP

Fractions from the 5 % - 25 % iodixanol gradient were analysed by western blot, as described in section 4.6.2, to confirm the presence of HAZV N in the fraction containing infectious HAZV particles (Figure 4.17 (A)). When probed with anti-HAZV N, a band of the correct molecular weight (54 kDa) was identified in the 'aspirated HAZV' lane. HAZV N was also detected to a much lesser extent in lanes corresponding to fraction 2 and fraction 4; these may represent isolated RNPs, or virus particles that migrate to a different density for unknown reasons. Interestingly, when fractions were probed with anti-HSP70 antibodies, HSP70 could also be detected in the fraction containing infectious HAZV, suggesting HSP70 co-purified with infectious HAZV particles (Figure 4.17 (A)).

Next, HAZV N was IP'd from the fraction containing infectious HAZV using anti-HAZV N, as described in section 2.2.7.2. Control samples used in this experiment were the fraction containing infectious HAZV IP'd with unbound Dynabeads, and fraction 3 IP'd with anti-HAZV N bound Dynabeads. Immunoprecipitates were analysed by western blot using anti-HAZV N, and anti-HSP70 antibodies; HAZV N IP'd specifically with anti-HAZV N (Figure 4.17 (B) boxed lane); there was no HAZV N detected in the control lanes. HSP70 also immunoprecipitated with HAZV N from the fraction containing infectious HAZV (Figure 4.17 (B) boxed lane); there was very little or no HSP70 detected in the control lanes. This suggested that HAZV N remained associated with HSP70 during multiple stages of the virus replication cycle; HAZV N may even remain interacting with HSP70 within secreted HAZV particles.

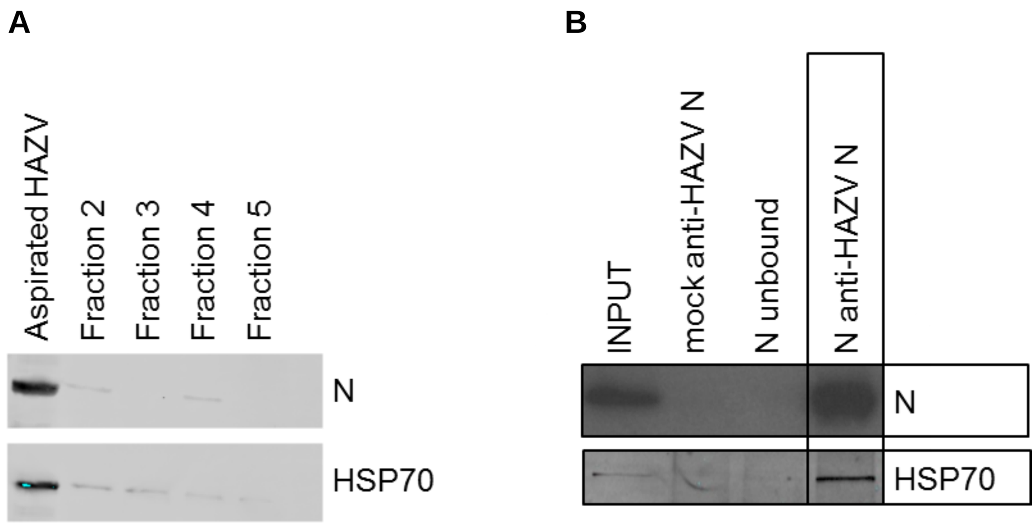


Figure 4.17 HSP70 in the fraction containing infectious HAZV, and the co-IP of HSP70 with HAZV N from this fraction

HAZV N and HSP70 were both detected in the fraction containing infectious HAZV when fractions from the iodixanol gradient were analysed by western blot (A). IP of HAZV N from the fraction containing infectious HAZV confirmed the co-IP of HSP70 with HAZV N (B), boxed lane, suggesting it is possible that HAZV N and HSP70 remained associated during multiple stages of the virus replication cycle, and might be occasionally packaged together into secreted HAZV particles.

4.7.3 Mass spectrometry analysis of the fraction containing infectious HAZV

HAZV N and HSP70 were both detected by western blot analysis in the fraction containing infectious HAZV. To identify other proteins in this fraction, that were visible by Coomassie and silver stain analysis (Figure 4.16), mass spectrometry (MS) analysis of the fraction containing infectious HAZV was carried out by Dr Stuart Armstrong at Liverpool University. Briefly; the fraction containing infectious HAZV and fraction 3 were denatured, then ~11 µg total protein from these fractions was sent for MS analysis. Equivalent amounts of denatured protein from fraction 3 (~11 µg) was sent as a negative control, to determine whether any cellular proteins were present throughout the iodixanol gradient.

All four structural proteins of HAZV were identified in the fraction containing infectious HAZV (table 7); in good agreement with the Coomassie stained gel in Figure 4.16 HAZV N was identified as the most abundant protein in the whole fraction. Of the HAZV structural proteins, the glycoproteins G_N and G_C were identified as the next most abundant, with the L protein detected as the least abundant viral protein in purified HAZV particles. Approximate protein abundance was estimated using the emPAI score (the exponentially modified protein abundance index) which is calculated automatically during protein identification with the Mascot database search engine. The emPAI score is based on protein coverage by the peptide matches in a database search result to determine an approximate, label free, relative quantitation of the proteins in a mixture; the number of identified peptides correlates well with protein abundance (Ishihama, Oda et al. 2005). The emPAI score is however, only a crude estimate of protein abundance, providing a guideline to the abundance of a specific protein within a mixture, and is not as precise as SILAC-based quantitation methods.

Table 7 HAZV proteins detected in the fraction containing infectious HAZV.

The protein accession number can be entered into the UniProtKB database to obtain a description of the protein. The protein mass is estimated based on the Mascot database prediction and the emPAI score provides an approximate estimate of the abundance of each protein.

Protein_accession number	Protein_description	Protein_mass (Da)	emPAI
M4PWE6	Nucleocapsid protein	54639	16.61
A6XIP3	Glycoproteins G_N and G_C	161105	1.23
A6XA53	L protein	450300	0.01

Including HAZV N, G_N, G_C and L protein, MS analysis identified 264 proteins in the fraction containing infectious HAZV, and 82 proteins in fraction 3. Of these 82 proteins identified in fraction 3, only 13 were not identified in the fraction containing infectious HAZV. The proteins identified in fraction 3 could therefore be used as a guide to identify proteins that were present in multiple fractions, and therefore unlikely to be contained within, or associated with, purified HAZV particles. However as HAZV N, G_N

and G_C are also identified in fraction 3 (albeit with much lower emPAI scores; G_N and $G_{C:}$ 1.23 and 0.06, HAZV N: 16.61 and 0.51, respectively) it is harder to confirm that these proteins were not associated with components of the virus particles, or the particles themselves. Proteins identified by MS, and thought to be present in the fraction containing infectious HAZV should be validated by western blot analysis. A complete list of all proteins identified in the fraction containing infectious HAZV is provided in appendix III.

Several members of the HSP70 family were identified in the fraction containing infectious HAZV; table 8 summarises the identity of these HSP70 family members, their emPAI scores, and a brief description of their function (adapted from the UniProt database). HSC70 was identified in the fraction containing infectious HAZV, in good agreement with Figure 4.17 (A); where western blot analysis of the fraction containing infectious HAZV detected a protein of the predicted molecular weight using the anti-HSP70 antibody that recognised both HSC70 and HSP70. It is likely the antibody detected only HSC70 in Figure 4.17. The identification of HSC70 in the fraction containing infectious HAZV also supported both the original MS data from both EGFP-CCHFV N IPs, which identified HSC70 in EGFP-N immunoprecipitates with high SILAC ratios (27.11, 24.72), and all subsequent IPs of both CCHFV N and HAZV N which identified HSP70 in N-immunoprecipitates using anti-HSP70 that also recognises HSC70. It is possible that HAZV N and HSC70 remained associated during multiple stages of the HAZV replication cycle, such that HSC70 might be packaged within secreted HAZV particles; alternatively HSC70 may be associated with the outside of secreted HAZV particles.

Table 8 Summary of members of the HSP70 family identified in the infectious HAZV containing fraction

The protein accession number can be entered into the UniProtKB database to obtain a description of the protein. The emPAI score provided an approximate estimate of the abundance of each protein. A brief description of the function of each protein is also provided, based on the UniProt database description.

Protein accession number	Protein_description	emPAI	Function
P11142	Heat shock cognate 71 kDa protein (HSC70 - heat shock cognate protein 70)	0.97	Constitutively expressed molecular chaperone; facilitates correct folding of nascent proteins, helps disrupt multi-protein complexes, regulates cell signaling and apoptosis, amongst many other functions.
P17066	Heat shock 70 kDa protein 6 (Heat shock 70 kDa protein B)	0.2	Molecular chaperone that stabilises pre-existing proteins against aggregation, and helps nascent proteins fold correctly. Expressed in response to the unfolded protein response, and cellular stress.
P34931	Heat shock 70 kDa protein 1-like	0.15	Molecular chaperone, increased expression following induction of the unfolded protein response. Interacts with PARK2 (E3 ubiquitin ligase).
P11021	78 kDa glucose-regulated protein (Bip)	0.09	Endoplasmic reticulum (ER) located molecular chaperone. Facilitates the assembly of multimeric protein complexes within the ER.
Q9Y4L1	Hypoxia up-regulated protein 1	0.06	Located in the lumen of the ER and is induced by hypoxia. Highly expressed in tissues that have a well developed ER and synthesise large amounts of secretory proteins. Highly expressed in the liver.

4.8 SECTION SUMMARY

This section presents the use of actively replicating HAZV, to investigate the interaction between HSP70 and HAZV N. Conditions for HAZV infection were optimised to enable stocks of HAZV to be grown, these were used to infected SW13 cells and assay the ability of anti-HAZV N to detect HAZV N synthesised during infection, through western blot and indirect IF analysis. HAZV N was IP’d from HAZV infected cells using anti-HAZV N and immunoprecipitates were analysed by western blot; this confirmed the IP of HAZV N, and the co-IP of HSP70. The association between HAZV N and HSP70 was supported by indirect IF analysis, where HAZV N could be observed co-localising with HSP70 in certain cytoplasmic puncta. Secreted HAZV particles were purified from the supernatant of infected cells by PEG6000 precipitation and density gradient centrifugation. The fraction containing infectious HAZV from the gradient was analysed by western blot; both HAZV N and HSP70 were detected in this fraction; furthermore HSP70 immunoprecipitated with HAZV N from this fraction. MS analysis of the fraction containing infectious HAZV identified 264 cellular proteins, as well as all four structural HAZV proteins and HSC70, which was also detected during western blot analysis of this fraction with the anti-HSP70 antibody. The ratio of infectious HAZV particles: defective, non-infectious particles in this fraction is unknown however.

The results in this section suggested HSP70/HSC70 interacted with HAZV N at multiple stages of the virus replication cycle, and may be packaged into, or associated with secreted HAZV particles. There is growing evidence that enveloped virus particles carry multiple host proteins both internally and externally during infection (Cantin, Methot et al. 2005). HSP70 and/or HSC70 is associated with purified RSV particles, rabies virus particles, HCV particles, Newcastle disease virus particles, and herpes simplex virus (HSV) particles, amongst others (Sagara and Kawai 1992; Cantin, Methot et al. 2005; Parent, Qu et al. 2009; Radhakrishnan, Yeo et al. 2010; Stegen, Yakova et al. 2013). This suggests that association of members of the HSP70 family with extracellular virus particles may be common strategy adopted by viruses of multiple different families to aid virus replication. Few studies have investigated the significance of the incorporated fraction of cellular proteins to virus infectivity; (Stegen, Yakova et al. 2013) however were able to show that depletion of HSC70 in purified HSV particles using siRNA reduced the infectious titre of these particles by > 50 %. Various studies have depleted HSP70 and/or HSC70 in cells using siRNA, or used small molecule inhibitors of HSP70/HSC70, and assayed the effect on virus replication; for example titres of rabies virus and HCV both decreased when HSP70 or

HSC70 was depleted from infected cells (Lahaye, Vidy et al. 2012; Khachatoorian, Ganapathy et al. 2014). The following section presents experiments designed to assay the effect on HAZV replication of depleting HSP70 from SW13 cells using siRNA, or interfering with the function of these proteins using small molecule inhibitors of HSP70 and HSC70.

4.9 RESULTS: THE EFFECT OF DEPLETING OR INHIBITING HSP70/HSC70 ON HAZV REPLICATION

This section describes preliminary experiments to determine the functional effects on HAZV replication of depleting HSP70 levels with siRNA, or inhibiting HSP70/HSC70 using the small molecule inhibitors, VER155008 and pifithrin-µ.

4.9.1 The effect of HSP70 depletion by siRNA on HAZV replication

In this section siRNA (short interfering RNA) targeting HSP70 (HSPA1A gene) was utilised to examine the role of HSP70 during HAZV replication. HSP70 was chosen for depletion as the SILAC ratios for HSP70 in the initial MS analysis of EGFP-CCHFV N IPs were higher than the HSC70 ratios (37.85 and 40.11, compared to 27.11 and 24.72 respectively) and it was thought CCHFV N (and by inference HAZV N) might preferentially interact with HSP70. In addition, multiple studies have shown that depletion of HSC70 results in upregulation of HSP70 expression (Schlecht, Scholz et al. 2013; Khachatoorian, Ganapathy et al. 2014); if recruitment of HSP70 and HSC70 by HAZV is interchangeable due to HAZV replication requiring a conserved function possessed by both HSC70 and HSP70, any effects on HAZV replication due to depletion of HSC70 may be masked by the upregulation of HSP70.

SW13 cells were transfected with 80 pmole HSP70 siRNA per well to knock down HSP70, or 80 pmole scrambled siRNA was transfected as a negative control for off-target effects due to siRNA transfection. RNAiMAX, a cationic lipid-based transfection reagent was used to transfect SW13 cells with siRNA, as described in section 2.2.5.2. Following cell entry siRNA (which consists of short, 21 nucleotides long strands of dsRNA, complimentary to nucleotide sequences of specific mRNA molecules) is detected by the cellular protein DICER, and transferred to ARGONAUT, to form the RISC complex (RNA induced silencing complex). ARGONAUT breaks the hydrogen bonds between the dsRNA to generate a ssRNA strand, the ‘guide strand’ which is complimentary to the mRNA of the specific gene to be silenced. The guide strands anneals to its target mRNA, which is then degraded by the exonuclease activity of the RISC complex. Degradation of mRNA ensures the protein encoded by a specific gene (in this case HSP70 encoded by HSPA1A) is not translated, and over time, cellular protein turnover ensures depletion of the specific protein.

Depletion of cellular proteins can be cytotoxic, which can result in reduced virus replication. SW13 cells transfected with HSP70 siRNA, or scrambled siRNA, were MTT assayed 24 hrs and 48 hrs after transfection, to ensure depletion of HSP70 was not cytotoxic. The MTT assay is a colorimetric assay to determine cell viability, based on the cleavage of yellow tetrazolium salt (MTT) by mitochondrial enzymes to form a soluble, blue formazan product; the amount of formazan produced is directly proportional to the number of living, not dead cells, and can be detected by absorption of light at 570 nm (Figure 4.18 (C)) (Sylvester 2011).

Depletion of HSP70 by siRNA transfection was not cytotoxic (Figure 4.18 (C)) and 24 hrs after transfection, resulted in a 67.8 % reduction of HSP70 protein levels, relative to HSP70 protein levels from scrambled siRNA transfected cells (Figure 4.18 (A) and (B)). HSP70 protein levels were quantified using StudioImage software, after scanning membranes probed with fluorescent secondary antibodies (which bound primary anti-GAPDH and anti-HSP70 antibodies; ‘anti-HSP72’ antibodies from Enzo Life sciences were utilised, which specifically detect the HSPA1A gene product) using the LiCor Odyssey Sa Infrared imaging system. Fluorescent signals from HSP70 were normalised to GAPDH signals, and the geometric mean of HSP70 signals from the three repeat siRNA experiments using either HSP70 siRNA, or scrambled siRNA, was calculated. The geometric mean was used to determine the difference in HSP70 protein levels in HSP70 siRNA, compared to scrambled siRNA, transfected cells.

However, 48 hrs after transfection expression levels of HSP70 had increased, and there was only a 45.2 % reduction of HSP70 protein levels, relative to HSP70 protein levels from scrambled siRNA transfected cells (Figure 4.18 (A) and (B)). Re-transfecting cells with HSP70 siRNA at the 24 hr time point did not result in increased HSP70 depletion at future time points; neither did altering the amount of siRNA or RNAiMAX in the transfection mix. This assay should be repeated using different siRNA sequences targeting HSP70, or a mixed pool of siRNA sequences targeting multiple regions of the HSP70 mRNA, to determine the effect of more efficient HSP70 depletion on HAZV replication.

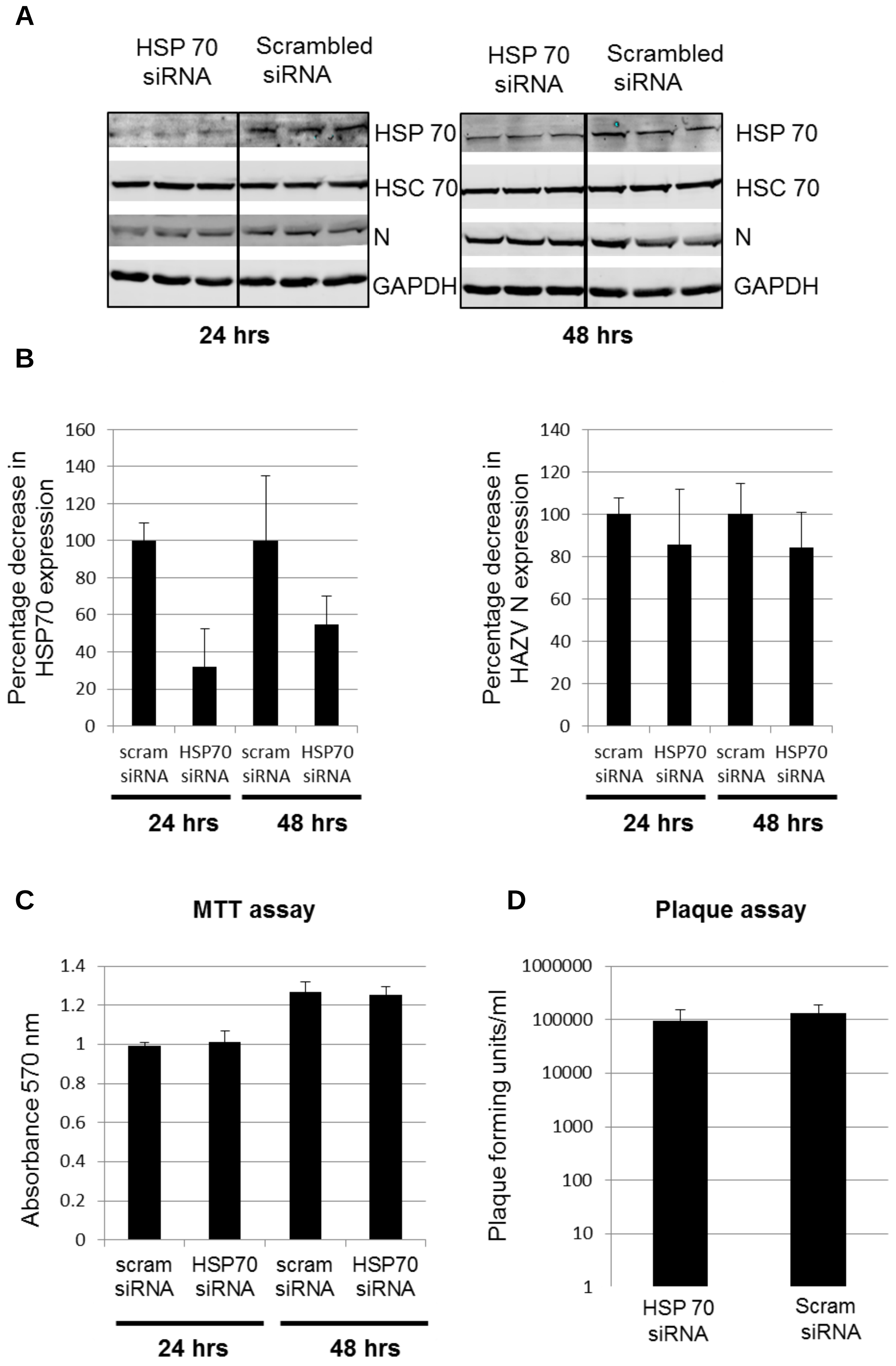
Depletion of HSP70 was highly specific, and did not affect expression levels of HSC70 (detected using antibodies specific to HSC70), or GAPDH (Figure 4.18 (A)).

Cells were infected with HAZV at an MOI 1 following overnight transfection with HSP70 siRNA, or scrambled siRNA; when cell supernatant and proteins were harvested 24 hrs after siRNA transfection, HAZV had been replicating in SW13 cells for 6 hrs. After 6 hrs of replication HAZV N could be detected by western blot with anti-HAZV N in the SW13 cell lysate (Figure 4.18 (A)), however secreted HAZV could not be detected by plaque assay. There were slightly reduced levels of HAZV N in HSP70 depleted cells 6 hrs after infection, compared to scrambled siRNA transfected cells; HAZV N expression was at 85.4 % of levels of HAZV N detected in lysates from scrambled siRNA transfected cells. HAZV N expression levels were quantified as described for HSP70, after normalisation of HAZV N fluorescent signals to GAPDH signals.

Cell lysates harvested 48 hrs after siRNA transfection (30 hrs after HAZV infection) were analysed by western blot; the level of HSP70 depletion was not as pronounced as at 24 hrs post-transfection (HSP70 levels were reduced by 45.2 %, compared to 67.8 % at 24 hrs post-transfection). However the abundance of HAZV N in HSP70 siRNA-transfected cells at 48 hrs was still slightly reduced compared to scrambled siRNA transfected cells (HAZV N proteins levels remained at 84 % of those detected in lysates from scrambled siRNA transfected cells) (Figure 4.18 (B)). Analysis of secreted infectious HAZV by plaque assay also revealed a slight decrease in titre in HAZV secreted from HSP70 depleted cells compared to scrambled siRNA transfected cells; HAZV secreted from HSP70 depleted cells had an average titre of 9.4 x 10^4 pfu/mL, and HAZV secreted from scrambled siRNA transfected cells had an average titre of 1.3 x 10^5 pfu/mL (Figure 4.18 (D)).

Figure 4.18 siRNA mediated depletion of HSP70 and effects on HAZV replication

(Below) siRNA mediated depletion of HSP70 in SW13 cells was not cytotoxic when compared to scrambled siRNA transfection, when cell viability was measured by MTT assay (C). Depletion of HSP70 was specific, (expression levels of HSC70 and GAPDH were unchanged) and the greatest depletion of HSP70 occurred 24 hrs after transfection with HSP70 siRNA; expression levels of HSP70 had increased again by 48 hrs post-transfection (A) and (B). HAZV N protein expression levels were slightly reduced in HSP70 depleted cells compared with cells that had been transfected with scrambled siRNA (B) and titres of infectious HAZV secreted from HSP70 depleted cells were also slightly reduced compared to scrambled siRNA transfected cell supernatant (D).


A
HSP 70 siRNA
Scrambled siRNA
HSP 70
HSC 70
N
GAPDH
24 hrs
48 hrs
B
Percentage decrease in HSP70 expression
160
140
120
100
80
60
40
20
0
scram siRNA
HSP70 siRNA
24 hrs
48 hrs
Percentage decrease in HAZV N expression
C
MTT assay
Absorbance 570 nm
1.4
1.2
1
0.8
0.6
0.4
0.2
0
D
Plaque assay
Plaque forming units/ml
1000000
100000
10000
1000
100
10
1
HSP 70 siRNA
Scram siRNA

These data suggested that the interaction between HSP70 and HAZV N was not detrimental to virus replication, as no increase in HAZV replication was observed when HSP70 was depleted. The reduction in expression levels of HAZV N, and the decrease in secreted infectious HAZV titre, between HSP70 siRNA transfected cells compared to scrambled siRNA transfected cells, were modest (15 % reduction in HAZV N expression and 28 % reduction in virus titre). It could be that a more efficient depletion of HSP70 could result in a greater reduction in HAZV N expression and secreted infectious HAZV titre. Alternatively, siRNA mediated depletion of HSC70 could have a greater effect on the HAZV replication cycle, and HAZV secretion, then depletion of HSP70, as has been shown for HCV (Khachatoorian, Ganapathy et al. 2014).

4.9.2 Effect on HAZV replication of inhibiting HSP70/HSC70 with VER155008 and pifithrin-µ

VER155008 (VER) and pifithrin-µ (PIF) (a commonly used synonym for PIF is**:** 2-Phenylethynesulfonamide (PES)) are both small molecule inhibitors of HSP70/HSC70. All functions of HSP70 family members are controlled by binding of HSP70 to protein targets via a C-terminal substrate binding domain (SBD); this binding is allosterically controlled by a nucleotide (ATP) binding in the N-terminal ATP-ase domain (the nucleotide binding domain (NBD)) (Schlecht, Scholz et al. 2013). VER was developed as a small molecule inhibitor of HSP70 ATPase activity; VER acts as an ATP-competitive inhibitor of HSP70 by binding to the HSP70 NBD and arresting it in a ‘half-open’ conformation, thereby preventing ATP from binding, and inhibiting the allosteric control between the NBD and SBD (Massey, Williamson et al. 2010; Schlecht, Scholz et al. 2013). PIF is thought to target the SBD of HSP70; interaction of PIF with the HSP70 SBD disrupts the association between HSP70 and several of its cofactors and client proteins, thereby inhibiting HSP70 function (Kondoh and Osada 2013). VER and PIF are not specific for individual HSP70 family members, they target both HSP70 and HSC70 (Schlecht, Scholz et al. 2013).

The cytotoxic effects of VER and PIF on SW13 cells were estimated by MTT assay over multiple concentrations of VER and PIF; VER and PIF were dissolved in DMSO, then diluted in complete media and applied to SW13 cells for 24 hrs prior to the MTT assay (Figure 4.19). (The highest concentration of VER was limited by the solubility of this compound in DMEM). This defined working concentrations of VER and PIF which did not affect cell viability, which were then used to assay the ability of HAZV to replicate in the presence of these compounds.

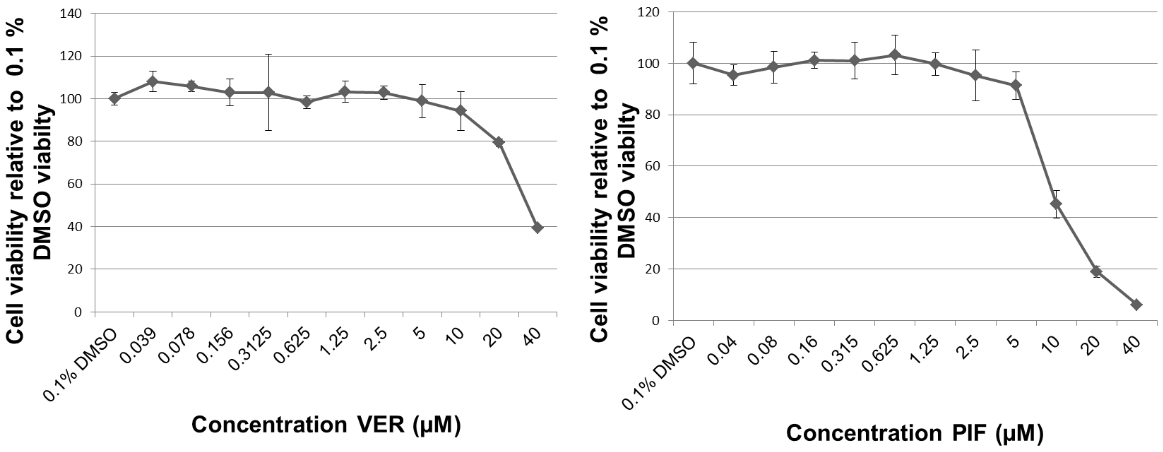


Figure 4.19 Cytotoxic effects of VER and PIF measured by MTT assay over multiple compound concentrations.

SW13 cells were treated with VER and PIF for 24 hrs over a range of concentrations (40 µM – 0.04 µM), and cell viability was measured relative to cells treated with 0.1 % DMSO (a DMSO concentration greater than that present at the highest concentration of compound). Cell viability relative to 0.1 % DMSO was plotted against compound concentration.

SW13 cells were pre-treated with VER and PIF for 1 hr, prior to infection with HAZV at an MOI of 1. SW13 cells were infected as described in section 2.2.9.1, however after HAZV adsorbed to cells for 1hr and HAZV diluted in SFM was removed, media containing 2 % FBS with the appropriate concentration of VER and PIF was layered over cells. HAZV was allowed to replicate in SW13 cells in the presence of various concentrations of VER and PIF for 24 hrs, cell lysates and supernatants were then harvested; HAZV N was detected by western blot analysis, and secreted HAZV by plaque assay (Figure 4.20).

HAZV N expression decreased in the presence of increasing concentrations of PIF, such that HAZV N expression was not detected in the lysates of cells grown in 6 µM PIF for 24 hrs (Figure 4.20 (B)); this corresponded well with plaque assays of HAZV secreted from SW13 cells treated with 0.5 µM and 2 µM PIF. The amount of HAZV secreted from SW13 cells grown in the presence 0.5 µM PIF for 24 hrs was decreased by 75 % relative to amount of HAZV secreted from cells grown in the presence of 0.2 % DMSO, and HAZV secreted from cells grown in 2 µM PIF for 24 hrs was undetectable by plaque assay (Figure 4.20 (C)). In contrast, although levels of

HAZV N expression decreased in the presence of increasing concentrations of VER when assayed by western blot (Figure 4.20 (B)), HAZV N expression was still detectable in the lysates of cells grown in the presence of 20 µM VER for 24 hrs. However plaque assays of HAZV secreted from SW13 cells treated with 16 µM VER for 24 hrs, revealed a 95 % decrease in HAZV titre, when compared to HAZV secreted from cells treated with 0.2 % DMSO.

MTT assay of uninfected SW13 cells treated with corresponding concentrations PIF (Figure 4.20 (A)) showed little cytotoxicity that could account for the decrease in HAZV replication observed. Viability of cells treated with between 6 µM and 20 µM VER decreased to 80 % of the viability level of cells treated with 0.2 % DMS0 (a DMSO concentration greater than that present in any dilution of VER or PIF). This decrease in cell viability may have contributed to the decrease in HAZV titres observed at 16 µM VER, however the decrease in cell viability is not as great as the decrease in virus titre at 16 µM VER.

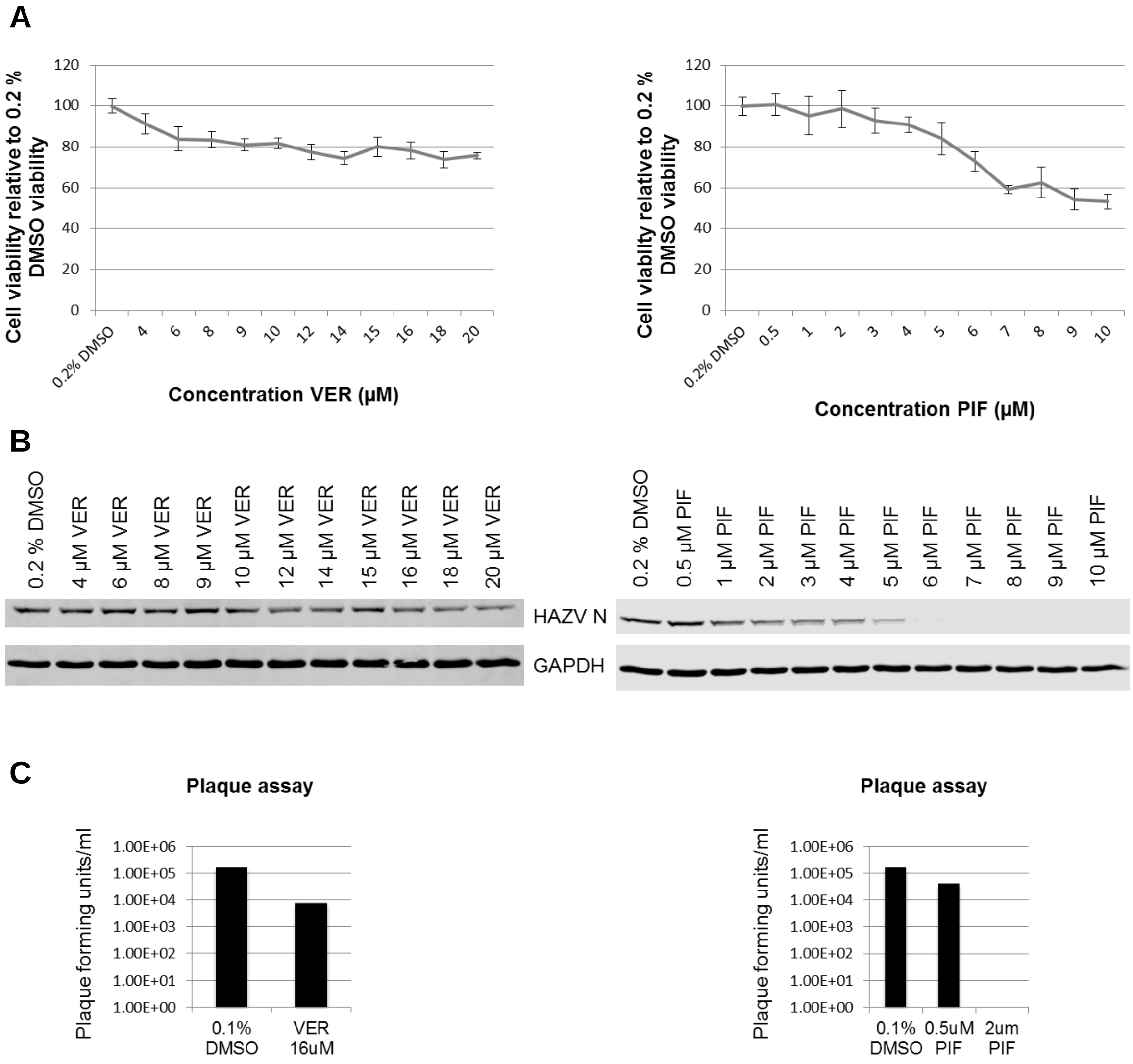


Figure 4.20 The effect of VER and PIF on HAZV replication

SW13 cells were treated with VER and PIF, at concentrations ranging from 4 µM – 20 µM, and 0.5 µM to 10 µM, respectively. After 24 hrs incubation of VER and PIF with SW13 cells in the absence of HAZV, cytotoxicity was estimated by MTT assay (A). SW13 cells grown in the presence of PIF and VER were then infected with HAZV for 24 hrs, and virus replication efficiency was determined through western blot analysis of HAZV N and plaque assay (B) and (C). Plaque assays were performed once. Western blot analysis was performed after total protein concentration of cell lysates was determined by BCA assay; then equal amounts of protein were loaded onto SDS PAGE gels, prior to transfer of protein to FL-PVDF membranes and probing with anti-HAZV N and anti-GAPDH antibodies.

The decrease in HAZV N expression, and decrease in titre of secreted HAZV from cells treated with PIF and VER, in comparison to cells treated with 0.2 % DMSO, is consistent with a positive role for HSP70/HSC70 in the HAZV replication cycle. However, in this assay, the decrease in HAZV N expression and HAZV titre in the presence of PIF and VER, could not be directly linked to the involvement of HSP70/HSC70 in the HAZV replication cycle. This was due to the lack of a specific cellular HSP70/HSC70 function to assay, in order to determine the efficacy of HSP70/HSC70 inhibition by PIF and VER. HSP70 and HSC70 behave in a relatively non-specific manner in the control of protein folding, therefore there are currently no reliably identified ‘client proteins’ of HSP70 or HSC70 (Budina-Kolomets, Balaburski et al. 2014). Degradation of chaperone client proteins is an established method to assay the inhibition of chaperones by small molecule inhibitors; inhibition of HSP90 for example, results in the ubiquitin dependent proteosomal degradation of HSP90 client proteins, including Her2, CDK4, Raf1 and AKT (Massey, Williamson et al. 2010). Given that HSP70 has been shown to play a key role in the HSP90 chaperone machinery (inhibition of HSP90 results in up regulation of HSP70 expression, in addition to client protein degradation), reduced levels of the HSP90 client protein Raf1 have been used to measure the efficacy of HSP70 inhibition by VER in the past (Massey, Williamson et al. 2010). However concentrations of VER at which Raf1 expression levels are reduced are cytotoxic, and any effect on virus replication could be attributed to HSP90 inhibition.

4.10 CHAPTER SUMMARY AND DISCUSSION

This chapter presents the identification of cellular proteins that interact with EGFP-CCHFV N, and validation of the interaction of HSP70 with un-tagged CCHFV N and HAZV N. This chapter also presents the purification of secreted HAZV particles from the supernatant of infected cells, the detection of HSC70 within the fraction containing infectious HAZV by both western blot and MS analysis, and the co-IP of HSP70 with HAZV N from this fraction. This chapter then goes on to describe preliminary experiments to dissect the role of HSP70/HSC70 within the HAZV replication cycle using siRNA and small molecular inhibitors of HSP70/HSC70.

The N proteins of other bunyaviridae interact with components of their host cells, including cellular proteins, and these interactions can have important consequences both for viral replication and for the cell. The hantavirus N protein, for example, has been shown to have multiple roles in the host cell; hantavirus N interacts with the 5’ caps of cellular RNAs in cytoplasmic P bodies and protects them from degradation (Mir, Duran et al. 2008). As these cellular 5’ capped oligoribonucleotides are then used to prime viral transcription, it is thought that the hantavirus N protein therefore has a role to play in the cap snatching mechanism essential for initiation of viral transcription. In addition, the hantavirus N protein interacts with the ribosomal protein S19, a component of the 40S ribosomal subunit (Haque and Mir 2010). This interaction recruits the 40S ribosomal subunit to viral mRNA transcripts and initiates N-mediated translation initiation. In this way, the hantavirus N protein is thought to functionally replace the entire cellular eIF4F cap binding and translation initiation complex, to preferentially enhance translation initiation of viral mRNA transcripts (Cheng, Haque et al. 2011). The N protein of BUNV interacts with the cellular protein poly (A) binding protein (PABP), which helps to shuttle polyadenylated mRNAs from the nucleus to the cytoplasm. PABP is predominately cytoplasmic, however BUNV N co-localises with PABP in the cytoplasm of infected cells at early timepoints after infection, and at later time points, causes a dramatic re-distribution of PABP to the nucleus. It is thought this interferes with poly (A) dependent cellular translation initiation, and thus aids viral poly (A) independent translation initiation (Blakqori, van Knippenberg et al. 2009).

To determine whether CCHFV N could also have a role within the host cell, in a similar manner to the hantavirus and BUNV N proteins, SILAC-based MS analysis coupled to GFP-Trap IPs was used for the unbiased identification of cellular proteins that immunoprecipitated with EGFP-CCHFV N from transfected 293T cells. The use of SILAC enabled the effective discrimination between cellular proteins that interacted specifically with EGFP-CCHFV N, and those that bound non-specifically to components of the GFP trap. After EGFP-CCHFV N was IP’d from cell lysates, MS analysis was performed twice on the same sample; in total 192 cellular protein were identified and quantified that immunoprecipitated with EGFP-CCHFV N; 112 proteins were common to both analyses, 46 proteins unique to the first analysis and 34 proteins unique to the second. The difference in MS results generated from the same sample emphasised the importance of validating potential interacting proteins using an independent method. Independent western blot analysis confirmed the co-IP of several cellular proteins with high SILAC ratios with EGFP-CCHFV N, including actin (a known cellular interacting partner of CCHFV N) (Andersson, Simon et al. 2004), vimentin, p53 and HSP70/HSC70. The identification of actin by MS, and confirmation of the co-IP of actin with EGFP-CCHFV N provided a measure of confidence in the MS results. Furthermore, the absence of GAPDH and B23 (both proteins that were identified with low SILAC ratios in the MS analysis) in immunoprecipitates from GFP-

Trap IP'd EGFP-CCHFV N supported the 6-fold cut off ratio that was chosen to differentiate between proteins that immunoprecipitated specifically with EGFP-CCHFV N, and those that bound non-specifically to components of the GFP-Trap.

HSP70 was chosen to be further validated as an interacting partner of untagged CCHFV N due to the high SILAC ratios of both HSP70 and HSC70, and the abundance of HSP70 that immunoprecipitated with EGFP-CCHFV N when analysed by western blot. Growing evidence from the literature indicates HSP70 interacts with viral proteins from several different families including: the nucleocapsid protein of rabies virus (RV), NS5A of HCV, the PB2 and PB1 monomers of influenza A virus (IAV) and both the nucleoprotein and phosphoprotein of RSV (Brown, Rixon et al. 2005; Lahaye, Vidy et al. 2012; Oliveira, Simabuco et al. 2013; Khachatoorian, Ganapathy et al. 2014; Manzoor, Kuroda et al. 2014). HSP70 was shown to have a positive functional role in the replication cycles of RV, HCV and IAV, as depletion of HSP70 using siRNA resulted in reduced virus transcription and replication. However increased HSP70 expression during Hantaan virus infection decreased virus replication, and the interaction between HSP70 and the Hantaan virus N protein is thought to be detrimental to virus replication; this suggests the role of HSP70 in virus replication cycles is not always beneficial for the virus (Yu, Ye et al. 2009). The choice of HSP70 to be further validated was therefore also based on previous studies that had shown HSP70 to be involved in the replication of several different viruses, raising the possibility that HSP70 could also be involved in CCHFV replication.

To confirm the co-IP of HSP70 was not an artefact of overexpression of the EGFP-CCHFV N fusion protein, untagged CCHFV N was IP'd from cells transfected with pCAGGs-CCHFV N, and from CCHFV infected cells, using anti-CCHFV N conjugated to Dynabeads. In IP experiments from both CCHFV infected and pCAGGs-CCHFV N transfected cells, HSP70 immunoprecipitated with CCHFV N, as verified by western blot analysis. The co-IP of HSP70 with CCHFV N from CCHFV infected cells suggested that HSP70 interacted with CCHFV N during infection, and therefore that HSP70 may play a role in the replication cycle of CCHFV.

One of the main factors that limits investigations into the CCHFV replication cycle is the requirement to work under containment levels 4 conditions when manipulating CCHFV, due to the highly pathogenic and lethal nature of CCHF. In 2011, the discovery that HAZV infection of IFNα/βR$^{-/-}$ mice resulted in very similar symptoms and pathology to CCHFV infection, suggested that HAZV infection in this small animal system may be a valid model to act as a CCHFV surrogate (Dowall, Findlay-Wilson et al. 2012). Data presented in chapter 3 indicated that CCHFV N and HAZV N are structurally very similar, and support the use of HAZV as a model for

CCHFV infection. HAZV and CCHFV are classified in the same serogroup within the same bunyavirus genus, which suggests these viruses adopt similar replication strategies; however in contrast to CCHFV, HAZV can be manipulated under containment level 2 conditions, and is not known to be pathogenic in humans. Experiments were therefore carried out to determine whether HSP70 also interacted with HAZV N, as the potential functional role of HSP70 within the replication cycle of HAZV could be more easily investigated under containment level 2 conditions.

Optimal conditions for growth of HAZV stocks were determined; infection of SW13 cells with HAZV at an MOI of 0.001 and harvesting cell supernatants 4 days after infection, resulted in the highest HAZV titres. HAZV infection of SW13 cells resulted in expression of HAZV N, which could be detected by western blot and IF analysis with anti-HAZV N antibodies (prepared using recombinant HAZV N protein, as described in section 3.2.4); HAZV N had a molecular weight of ~54 kDa, and was detected in the cytoplasm of infected cells, with a similar subcellular distribution to CCHFV N. In good agreement with CCHFV N IP's, HSP70 also immunoprecipitated with HAZV N from the lysates of infected cells, and HSP70 partially co-localised with HAZV N in the cytoplasm. These data supported a role for HSP70 during the replication cycle of HAZV.

Multiple different viruses (including several whose component proteins interact with members of the HSP70 family) have HSP70 family members packaged into, or associated with, secreted purified virus particles, including: RSV, rabies virus, HCV, HIV and Newcastle disease virus, amongst others (Sagara and Kawai 1992; Chertova, Chertov et al. 2006; Parent, Qu et al. 2009; Radhakrishnan, Yeo et al. 2010; Ren, Xue et al. 2012). Others viruses whose secreted virions have been analysed by mass spectrometry, but have not incorporated HSP70 include influenza virus (Shaw, Stone et al. 2008). Whilst not all host proteins are necessarily incorporated specifically into secreted virus particles, the association of HSP70 with virus particles from so many different families possibly indicates a conserved involvement of HSP70 in the replication cycle of multiple different viruses. As HSP70 immunoprecipitated with HAZV N, the association between HSP70 and purified HAZV particles was investigated; secreted HAZV particles from the supernatant of infected cells were purified by PEG 6000 precipitation, followed by density gradient centrifugation through a 5 %- 25 % iodixanol gradient. When the fraction containing infectious HAZV was analysed by western blot; HAZV N could be detected in this fraction, as could HSP70.

Coomassie and silver stain analysis of the fraction containing infectious HAZV revealed the presence of multiple protein bands, not all of which could be attributed to HAZV proteins. There is growing evidence that multiple host cell proteins are incorporated into, or onto newly formed enveloped viruses; it is thought that given the simplicity of viral proteomes, and the complexity of cell entry and exit, secreted virus particles may incorporate cellular proteins to aid these processes, and virion associated cellular proteins may therefore represent novel antiviral targets (Stegen, Yakova et al. 2013). MS analysis was used to identify cellular proteins in the fraction containing infectious HAZV.

MS analysis identified all 4 HAZV proteins (N, G_N, G_C and L) in the fraction containing infectious HAZV, with N present in the highest abundance. MS analysis also identified several members of the HSP70 family, including HSC70, which was presumed to be detected in all previous IPs with an anti-HSP70 antibody that bound both HSP70 and HSC70, and in western blot analysis of this fraction. Furthermore HSP70 immunoprecipitated with HAZV N from the fraction containing infectious HAZV, which suggested that HAZV N associated with HSP70 at multiple stages of the HAZV replication cycle.

The role of HSP70 within the replication cycle of HAZV was therefore investigated using siRNA that targeted HSP70 (gene HSPA1A) and the small molecule inhibitors of HSP70 and HSC70, VER and PIF. HSP70 expression was reduced by siRNA treatment to 33 % and 54 % of levels observed in scrambled siRNA treated cells at 24 hrs and 48 hrs post-transfection, respectively. Western blot analysis of HAZV N expression levels in HSP70 depleted cells 24 hrs and 48 hrs after transfection, revealed a ~15 % decrease in HAZV N protein levels compared to scrambled siRNA transfected cells. In agreement with this slight decrease in HAZV N expression, there was a modest decrease in the titre of secreted HAZV 48 hrs after transfection with HSP70 siRNA, compared to scrambled siRNA.

Depletion of HSP70 expression in cells did not result in an increase in HAZV replication, indicating the interaction between HSP70 and HAZV N was likely not detrimental to the HAZV replication cycle. Complete knock down of HSP70 could not be achieved in this study, but could result in a greater reduction in HAZV replication. Alternatively knock down of HSC70 could have a greater effect on the HAZV replication cycle, as has been shown for HCV (Khachatoorian, Ganapathy et al. 2014). Simultaneous knockdown of HSP70 and HSC70 could also have an additive effect on inhibition of HAZV replication.

Preliminary experiments with the HSP70 inhibitors VER and PIF, revealed a decrease in HAZV replication in the presence of these compounds; 16 µM VER reduced the titre of secreted HAZV by 95 %, and 0.5 µM PIF reduced the titre of secreted HAZV by 75 %, when compared to HAZV replication in the presence of 0.2 %

DMSO. This suggested a positive role for HSP70 within the HAZV replication cycle. The reduction in HAZV replication could not be solely attributed to the inhibition of HSP70/HSC70 however, as there are no reliably identified client proteins of HSP70, which are proteolytically degraded when HSP70/HSC70 is inhibited (Budina-Kolomets, Balaburski et al. 2014). Recent work has suggested that HSP90 client protein expression was reduced when HSP70 was inhibited, due to the closely linked nature of the HSP70 and HSP90 chaperone systems (Massey, Williamson et al. 2010). Expression levels of HSP90 client proteins could be analysed in the future to determine the efficacy of HSP70 inhibition, however effects on HAZV replication could potentially then be attributed to HSP90 inhibition. Interestingly HSP90 was also detected in the fraction containing infectious HAZV, which suggests HSP90 may also have a role within the HAZV replication cycle.

HSP70 and HSC70 are essential cellular ‘housekeeping’ proteins, therefore inhibiting HSC70/HSP70 function can be cytotoxic; in this study there was a 20 % reduction in cell viability in the presence of 16 µM VER, and PIF was cytotoxic at concentrations >5 µM. HSP70 inhibitors were designed to kill cancerous cells; several different types of cancer have upregulated expression levels of HSP70; knock down of HSP70, or inhibition of HSP70 results in the death of these cells (Nylandsted, Brand et al. 2000). Although it is thought that host directed antiviral therapeutics have the advantage of the slower mutation rate of the virus host, which should prevent virus escape mutants developing as quickly, a disadvantage of these therapeutics is that inhibition of cellular proteins can also be toxic to the host cell. The therapeutic window of the HSP70 inhibitors utilised in this study as anti-viral agents appeared to be relatively small. On the other hand the SW13 cell line utilised may depend on elevated expression levels HSP70 for survival; inhibition of HSP70 in a primary cell line may be less toxic.

HSP70 and HSC70 have been shown to directly aid virus replication at multiple stages of different virus replication cycles ranging from cell entry, genome replication, RNP formation and virion assembly (Mayer 2005). In the paramyxovirus, canine distemper virus, it was noted that induction of the stress response promoted the cytopathic effects of canine distemper virus infections, and that HSP70 and HSC70 associated with viral RNPs (Oglesbee, Ringler et al. 1990). The addition of anti-HSP70 antibodies reduced RNA polymerase activity associated with purified nucleocapsid particles, demonstrating the importance of the association between HSP70 and viral RNPs for genome transcription and replication (Oglesbee, Liu et al. 1996). More recently it was shown that the addition of anti-HSP70 antibodies to purified RSV polymerase complexes on lipid raft membranes also decreased RNA polymerase activity (Brown, Rixon et al. 2005). It was hypothesised that HSP70s are required to interact with and remodel the N protein to make the RNA accessible to the polymerase, or that HSP70 and HSC70 could directly chaperone the polymerase, and thereby enhance its activity (Mayer 2005). A similar mechanism could exist for CCHFV and HAZV.

HSP70 and HSC70 share significant (85 %) homology; HSC70 is the constitutively expressed homologue of HSP70. HCV NS5A interacts with both HSC70 and HSP70, and recently the differential roles of HSP70 and HSC70 in the replication cycle of HCV have begun to be dissected. siRNA mediated knockdown of both HSP70 and HSC70 resulted in a decrease in titre of secreted HCV; and simultaneous knock down of HSP70 and HSC70 had an additive effect on the decrease in HCV titre. Interestingly, HSP70 was shown to be involved in virus transcription and translation, whereas HSC70 was shown to have an important role in the assembly of virus particles (Khachatoorian, Ganapathy et al. 2014). This could also be the case during HAZV infection as HSC70, not HSP70 was identified in the fraction containing infectious HAZV.

In addition to direct involvement in the virus replication cycle, the interaction between HSP70 and viral proteins, can modulate cellular pathways regulated by HSP70. For example, IAV M1 protein interacts with HSP70, and by doing so disrupts the interaction between HSP70 and Apaf1, thereby enhancing the effects of apoptosis inducers; siRNA knock down of M1 during infection inhibited virus-induced apoptosis, but expression of M1 itself did not induce apoptosis (Halder, Bagchi et al. 2011).

There is currently a lack of effective therapeutics to treat CCHF; any insight gained into the replication cycle of CCHFV (and by inference, HAZV) could help reveal new strategies to combat these viruses. Members of the HSP70 family interact with viral proteins from multiple virus families, and are associated with or packaged into secreted virus particles. A growing number of studies are addressing the importance of HSP70 to the replication cycle of these viruses, using siRNA approaches, and small molecule inhibitors of HSP70 family members. A reduction in expression, or inhibition, of members of the HSP70 family resulted in a decrease in virus replication in many cases. This suggested that targeting HSP70 family members may be a promising strategy to inhibit replication of viruses from multiple different families, including CCHFV and HAZV.

CHAPTER 5: THE INDUCTION OF APOPTOSIS BY HAZV

5.1 CHAPTER INTRODUCTION

Apoptosis is a tightly regulated mechanism of controlled cell death, which can be activated in response to viral invasion, and is often a critical factor determining the outcome of infection. Hallmarks of cell death by apoptosis include expression of active caspases, DNA fragmentation, membrane ‘flipping’ (the expression of the normally inward facing phosphatidylserine on the plasma membrane outer leaflet), expression of Annexin V on the plasma membrane outer leaflet, cell shrinkage and membrane ‘blebbing’ as well as extensive protein cross linking (Elmore 2007). There are two main apoptotic pathways, the intrinsic pathway, which is activated in response to intracellular stimuli, and the extrinsic pathway, which is induced as a result of the activation of cell surface ‘death receptors’. Activation of various different cellular pathways results in the induction of apoptosis during virus infection, including activation of cell surface ‘death receptors’, the ER stress response and unfolded protein response (UPR), and activation of IRF 3, amongst others. Apoptosis can either be detrimental or beneficial to virus replication cycles, depending on the virus in question; in some cases apoptosis is considered to be an antiviral response to infection, with cell death limiting the spread of progeny virions. In other cases, viruses are thought to deliberately induce apoptosis to aid virus dissemination.

This chapter describes the activation of caspases-9, -3, and -7, and the induction of apoptosis by HAZV, over a 72 hr time course of infection. This chapter also describes the cleavage of HAZV N over this 72 hr time course, and investigations into the cellular proteins responsible for HAZV N cleavage using the inhibitor of caspase-3, Z-DEVD-FMK, and the inhibitor of executioner caspases, Z-FA-FMK. This chapter then goes on to examine the effect of these inhibitors (Z-DEVD-FMK and Z-FA-FMK) on HAZV replication.

5.1.1 Summary of the induction of apoptosis by CCHFV, and cleavage of CCHFV N by caspase-3

It has previously been reported that CCHFV induces apoptosis at a relatively late time point after infection; activation of procaspase-3 and TUNEL positive cells were detected 48 hrs after CCHFV infection, and increased at 72 hrs after infection, when major cytopathic effect (CPE) was observed (Karlberg, Tan et al. 2011). Induction of apoptosis by CCHFV was thought to occur after activation of the ER stress response and UPR; CCHFV was therefore thought to induce apoptosis via the intrinsic mitochondrial pathway (Figure 1.14) (Rodrigues, Paranhos-Baccala et al. 2012). During CCHFV infection it was shown that active caspase-3 cleaved CCHFV N at a conserved DEVD motif, located at the apex of the arm domain (Figure 5.1 (A)) to generate a 30 kDa and a 22 kDa CCHFV N fragment, that were detectable by western blot analysis (Karlberg, Tan et al. 2011). Cleavage of CCHFV N occurred in infected cells during apoptosis, when CCHFV N was transfected into cells and apoptosis was induced, and *in vitro* using recombinant CCHFV N and recombinant caspase-3 expressed in *E. coli*. After cleavage by caspase-3, the N- and C-terminal CCHFV N fragments remained associated with each other, and eluted from a size exclusion column in a single peak (Carter, Surtees et al. 2012).

The CCHFV N DEVD motif is conserved among all currently sequenced CCHFV isolates and site directed mutagenesis confirmed the DEVD motif as the specific site for caspase-3 cleavage; however during infection, only a small proportion of total CCHFV N was cleaved by caspase-3. It could be that, given the location of the DEVD motif on the apex of the arm domain, only monomeric CCHFV N is a substrate for active caspase-3. The majority of CCHFV N is thought to be multimerised and bound to RNA in an infected cell and the crystal structure suggests that the DEVD motif is inaccessible in RNPs. Nevertheless, inhibition of caspase-3 using Z-DEVD-FMK resulted in an increase in secreted CCHFV titre, which suggested that cleavage of CCHFV N by caspase-3, and the induction of apoptosis, formed part of the cellular defence against CCHFV infection.

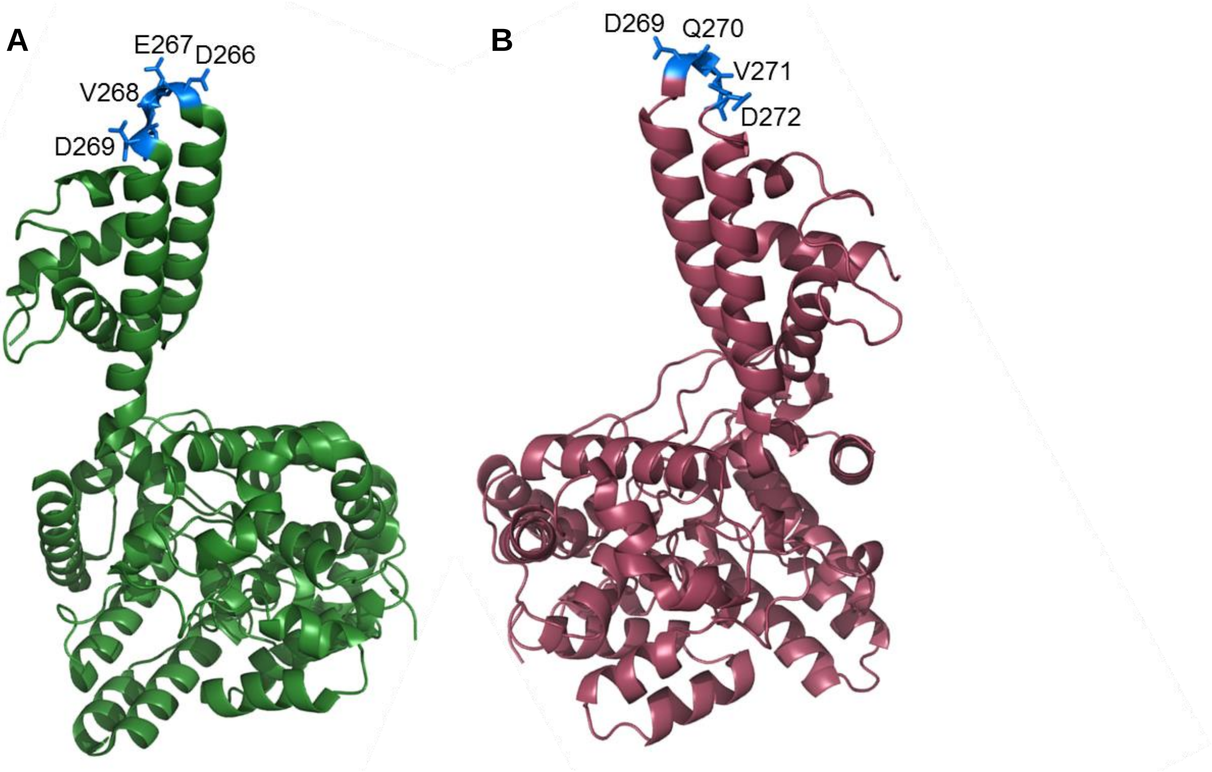


Figure 5.1 CCHFV N and HAZV N with DEVD and DQVD motifs highlighted

CCHFV N (A) contains a conserved DEVD (Asp, Glu, Val, Asp) motif on the apex of the arm domain (blue) that is cleaved by active caspase-3 during apoptosis to generate two fragments of 30 kDa and 22 kDa. HAZV N (B) possesses a DQVD (Asp, Gln, Val, Asp) motif on the apex of the arm domain (blue), in a similar position to the DEVD motif on the apex of the arm domain of CCHFV N.

5.1.2 The crystal structure of HAZV N revealed a DQVD motif on the apex of the arm domain

Comparison of CCHFV N and HAZV N (Figure 1.12) primary sequences revealed that HAZV N did not possess a DEVD motif, but instead contained a DQVD (Asp, Gln, Val, Asp) sequence. Analysis of the crystal structure of HAZV N showed this DQVD sequence resided on the apex of the arm domain, in approximately the same position as the DEVD motif of CCHFV N (Figure 5.1 (B)). DQVD is not a characterized caspase-3 cleavage motif, however there are many uncharacterized caspase-3 cleavage motifs, whose amino acid composition has yet to be determined. Caspases recognize a tetra-peptide repeat Asp-x-x-Asp, therefore DQVD could potentially represent a currently uncharacterized caspase-3 cleavage site. Similar to CCHFV N, the exposed position of the DQVD motif on the apex of the arm domain suggested the motif would be accessible to cellular caspases during infection. However little is known about the effect of HAZV replication on the host cell environment; and it is not known whether HAZV infection results in activation of cellular caspases or the induction of apoptosis. Experiments were therefore carried out to establish whether cellular caspases were activated during HAZV infection, and whether this led to the induction of apoptosis and cleavage of HAZV N.

5.2 RESULTS: HAZV INFECTION RESULTED IN ACTIVATION OF CELLULAR CASPASES AND THE INDUCTION OF APOPTOSIS

5.2.1 HAZV infection resulted in activation of caspase-9, -3, and -7

Caspases are constitutive components of mammalian cells that are widely expressed in most cells in an inactive proenzyme form. The caspase family of proteins can be broadly divided into three categories; the initiator caspases (capase-2, -8, -9 and -10), the executioner caspases (caspase-3, -6, and -7) and the inflammatory caspases (capase-1, -4 and -5). Upon activation of the intrinsic or extrinsic apoptotic pathways, the initiator caspases are cleaved and activated; some procaspases can aggregate and autoactive (self-cleave), other procaspases are cleaved by active initiator caspases. Once activated, one of the main proteolytic targets of the initiator caspases are the executioner procaspases; procaspase-3 for example can be cleaved by any of the initiator caspases. Cleavage of executioner procaspases results in their activation; active executioner caspases then cleave both other executioner procaspases, and some initiator procaspases. This proteolytic cascade amplifies the apoptotic signal and leads to rapid cell death (Elmore 2007). In addition to procaspases, active executioner caspases have multiple other proteolytic targets; for example caspase-3 and -7 cleave poly-ADP ribose polymerase (PARP) and inhibitor of caspase activated DNase (ICAD).

To determine whether HAZV infection resulted in cleavage and activation of initiator or executioner caspases, SW13 cells were infected with HAZV at an MOI 1, or mock infected with SFM. At 24, 48 and 72 hrs post infection, cells were harvested; both adherent cells and culture media containing detached cells were harvested together. Following cell lysis, total protein concentration was determined by BCA assay, and 20 µg protein loaded per well after samples were denatured in LDS buffer for 10 minutes. Following transfer of proteins to FL-PVDF membranes and blocking, membranes were probed with primary antibodies that targeted the following proteins; procaspase-3, cleaved caspase-3, procaspase-7, cleaved caspases-7, procaspase-9, cleaved caspase-9, PARP, cleaved PARP, anti-HAZV N and GAPDH. Primary antibodies were detected using fluorescently conjugated secondary antibodies, and membranes were analysed with a LiCor Odyssey Sa Infrared imaging system in the 700 and 800 channels. The results of this analysis are shown in Figure 5.2; HAZV N was detected with anti-HAZV N 24, 48, and 72 hrs after infection.

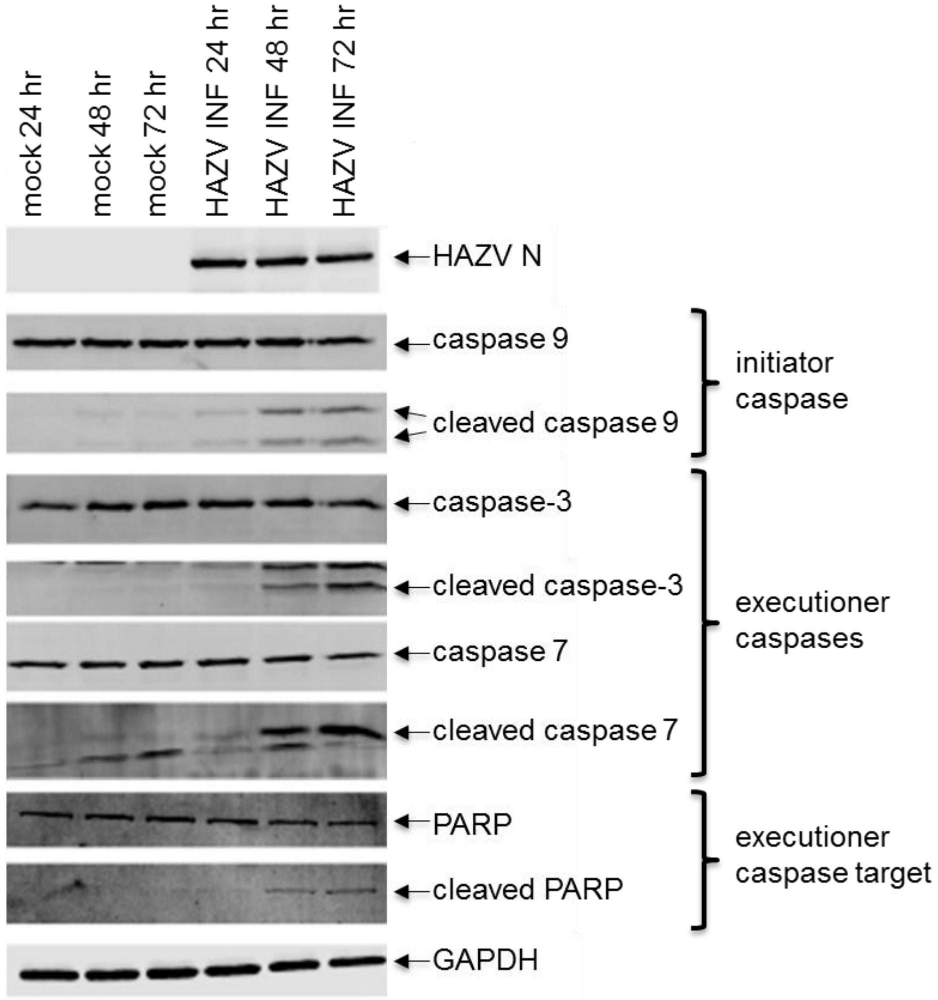


Figure 5.2 HAZV infection resulted in activation of procaspases-3, -7, and -9

HAZV infected or mock infected cells were harvested 24, 48, and 72 hrs after infection, and lysates were analysed by western blot for the proenzyme and active form of several cellular caspases including caspases-3, -7, and -9. Abundance of uncleaved and cleaved PARP was also analysed, and GAPDH expression levels were used as a loading control. Anti-HAZV N confirmed successful HAZV replication.

The constitutive proenzymes of caspases-3, -7, and -9 were detected in both mock infected and HAZV infected cells 24, 48, and 72 hrs after infection. Procaspases-3, -7 and -9 were detected in similar abundances in all samples, and expression levels of these proteins remained relatively constant during the time course of the experiment in mock and HAZV infected cells. In contrast protein levels of cleaved, and therefore active, caspases-3, -7 and -9 increased dramatically 48 and 72 hrs post HAZV infection, compared to mock infected cells. Distinct bands corresponding to the correct molecular weight of cleaved caspases-3, -7 and -9 are detected in HAZV infected cells 48 and 72 hrs post infection, but are absent from, or present at very low levels in mock infected cell lysates at these time points. This indicated that HAZV infection specifically resulted in the activation of these enzymes, rather than their activation being a product of prolonged incubation in 12 well dishes (up to 72 hrs). There appeared to be a greater abundance of cleaved caspase-3 and caspase-7 at 72 hrs, compared to 48 hrs post HAZV infection, whereas expression levels of active caspases-9 did not appear to increase to the same extent at 72 compared to 48 hrs post infection. The increase in abundance of active caspase-3, -7, and -9 in HAZV infected cells at 48 and 72 hrs post infection could not be attributed to varying expression levels of the constitutive, proenzymes of these caspases, as the proenzymes are present in similar abundances in HAZV infected and mock infected cells.

Expression levels of one of the proteolytic targets of caspase-3 and caspase-7, PARP, was also analysed, to confirm activated caspase-3 and caspases-7 were functional in HAZV infected cells. Similar to the expression profiles of procaspases-3, 7-, and -9, levels of uncleaved PARP remained relatively constant in both mock and HAZV infected cells 24, 48, and 72 hrs after infection, whereas the abundance of cleaved PARP increased dramatically 48 and 72 hrs after HAZV infection, compared to mock infected cells. This indicated that active caspase-3 and -7 were functional, and cleaved at least one of their proteolytic targets. In all samples expression levels of GAPDH remained constant therefore changes in abundance of cellular proteins could not be attributed to unequal protein loading.

Taken together, these data indicated that HAZV infection resulted in the cleavage and activation of the initiator procaspase-9, and the executioner procaspase-3 and procaspase-7, 48 and 72 hrs after HAZV infection. Cleaved PARP was also detected in HAZV infected cells 48 and 72 hrs after infection, which indicated active caspase-3 and caspase-7 were functional.

5.2.2 HAZV infection resulted in DNA fragmentation

To confirm that the presence of cleaved PARP and active caspases-9, -3 and -7 in HAZV infected cells correlated with the induction of apoptosis, TUNEL (terminal dUTP nick-end labelling) assays were carried out 24, 48, and 72 hrs after HAZV infection or mock infection. The TUNEL assay detects intrachromosomal DNA strand breaks, one of the final stages of apoptosis that results in cell death. SW13 cells were grown on glass coverslips prior to HAZV infection at MOI 1, or mock infection; at 24, 48 and 72 hrs cells were fixed in 4 % paraformaldehyde, and after permeabilisation, TUNEL assays were carried out according to the manufacturer's instructions. TUNEL assay analysis was performed on the Zeiss LSM 510 laser scanning confocal microscope; 5 images were taken of HAZV-infected and mock-infected cells at each time point, and the total number of cells was used to calculate the number of TUNEL positive nuclei relative to DAPI-stained nuclei at 24, 48, and 72 hrs post infection. The results of this analysis are shown in Figure 5.3 (B), along with representative confocal images at 24, 48, and 72 hrs post HAZV- or mock infection (Figure 5.3 (A)).

Confocal microscopy analysis revealed HAZV infection resulted in a greater proportion of TUNEL positive cells compared to mock-infected cells at 48 and 72 hrs post-infection. At 24 hrs post HAZV infection, the relative proportion of TUNEL positive cells compared to DAPI positive nuclei was 2.7 %, this increased to 5.3 % 48 hrs post infection, and by 72 hrs post infection, 23 % of cells in wells that had been infected with HAZV were TUNEL positive. The increase in the proportion of TUNEL positive nuclei at each time point indicated HAZV infection resulted in apoptosis at 48 hrs post infection, which was more pronounced at 72 hrs post infection. In contrast the relative proportion of TUNEL positive nuclei compared to DAPI positive nuclei in mock infected cells was 5.1 % at 24 hrs, this decreased to 2.3 % 48 hrs post mock infection, and by 72 hrs post mock infection, the relative proportion of TUNEL positive nuclei was 3 % in mock infected cells. The greater proportion of TUNEL positive nuclei (23 % nuclei) in HAZV infected cells compared to mock infected cells (3 % nuclei) at 72 hrs post infection indicated HAZV infection resulted in apoptosis at this time point; 48 hrs after infection the difference between the number of TUNEL positive cells in HAZV infected wells compared to mock was not as great (5.3 % compared to 2.3 %), possibly indicating the induction of apoptosis was not as pronounced at this time point.

This study did not investigate whether the induction of apoptosis was a direct result of HAZV infection, or an indirect consequence of cellular processes (such as the IFN response) activated by HAZV infection. These data indicated that HAZV infection resulted in the induction of apoptosis at a relatively late time point after infection, in a similar manner to CCHFV. A greater proportion of TUNEL positive cells were detected 48 and 72 hrs after HAZV infection compared to mock-infected cells; active caspase-3, caspase-7 and cleaved PARP were also increased in abundance in HAZV infected cells compared to mock infected cells at these time points. There may have been a slight increase in abundance of the initiator capase-9 24 hrs post HAZV infection compared to mock-infected cells, however activation of the executioner caspases and increased detection of intrachromosomal DNA strand breaks were not observed until 48 hrs after HAZV infection.

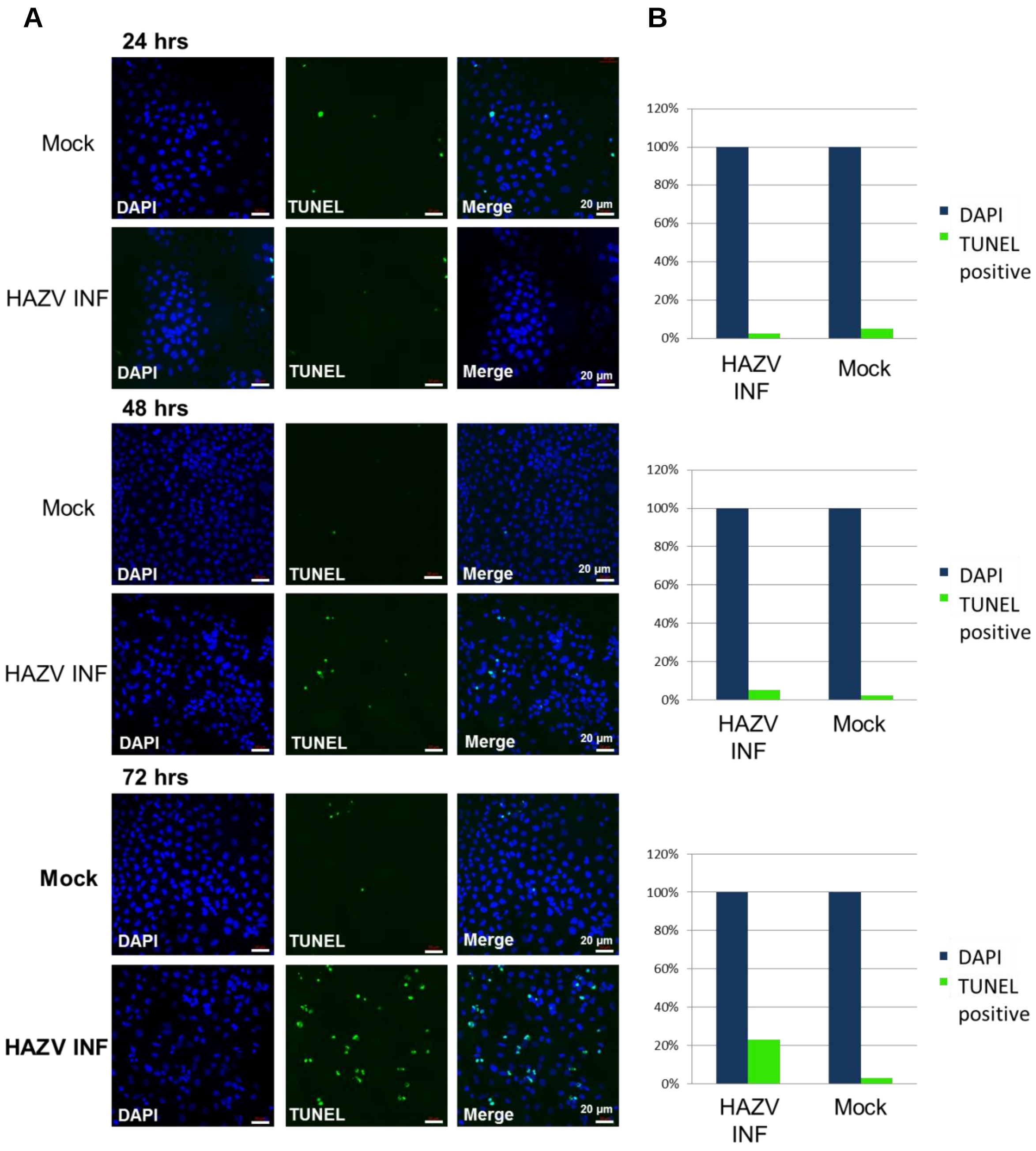


Figure 5.3 TUNEL assay confirmed the induction of apoptosis in HAZV infected cells 48 hrs and 72 hrs post HAZV infection

HAZV infection resulted in a greater proportion of TUNEL positive cells 48 and 72 hrs after infection, compared to mock infected cells ((A) and (B)). In all cases green nuclei represent TUNEL positive cells. This indicated HAZV infection induced apoptosis at relatively late time points, similar to CCHFV infection.

5.3 RESULTS: THE CLEAVAGE OF HAZV N OVER THE 72 HR TIME COURSE OF INFECTION

HAZV infection of SW13 cells resulted in the activation of caspases-9, -3, and -7, cleavage of PARP, and the induction of apoptosis 48 and 72 hrs after infection. To investigate whether HAZV N was cleaved over this 72 hr time course, SW13 cells were infected with HAZV at an MOI 1, or mock infected with SFM. Cells were lysed and proteins were harvested at 24, 48, and 72 hrs post infection; proteins were quantified by BCA assay, then 5 µg protein loaded per lane and resolved by SDS PAGE prior to western blot analysis, as described in section 2.2.6.5. Membranes were probed with primary anti-HAZV N and anti-GAPDH, which were detected with fluorescently conjugated secondary antibodies and analysed using the LiCor Odyssey Sa Infrared imaging system.

Western blot analysis revealed that HAZV N was cleaved over the 72 hr time course of infection, and several cleavage products were detected (Figure 5.4). The predominant protein species detected at all time points is a band of ~54 kDa, which corresponds to the predicted molecular weight of full length HAZV N (Figure 5.4, arrow). At 24 hrs post infection a cleavage product of ~40 kDa was observed (Figure 5.4 *), and a very weak band corresponding to a molecular weight of ~20 kDa was faintly visible (Figure 5.4 **). This band with an approximate molecular weight of ~20 kDa increased in intensity at 48 and 72 hrs post infection, which suggested this HAZV N cleavage product increased in abundance at these time points. A band with an approximate molecular weight of ~ 30 kDa was also detected at 48 hrs and 72 hrs post infection (Figure 5.4 ***). Bands of corresponding molecular weights were absent from mock infected cells, which suggested these were cleavage products of HAZV N, rather than cross reacting cellular proteins. The increase in abundance of the ~20 kDa band corresponded with the induction of apoptosis, described in the previous section. This suggested the ~20 kDa cleavage product may be generated by proteins that are activated, or have increased expression levels, during apoptosis.

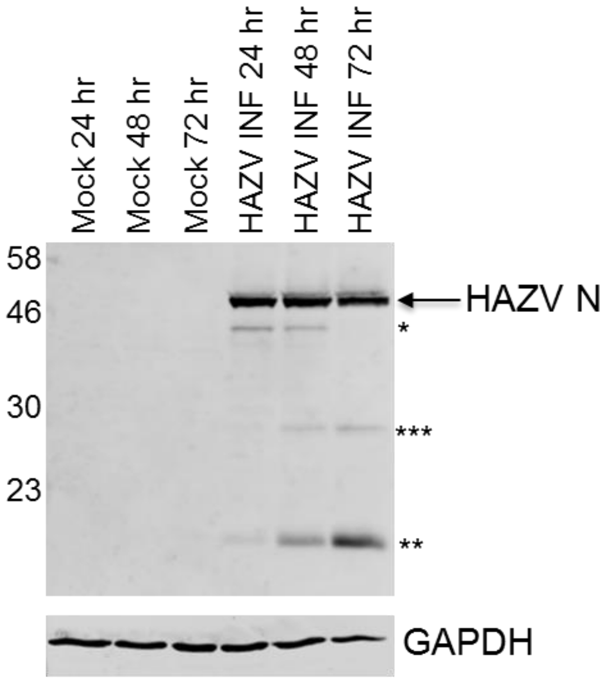


Figure 5.4 Cleavage products of HAZV N over a 72 hr time course of infection

Several HAZV N cleavage products were detected with anti-HAZV N in cell lysates harvested 24, 48, and 72 hrs post HAZV infection. HAZV N cleavage products with approximate molecular weights of 20 kDa (**) and 30 kDa (***) were increased in abundance 48 and 72 hrs post infection, and a cleavage product with an approximate molecular weight of 40 kDa (*) was detected 24 hrs post infection.

5.3.1 Treatment of cells with the caspase-3 inhibitor (Z-DEVD-FMK) and the executioner caspase inhibitor (Z-FA-FMK)

Cleavage products of HAZV N were detected in infected cell lysates with anti-HAZV N over the 72 hrs time course of infection; however, the cellular factor required to produce these cleavage products and the specific cleavage site in HAZV N was unknown. To determine whether, in a similar manner to CCHFV N, HAZV N could be cleaved by active caspase-3 at 48 and 72 hrs post infection, SW13 cells were treated with the caspase-3 inhibitor Z-DEVD-FMK during a 72 hr time course of HAZV infection. Cells were also treated with Z-FA-FMK, an inhibitor of the executioner caspases, to confirm the results obtained with Z-DEVD-FMK. Z-DEVD-FMK and Z-FA-FMK are both irreversible cysteine protease inhibitors; these compounds are fluoromethyl ketone derivatives of peptides that resemble the cleavage sites of known caspase substrates, and irreversibly alkylate the cysteine residues in the active site of caspases (Schotte, Declercq et al. 1999). Z-DEVD-FMK shows greater specificity for caspase-3, whereas Z-FA-FMK can inhibit the DEVD-ase activity of caspase-2, -3, -7, and -6 (Lopez-Hernandez, Ortiz et al. 2003).

Initially, uninfected SW13 cells were treated with 20 μM and 40 μM Z-DEVD-FMK and Z-FA-FMK; 20 μM was the concentration used to treat SW13 cells during CCHFV infection in the study carried out by (Karlberg, Tan et al. 2011), and 40 μM was utilized to determine whether a dose dependent effect could be observed. 24, 48, and 72 hrs after treatment with these compounds, an MTT assay was performed (as described in section 2.2.12) to assess cell viability. Cell viability levels relative to cells treated with 0.1 % DMSO were determined for SW13 cells treated with 20 μM or 40 μM of each compound (Figure 5.5).

Growth of SW13 cells in the presence of 20 μM or 40 μM Z-DEVD-FMK resulted in cell viability levels that were ~80 % the level of cells treated with 0.1 % DMSO at each time point, when measured by MTT assay. A similar result was obtained when SW13 cells were treated with Z-FA-FMK, which indicated these compounds might have slightly affected cell growth or metabolism. However little difference was observed in viability levels between cells grown in the presence of 20 μM Z-DEVD-FMK or Z-FA-FMK and cells grown in the presence of 40 μM Z-DEVD-FMK or Z-FA-FMK, indicating the effect these compounds had on cell growth or metabolism was not dose-responsive.

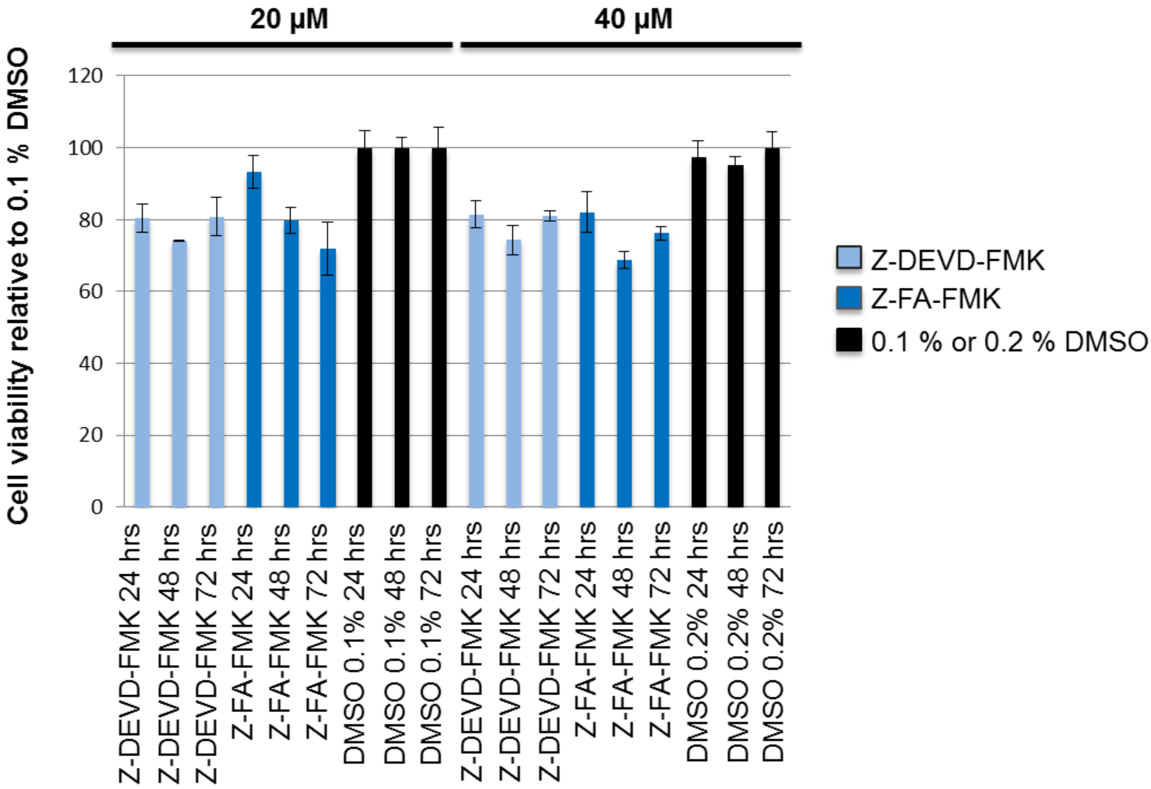


Figure 5.5 Cell viability in the presence Z-DEVD-FMK or Z-FA-FMK

Cells were treated with either 20 µM or 40 µM Z-DEVD-FMK, or Z-FA-FMK, or cells were treated with 0.1 % or 0.2 % DMSO in triplicate. 24, 48, and 72 hrs later, cell viability was examined by MTT assay; viability levels of cells treated with 20 µM or 40 µM Z-DEVD-FMK, or Z-FA-FMK are expressed relative to viability levels of cells treated with 0.1 % DMSO. The graph represents the average of each of the 3 absorbance readings.

SW13 cells were then infected with HAZV at an MOI 1, or mock infected; after adsorption of HAZV to cells for 1 hr at 37 °C, SFM containing HAZV was removed and replaced with DMEM containing 2 % FBS and the appropriate concentration of Z-DEVD-FMK or Z-FA-FMK. HAZV was allowed to replicate in the presence of these compounds for 24, 48, or 72 hrs, then proteins were harvested as previously described (section 2.2.6.1). Lysates of cells mock infected, or infected with HAZV that had been replicating in the presence of Z-DEVD-FMK or Z-FA-FMK for 24, 48 or 72 hrs were analysed by western blot using antibodies that targeted HAZV N, GAPDH, or cleaved PARP. Expression levels of cleaved PARP at 72 hrs post infection were examined to confirm the inhibition of caspase-3, or the executioner caspases; expression levels of cleaved PARP at 72 hrs were quantified with StudioImage software, as previously described (Section 4.9.1) and plotted relative to GAPDH levels (Figure 5.6 (D)).

The abundance of cleaved PARP was decreased in the presence of 20 µM or 40 µM Z-DEVD-FMK or Z-FA-FMK (on average 16 % and 11 % the level of GAPDH, respectively), compared to HAZV infected cells treated with 0.1 % or 0.2 % DMSO (on average 25 % the level of GAPDH) (Figure 5.6 (D)). This indicated PARP cleavage was impaired in the presence of both Z-DEVD-FMK and Z-FA-FMK, and that less PARP was cleaved in the presence of Z-FA-FMK. These observations were consistent with reports that PARP can be cleaved by both active caspase-3 and caspase-7. Inhibition of PARP cleavage was not complete even using 40 µM Z-FA-FMK, which suggested either incomplete inhibition of the executioner caspases, or PARP cleavage by another cellular protein. There was not a greater decrease in the amount of cleaved PARP produced in the presence of 40 µM Z-DEVD-FMK or Z-FA-FMK, compared to 20 µM Z-DEVD-FMK or Z-FA-FMK, which indicated that increasing the concentration of these compounds did not result in a greater inhibition of caspase-3 or the executioner caspases.

Examination of HAZV N cleavage in the presence of 20 µM or 40 µM Z-DEVD-FMK did not reveal an obvious decrease in abundance of the ~20 kDa cleavage product compared to cells treated with 0.2 % DMSO at 48 and 72 hrs post infection (Figure 5.6 (B) and (C)). This suggested HAZV N was not cleaved by active caspase-3. Interestingly, examination of HAZV N cleavage products produced in the presence of either 20 µM or 40 µM Z-FA-FMK, revealed an almost complete abolishment of the ~20 kDa band compared to cells treated with 0.2 % DMSO at 48 and 72 hrs post infection (arrow, Figure 5.6). This suggested that although HAZV N was possibly not cleaved by active caspase-3, HAZV N could instead have been cleaved by either another executioner caspase (caspases-6 or -7), or the downstream product of one of these other executioner caspases.

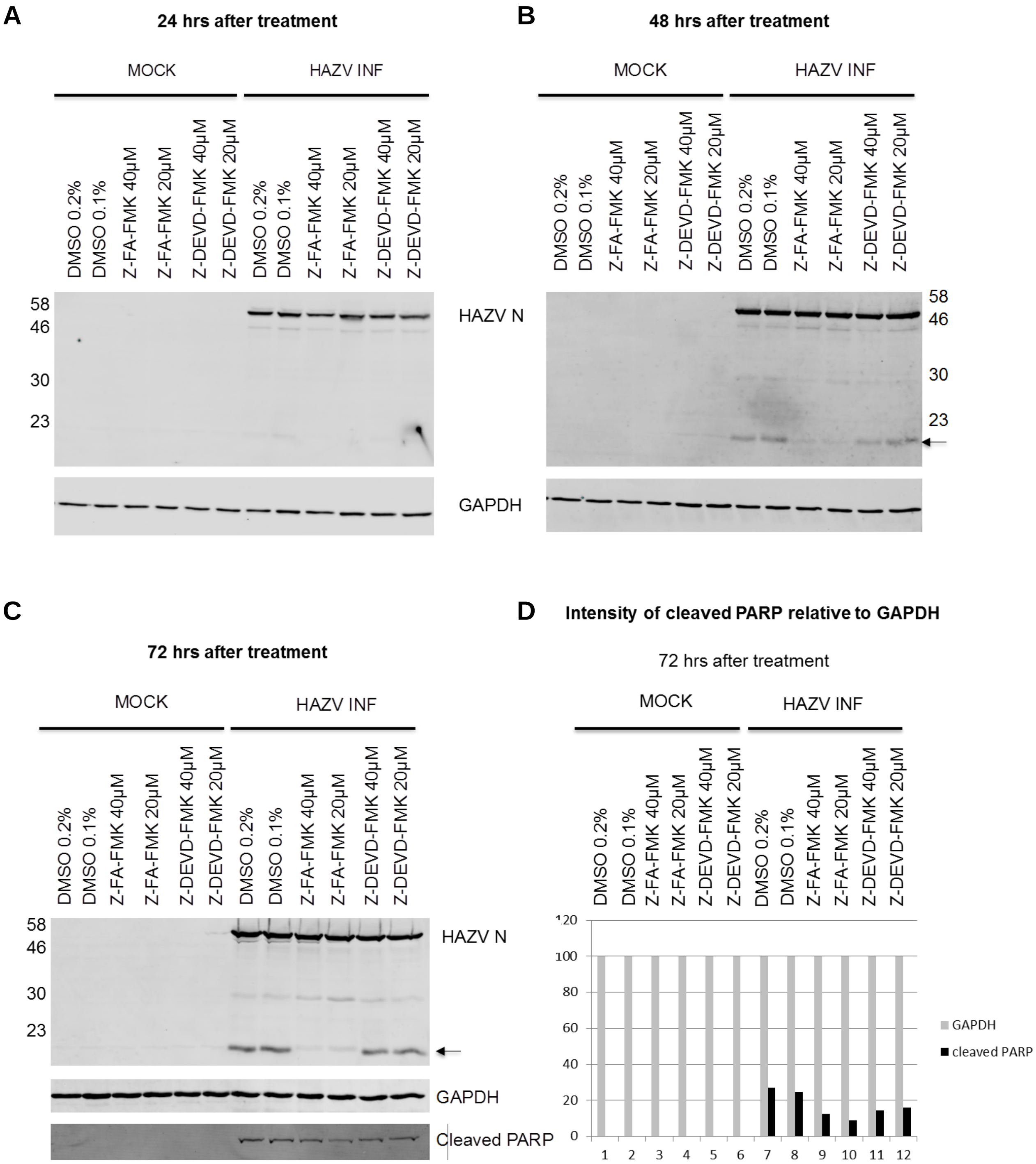


Figure 5.6. Cleavage products of HAZV N detected in the presence of 20 µM or 40 µM Z-DEVD-FMK or Z-FA-FMK.

HAZV N was cleaved during infection, with the cleavage product of ~20 kDa becoming more abundant 48 and then 72 hrs after infection. Inhibition of caspase-3 with 20 µM or 40 µM Z-DEVD-FMK resulted in a reduction in cleaved PARP levels, compared to cells treated with 0.2 % DMSO, and inhibition of the executioner caspases with Z-FA-FMK resulted in a slightly greater decrease in cleaved PARP *levels compared to cells treated with 0.2 % DMSO (D). Interestingly, treatment of cells with Z-FA-FMK resulted in an almost complete abolishment of the ~20 kDa HAZV N cleavage product compared to cells treated with 0.2 % DMSO (B) and (C). Treatment of cells with the caspase-3 inhibitor on the other hand, did not obviously decrease the intensity of the ~20 kDa HAZV N cleavage product band. This indicated HAZV N might have been primarily cleaved by either other executioner caspases, or their downstream substrates during infection, rather than by active caspase-3.*

5.3.2 The effect on HAZV replication of Z-DEVD-FMK and Z-FA-FMK

Treatment of cells with Z-FA-FMK prevented the production of the ~20 kDa cleavage product of HAZV N, observed 48 and 72 hrs after HAZV infection. This suggested that HAZV N may have been cleaved by one of the other executioner caspases, or one of their downstream targets during the 72 hrs time course of infection. Next, preliminary experiments were carried out to examine the effects of the presence of 20 µM Z-DEVD-FMK and Z-FA-FMK on HAZV replication. As previously described (section 2.2.9.1) SW13 cells were infected with HAZV at an MOI 1, and after adsorption to cells for 1 hr at 37 °C, SFM containing HAZV was removed and replaced with media containing 20 µM Z-DEVD-FMK or Z-FA-FMK. HAZV was allowed to replicate for 24, 48, or 72 hrs in the presence of these compounds; cell culture supernatants were then harvested, and the amount of infectious HAZV present estimated by plaque assay, as described in section 2.2.9.2. The estimated infectious titre of HAZV secreted from cells grown in 20 µM Z-DEVD-FMK or 20 µM Z-FA-FMK, or grown in the absence of these compounds was plotted against the number of hours of HAZV replication, and is shown in Figure 5.7.

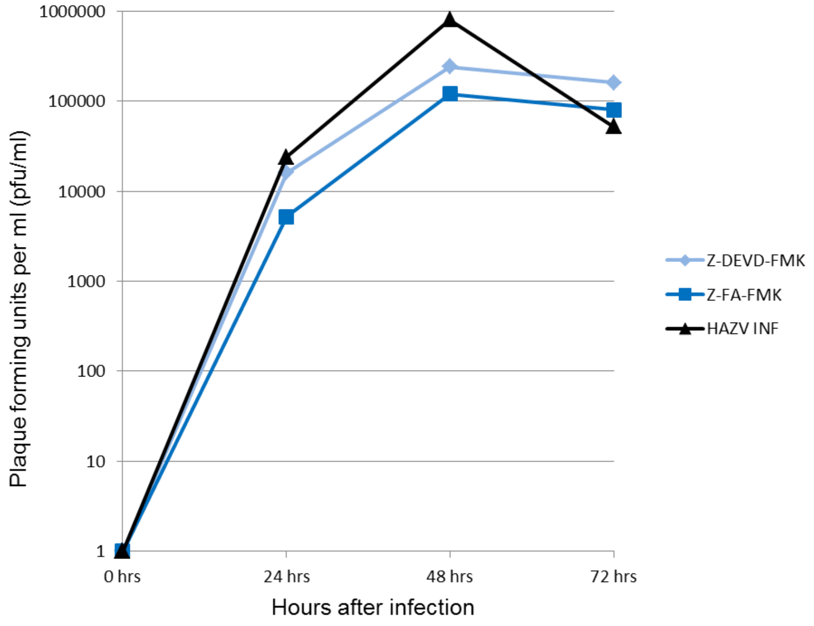


Figure 5.7 Titre of HAZV secreted from cells treated with Z-DEVD-FMK or Z-FA-FMK

SW13 cells were infected with HAZV at MOI 1, and HAZV was then allowed to replicate in the presence of 20 µM Z-DEVD-FMK or Z-FA-FMK, or in the absence of these compounds. Cell culture supernatants were harvested 24, 48, and 72 hrs after infection, and the titre of secreted HAZV was estimated once by plaque assay, then plotted against time after infection.

The titre of infectious HAZV secreted from SW13 cells in the presence of these compounds does not increase 24 and 48 hrs after infection, compared to untreated cells. This suggested that the activation of cellular caspases and cleavage of HAZV N observed 48 hrs after infection was not detrimental to virus infection, as secreted HAZV titre did not increase when these processes were impaired. Plaque assays were performed once and gave the following results: 24 hrs after infection, 1.5 fold less infectious HAZV was secreted into the supernatant of cells treated with 20 µM Z-DEVD-FMK and compared to untreated cells (1.6 x 10^4 pfu/mL compared to 2.4 x 10^4 pfu/mL), and 4.6 fold less infectious HAZV was secreted into the supernatant of cells treated with Z-FA-FMK, compared to untreated cells (0.52 x 10^4 pfu/mL compared to 2.4 x 10^4 pfu/mL). The titre of secreted HAZV from cells treated with Z-DEVD-FMK and Z-FA-FMK increased 48 hrs after infection (2.4 x 10^5 pfu/mL and 1.2 x 10^5 pfu/mL, respectively), but remained below the levels of HAZV secreted from untreated cells (8 x 10^5 pfu/mL). In contrast, at 72 hrs post infection, the titre of infectious HAZV secreted from untreated cells (5.2 x 10^4 pfu/mL) had decreased to slightly below the levels of HAZV secreted from cells grown in the presence of Z-DEVD-FMK or Z-FA-FMK (1.6 x 10^5 pfu/mL and 8 x 10^4 pfu/mL, respectively).

The decrease in titre of HAZV secreted from cells grown in the presence of these compounds at 24 and 48 hrs post infection could be a reflection of the slight decrease in cell viability observed when SW13 cells were treated with Z-DEVD-FMK or Z-FA-FMK. However overall, these data suggested that activation of cellular caspases and cleavage of HAZV N was not detrimental to the replication cycle of HAZV.

5.4 CHAPTER SUMMARY AND DISCUSSION

This chapter presents the activation of procaspases-9, -3, and -7, and the induction of apoptosis during the replication cycle of HAZV, 48 and 72 hrs after HAZV infection. This chapter also presents the cleavage of HAZV N during the replication cycle of HAZV, and the inhibition of this cleavage with the executioner caspase inhibitor Z-FA-FMK, as well as the effect of the caspase-3 inhibitor Z-DEVD-FMK, and Z-FA-FMK on the replication cycle of HAZV.

The crystal structure of HAZV N, presented in chapter 3, revealed the presence of a DQVD motif on the apex of the arm domain of HAZV N, in a similar position to a DEVD motif on the apex of the arm domain of CCHFV N, which is cleaved by active caspase-3 during the CCHFV replication cycle. The DQVD motif of HAZV N is not a characterized caspase-3 cleavage site, however this motif contains the minimum requirement for cleavage by caspase-3, which recognizes Asp-X-X-Asp. In this chapter HAZV infection was shown to result in the activation of the procaspases-9, -3, and -7 by western blot analysis 48 and 72 hrs after HAZV infection. In addition, protein levels of cleaved PARP (one of the proteolytic targets of active caspase-3 and -7) increased in abundance in the lysates of HAZV infected cells 48 and 72 hrs after infection, which indicated active caspase-3 and -7 were functional, and effectively cleaved their target substrates.

The induction of apoptosis in HAZV infected cells 48 and 72 hrs after infection was confirmed using a TUNEL assay, which detected intrachromosomal DNA strand breaks, and indicated that one of the final stages of apoptosis had been initiated. The percentage of TUNEL positive nuclei relative to DAPI positive nuclei in HAZV infected cells 48 and 72 hrs after HAZV infection was 5.3 % and 23 %, compared to mock infected cells, which was 2.3 % and 3 % at the same time points. This indicated the induction of apoptosis was specific to HAZV infected cells. TUNEL positive nuclei were detected by microscopy analysis using a Zeiss LSM 510 confocal microscope; due to the method used to prepare these cells (growth of SW13 cells on glass coverslips, infection with HAZV or mock infection, followed by cell fixation, permeabilisation and TUNEL assay) a proportion of non-adherent TUNEL positive cells would have been removed with the media at the cell fixation stage. The results presented in Figure 5.3 could therefore be an underestimation of the percentage of TUNEL positive cells induced during HAZV infection.

Multiple viruses induce apoptosis during their replication cycles, including several other members of the *Bunyaviridae* family; such as the orthobunyaviruses LACV, oropouche virus (OROV) and BUNV and the hantaviruses, Hantaan, Sin Nombre, Andes, and Seoul virus (Pekosz, Phillips et al. 1996; Kang, Park et al. 1999; Kohl, Clayton et al. 2003; Markotic, Hensley et al. 2003; Acrani, Gomes et al. 2010). In the case of BUNV, apoptosis was linked to activation of IRF-3 and downstream IRF-3 signalling; apoptosis was induced 48 hrs after infection with BUNV and was inhibited at earlier time points by the activity of the BUNV NSs protein. CCHFV has also been shown to delay expression of IFN and IFN stimulated genes, possibly through interfering with IRF-3 activity, until 48 hrs after infection (Andersson, Karlberg et al. 2008). It could be the delay in apoptosis until 48 hrs after CCHFV infection is linked to the delayed activation and function of IRF-3 and IFN during CCHFV replication. CCHFV induced apoptosis has however, also been linked to ER stress and the UPR (Rodrigues, Paranhos-Baccala et al. 2012).

HAZV induced apoptosis at the same relatively late time point as CCHFV (48 hrs after infection), therefore HAZV may also possess a mechanism to delay the induction of apoptosis until this time point. Interestingly, 24 hrs after HAZV infection or mock infection, there was a higher proportion of TUNEL positive nuclei in mock infected cells (5.1 %) compared to HAZV infected cells (2.7 %); it is possible HAZV is able to counteract some of the possibly-stress induced cell death observed at this time point. This study did not examine whether apoptosis induced by HAZV was a direct result of HAZV infection and replication, or a consequence of indirect processes induced by HAZV infection, such as activation of the IFN response. Interestingly the induction of apoptosis by LACV was linked to inhibiting the anti-apoptotic activity of HSC70, and the M1 protein of IAV exerts a proapoptotic effect by binding to HSP70 and disrupting the interaction between HSP70 and Apaf-1 (Colon-Ramos, Irusta et al. 2003; Halder, Bagchi et al. 2011). The contribution towards the induction of apoptosis, if any, of HAZV N interacting with HSP70/HSC70 is unknown; however it is possible that HAZV N binding to HSP70 may enhance the apoptotic consequences of HAZV infection.

Western blot analysis revealed that during a 72 hrs time course of HAZV infection, HAZV N was cleaved in infected cells; cleavage products of HAZV N were detected with anti-HAZV N 24, 48 and 72 hrs after infection. An HAZV N cleavage product of ~20 kDa increased in intensity over the time course of infection, which suggested this ~20 kDa cleavage product increased in abundance at 48 and 72 hrs post infection. As this increase in abundance corresponded with the induction of apoptosis, it could be that one of the cellular proteins that was activated or upregulated during apoptosis was responsible for cleavage of HAZV N to generate the ~20 kDa cleavage product, in an analogous fashion to the cleavage of CCHFV N by active caspase-3. Although the site at which HAZV N was cleaved to generate the ~20 kDa cleavage product was not identified in this study, cleavage of HAZV N at the DQVD motif is predicted to result in HAZV N fragments of ~30 kDa and 23 kDa; it is possible the observed HAZV N ~20 kDa fragment could be derived from HAZV N cleavage at the DQVD motif, and migrates with an apparent lower molecular weight by SDS PAGE.

SW13 cells were infected with HAZV in the presence of the caspase-3 inhibitor Z-DEVD-FMK, and executioner caspase inhibitor Z-FA-FMK, to examine the possible role of cellular caspases in the cleavage of HAZV N. Z-DEVD-FMK and Z-FA-FMK are both fluoromethyl ketone derivatives of peptides that resemble the cleavage site of known caspase substrates, and irreversibly inhibit caspases by alkylating the cysteine residue in the caspase active site (Schotte, Declercq et al. 1999; Lopez-Hernandez, Ortiz et al. 2003). Analysis of levels of cleaved PARP 72 hrs after HAZV infection in the presence of these compounds revealed a decrease in the levels of cleaved PARP in cells that that been treated with both Z-DEVD-FMK and Z-FA-FMK compared to untreated cells, which suggested these compounds were impairing the function of caspase-3 and the executioner caspases. A greater decrease could be observed in the expression levels of cleaved PARP in HAZV infected cells treated with Z-FA-FMK than in cells treated with Z-DEVD-FMK, consistent with the cleavage of PARP by both caspase-3 and -7.

Replication of HAZV in the presence of Z-FA-FMK, prevented the production of the ~20 kDa cleavage product of HAZV N, whereas replication of HAZV in the presence of Z-DEVD-FMK did not. This suggested the ~20 kDa cleavage product of HAZV N was the result of cleavage of HAZV N by executioner caspases-6 or -7, rather than caspase-3, or one of their downstream proteolytic substrates. However, Z-

FA-FMK can also inhibit cathepsin B (a lysosomal enzyme involved in protein turnover and metabolism), and papain, amongst other papain-like cysteine proteases (Schotte, Declercq et al. 1999). HAZV N could be cleaved by one of these other enzymes during apoptosis; cathepsin B for example has a role in the mitochondrial intrinsic apoptotic pathway and can cleave and activate the pro-apoptotic protein BID (Sandes, Lodillinsky et al. 2007). The production of the ~20 kDa cleavage product of HAZV N in the presence of Z-DEVD-FMK suggested that caspase-3 was not responsible for cleavage of HAZV N. This indicated that cleavage of HAZV N potentially did not occur at the DQVD site; it was not concluded in this study at which site HAZV N was cleaved during apoptosis, more work needs to be carried out to resolve this point, and possibly identify where HAZV N is cleaved.

The position of the DQVD site on the apex of the arm domain is adjacent to, and also contains residues thought to be involved in, the multimerisation of HAZV N, as described in section 3.4.1. The suggested mechanism of multimerisation involved residues both on the apex and other areas of the arm domain forming a hydrophobic pocket, in which Pro351 on the base of the globular domain of an adjacent monomer is buried. The potential involvement of DQVD residues in HAZV N multimerisation implies that only monomeric HAZV N could be cleaved at the DQVD motif. The role of monomeric HAZV N during HAZV replication is currently unclear; however monomeric HAZV N could be involved in the recruitment of cellular proteins (such as HSP70/HSC70 (described in chapter 4), or other manipulations of the host cell environment. However, if only monomeric HAZV N or CCHFV N can be cleaved at this site on the apex of the arm domain, it could be an explanation for why such a small proportion of CCHFV N is cleaved by active caspase-3 during CCHFV replication.

Viral proteins from multiple families of viruses are cleaved by cellular caspases during their replication cycles; the consequences of viral protein cleavage by cellular caspases have been shown to be both beneficial and detrimental to virus replication cycles, and to both inhibit and enhance apoptosis (Richard and Tulasne 2012). The baculovirus P35 and P49 proteins for example, are both cleaved by human cellular caspases. After cleavage of P35 the cleavage product remains irreversibly associated with the caspase active site and in this way P35 directly inhibits caspases-1, -3, -6, -7, -8 and -10. Inhibition of these caspases inhibits apoptosis induction and this inhibition of apoptosis is crucial for virus replication (Zhou, Krebs et al. 1998). On the other hand the NS1 protein of the parvovirus Aleutian mink disease virus possess two caspase cleavage sites; upon permissive infection cellular caspases are activated and cleave NS1 to release 5 protein products that are essential for virus amplification

(Best, Shelton et al. 2003). Without the induction of apoptosis and activation of cellular caspases this virus cannot replicate.

In many cases the consequences for viral replication of the induction of apoptosis remains unclear. OROV induces apoptosis, which is detectable 30 hrs post infection and is dependent on expression of viral proteins and viral genome replication. However the induction of apoptosis did not appear to affect the virus replication cycle, as inhibition of apoptosis with a pan-caspase inhibitor prevented apoptosis but did not affect progeny virus production (Acrani, Gomes et al. 2010).

To begin to examine the possible involvement of cellular caspases in the HAZV virus replication cycle, the titre of secreted HAZV that had been replicating in the presence of Z-DEVD-FMK, or Z-FA-FMK was assayed and compared to untreated cells. In both cases at 24 and 48 hrs post infection, the titre of secreted HAZV was lower from Z-DEVD-FMK and Z-FA-FMK treated cells compared to untreated cells. This implied that the activation of cellular caspases was not detrimental for HAZV replication, as no increase in titre was observed at these times points. The slight decrease in titre could be related to the slight decrease in cell viability (as measured by MTT assay) of cells grown in the presence of these compounds. Intriguingly the vOTU domain possessed by the L protein of CCHFV (and by sequence homology, HAZV) is a papain-like cysteine protease; members of this family of proteins are inhibited by Z-FA-FMK (Schotte, Declercq et al. 1999). It is conceivable that the vOTU of HAZV could be partially inhibited by Z-FA-FMK, and therefore this could possibly affect the replication of these viruses in the presence of this compound.

At 72 hrs post infection the titre of secreted HAZV was lower from untreated cells then from Z-DEVD-FMK or Z-FA-FMK treated cells, possibly due to the availability of more live cells at this time point in the presence of Z-DEVD-FMK and Z-FA-FMK, to allow continued HAZV replication. It would be interesting to expand this experiment to include later time points to examine whether HAZV replication continued to increase in the presence of Z-DEVD-FMK and Z-FA-FMK. If that was the case, it could be that apoptosis does represent a host cell antiviral mechanism to control HAZV infection, and, in an analogous manner to CCHFV, HAZV has developed a way of delaying the activation of apoptosis until sufficient virus replication has occurred to allow efficient progeny spread. However, the plaque assays were only carried out once, and should therefore be repeated. In addition, this study did not address the extent to which apoptosis was inhibited by Z-DEVD-FMK and Z-FA-FMK, therefore it was not possible to draw conclusions regarding the role of apoptosis in the replication cycle of HAZV.

Cell death by apoptosis can be advantageous for viruses; during apoptosis the entire cellular contents, including progeny virions, are packaged into membrane bound apoptotic bodies that are rapidly taken up by surrounding cells, a process which severely limits the inflammatory response, and allows the virus to spread undetected (Hay and Kannourakis 2002). Virus particles within enclosed vesicles are protected from inactivation by host antibodies and proteases; thus virus dissemination can be enhanced. IAV infection of cells induces apoptosis for example; inhibition of apoptosis results in decreased IAV replication, and conversely, increased apoptosis results in an increase in IAV replication. It has been suggested that the cell and virus struggle to control apoptosis, as appropriately timed apoptosis could be crucial for efficient IAV replication (McLean, Datan et al. 2009). It could be an analogous struggle occurs during CCHFV and HAZV replication; it has already been suggested that the outcome of CCHFV infection is dependent on the timing of the activation of the IFN response, and the same could be true of the activation of apoptosis. RNA viruses have high mutation rates, therefore maintaining the N protein caspase cleavage site seems unlikely if CCHFV and HAZV do not somehow profit from them.

CHAPTER 6: DISCUSSION

6.1 THE X-RAY CRYSTAL STRUCTURE OF HAZV N REVEALS AN RNA BINDING SITE AND PROVIDES NEW INSIGHTS INTO THE FUNCTION OF HAZV N AND CCHFV N

The high resolution crystal structures presented here of Apo-HAZV N and RNA bound-HAZV N reveal new and important insights into HAZV N protein structure and function. HAZV N is primarily alpha-helical and consists of two domains; a globular domain that contains amino acid residues from both the N- and C-termini of HAZV N, and an arm domain that extends away from the globular domain. The arm domain is linked to the globular domain by a disordered loop (which is not visible in either crystal structure), and a single alpha helix, helix α13. This is the first time a nairovirus N protein has been co-crystallised with RNA. A crevice in the globular domain that channels from the base of the globular domain almost to helix α13 was shown to be an RNA binding site (rather than the positively charged patch on the surface of HAZV N adjacent to the arm domain) and had three nucleotides of ssRNA bound in the RNA bound-HAZV N structure.

The structure of HAZV N is very similar to the structure of CCHFV N; the globular domains align very closely and although the position of the arm domain relative to the globular domain is greatly altered, the fold in the arm domain is the same. The change in the position of the arm domain may be linked to changes in primary amino acid sequence between the two proteins. CCHFV N and HAZV N have very similar functions within their viral replication cycles; the great structural similarity between CCHFV N and HAZV N suggests these proteins use a similar mechanism to perform these functions, and supports the use of HAZV as a model for CCHFV infection. Due to the highly pathogenic nature of CCHFV and the high mortality rate of CCHF, manipulation of infectious CCHFV must be carried out under containment level 4 conditions. HAZV belongs in the same serogroup as CCHFV, but is not pathogenic to humans, and can therefore be manipulated under containment level 2 conditions. However HAZV infection does result in a very similar disease course in the IFNα/βR$^{-/-}$ mouse model as CCHFV. It has been proposed that HAZV can be used as a model to examine the replication cycle of these viruses and test novel antivirals, therefore bypassing the necessity of working under containment level 4 conditions; the data presented in chapter 3 supports this use of HAZV as a model for CCHFV infection.

Currently few treatment options exist for CCHF, and most are palliative; therefore there is an urgent need for novel therapeutics against CCHFV. Given the structural similarity in the globular domains of HAZV N and CCHFV N, novel small molecule inhibitors that target the RNA binding crevice of HAZV N could also target the RNA binding crevice in CCHFV N.

The functions of nairovirus N proteins within their viral replication cycles are dependent on their ability to bind viral RNA. Therefore, the identification of an RNA binding site enables the potential exploitation of this knowledge for the structure guided design of small molecules to prevent N protein function, through abolishment of N's ability to bind RNA. Alternatively the N-N interaction site could also be targeted. Targeting nucleocapsid proteins for the development of novel antivirals is becoming more common; the N protein of HIV (Ramalingam, Duclair et al. 2011), IAV (Kao, Yang et al. 2010) and RSV (Chapman, Abbott et al. 2007) have all been shown to be effective antiviral targets. More recently, a high-throughput molecular target assay was used to identify compounds that could disrupt the interaction between RVFV N and RNA *in vitro* (Ellenbecker, Lanchy et al. 2012). The compound that was found to be the most effective at inhibiting RNA binding of RVFV N in the high throughput screening assay was suramin, an FDA approved drug used to treat trypanosomiasis. Interestingly, suramin was also shown to bind the RNA binding site in the N protein of the closely related phlebovirus, severe fever with thrombocytopenia syndrome virus (SFTSV), and to inhibit replication of SFTSV in Vero cells (Jiao, Ouyang et al. 2013). These studies support the hypothesis that antiviral compounds that target the RNA binding site of HAZV N may also be effective at targeting the RNA binding site in CCHFV N, and might possibly inhibit CCHFV (and HAZV) replication.

When the N-terminal RNA binding domain of LASV N is superposed on the globular domain of HAZV N, not only do the protein folds align very closely but the three nucleotides of RNA bound in the crevice in the globular domain of HAZV N overlap almost completely with the RNA in the RNA binding site of LASV N. This not only supports the position of RNA in the crevice in the globular domain of HAZV N as being correct, but suggests that LASV N, HAZV N, and CCHFV N may be more closely related to each other than to other members of the *Bunyaviridae* family, and implies that N proteins from multiple different sNSV families share common features, and possibly a common ancestor. The structural similarity between LASV N and HAZV N, and the striking change in the position of the arm domain observed between Apo-HAZV N and RNA bound-HAZV N suggests that the change in the position of the arm domain may possibly regulate the mechanism of RNA binding of HAZV N, analogous to the 'gating' mechanism described for LASV N (the position of helix α5 in

LASV N corresponds to helix α13 in HAZV N, which links the arm domain to the globular domain).

The crystal structures of the nucleocapsid proteins of various sNSVs have been solved, and they are thought to share common features and use common mechanisms to perform their functions. One of the primary functions of sNSV N proteins is to bind vRNA or cRNA to form functional RNPs; these RNPs act as the template for further RNA synthesis rather than naked RNA. The interaction of N with vRNA and cRNA is specific, N does not bind viral mRNA, or host cell RNA. RNPs also interact with viral glycoproteins during assembly and this interaction is thought to drive the budding of progeny virions into the Golgi cisternae. In order to carry out these functions, it has been proposed that sNSV N proteins possess common features such as; a mechanism to keep N monomeric, in order to prevent inappropriate multimerisation and RNA binding, a parallel orientation of subunits in the nucleocapsid which results in nucleocapsids with linear symmetry, and a cavity for the sequestration of RNA, that protects the RNA from degradation by RNA nucleases to a certain degree (Green, Cox et al. 2014; Reguera, Cusack et al. 2014).

The crystal structures of Apo-HAZV N and RNA bound-HAZV N presented here reveal potential mechanisms by which all of these criteria are met. The change in the position of the arm domain that could potentially represent a ‘gating’ mechanism of RNA binding analogous to LASV N, also reveals a mechanism for keeping monomeric HAZV N from binding RNA, as has been proposed for LASV N (Hastie, Liu et al. 2011). In lieu of a co-chaperone P protein, helix α6 and part of helix α5 occlude the RNA binding crevice of LASV N in the absence of RNA; this is proposed to help keep LASV N monomeric, and prevent LASV N from binding RNA inappropriately (Hastie, Liu et al. 2011). A similar mechanism has been proposed for RVFV N, where an N terminal loop involved in N multimerisation folds back on itself, and an alpha helix occludes the RNA binding site in the absence of RNA (Raymond, Piper et al. 2010; Raymond, Piper et al. 2012). It is not known what the subsequent trigger is for RNA binding, however it is hypothesised the L protein might induce conformational changes in N to enable RNA to bind. Given the limited coding capacity of bunyaviruses, the ability of N and L to regulate RNA binding of N is more economical than encoding an additional protein (such as the P protein of nsNSVs) to perform this function.

The crystal structure of Apo-HAZV N reveals a potential mechanism of HAZV N multimerisation involving the hydrophobic interaction between residues from the base of the globular domain of one monomer and the arm domain of the adjacent monomer, which results in the presence of linear chains of HAZV N throughout both

Apo-HAZV N and RNA bound-HAZV N crystals. This could reflect multimerisation events that occur when viral RNPs form, in which case viral RNPs would potentially possess a linear nucleocapsid symmetry, with a parallel orientation of N subunits. The involvement of residues on the apex of the arm domain of HAZV N and CCHFV N in N multimerisation (including some residues of the DQVD or DEVD motif) could possibly provide an explanation for the conservation of the DEVD motif in CCHFV N, despite its cleavage by capase-3. Finally, RNA bound-HAZV N reveals an RNA binding site deep within a crevice in the globular domain where the RNA is partially protected from degradation by RNA nucleases.

6.2 CCHFV N AND HAZV N INTERACT WITH HSP70/HSC70

The interaction between CCHFV N and HAZV N and HSP70/HSC70 presented here potentially reveals new insights into the replication cycles of these viruses. The successful recruitment of host cell factors is essential for virus proliferation; viruses are dependent on cellular proteins for every aspect of their replication cycles from cell entry, to genome replication, protein synthesis, virion assembly and secretion of progeny virions. It is often the role of viral proteins to recruit specific cellular factors to aid virus replication; indeed the function of a viral protein within the cellular environment can be defined by the multiple protein-protein interactions it makes. Therefore various studies have focused on examining viral protein ‘interactomes’, often through the use of mass spectrometry based proteomics. In order to identify cellular proteins that interacted with CCHFV N, we used a GFP-Trap to IP an EGFP-CCHFV N fusion protein from transfected cells, and SILAC based MS to identify cellular proteins that immunoprecipitated with EGFP-CCHFV N. SILAC enabled the differentiation between cellular proteins that interacted non-specifically with EGFP or components of the GFP-Trap, and those that interacted specifically with EGFP-CCHFV N. Although multiple cellular proteins were identified that potentially interacted with EGFP-CCHFV N, this study could not provide a complete ‘interactome’ of CCHFV N as the use of the EGFP tag potentially obscured cellular binding partners, as does the use of any antibody to IP protein complexes. Isolation of protein complexes by size exclusion chromatography based methods, and identification of their protein compositions by MS has been suggested as a method for elucidating comprehensive ‘interactomes’.

MS analysis revealed that HSP70 and HSC70 both immunoprecipitated with EGFP-CCHFV N, and were chosen for further examination. HSP70/HSC70 also immunoprecipitated with untagged CCHFV N both in the context of actively replicating CCHFV, and from pCAGGs-CCHFV N expressing cells. Furthermore, HSP70/HSC70 immunoprecipitated with HAZV N from HAZV infected cells. Members of the HSP70 family of proteins are molecular chaperones, located in the cytoplasm and most cellular organelles that play a central role in multiple cellular processes including protein translation, folding, degradation, intracellular trafficking, and the assembly and disassembly of oligomeric structures, amongst other roles (Mayer 2005). Members of the HSP70 family are frequently recruited by viruses and the precise role within virus replication cycles of HSP70 family members varies. Some viruses are thought to recruit HSP70 family members to assist with viral protein folding, as this can become a rate limiting step during replication, some are thought to be recruited to assist transcription by viral polymerases, and others are recruited to assist virus assembly. Several studies have reported an association between HSP70 and viral polymerase complexes of HCV, RSV (HSP70 is thought to be recruited to lipid raft domains, where viral replication complexes assemble), RV and IAV (Brown, Rixon et al. 2005; Lahaye, Vidy et al. 2012; Khachatoorian, Ganapathy et al. 2014; Manzoor, Kuroda et al. 2014). HSP70 has been shown to stimulate the viral polymerase *in vitro* activity of several paramyxoviruses, including canine distemper virus and RSV, however it is not clear whether HSP70 assists the folding of N in an active form, or whether HSP70 directly acts as a cofactor of the enzymatic process (Oglesbee, Liu et al. 1996; Brown, Rixon et al. 2005). HSP70 protein levels may or may not be upregulated in cells infected with viruses that recruit HSP70 as part of their replication strategy.

HSP70 and/or HSC70 interacts with NS5A of HCV, NS1 and M1 of IAV, N and P protein of RSV, and the N protein of RV, in addition to many other viral proteins (Halder, Bagchi et al. 2011; Lahaye, Vidy et al. 2012; Oliveira, Simabuco et al. 2013; Khachatoorian, Ganapathy et al. 2014; Manzoor, Kuroda et al. 2014). Similar to CCHFV N, NS5A of HCV interacts with both HSP70 and HSC70, and interestingly, HSP70 and HSC70 were shown to have different roles within the HCV replication cycle. HSP70 was involved in genome transcription and viral protein translation, whereas HSC70 was involved in virion assembly; HSC70 is also found within or associated with, purified HCV particles (Parent, Qu et al. 2009; Khachatoorian, Ganapathy et al. 2014). Generally the interaction of viral proteins with HSP70 and/or HSC70 and the recruitment of these chaperones, is beneficial to virus replication cycles; siRNA mediated knock down of HSP70 and/or HSC70 inhibits the replication of HCV, RV, HSV, PRRSV, and IAV (Stegen, Yakova et al. 2013; Gao, Xiao et al. 2014). In good agreement with siRNA studies, inhibition of HSP70 expression with the bioflavonoid quercetin inhibits the replication of RV, HCV, and PRRSV (Gonzalez, Fontanes et al. 2009). In some cases however, the interaction of HSP70 with viral proteins has been found to inhibit virus replication; infection of cells with Hantaan virus induces the expression of HSP70 4 hrs after infection. HSP70 interacts with Hantaan virus N protein, and overexpression of HSP70 reduces Hantaan virus replication and N protein expression (Yu, Ye et al. 2009). HSP70 has also been implicated in mediating the antiviral effects of prostaglandins (Santoro 1994).

Depletion of HSP70 from HAZV infected cells modestly reduced HAZV N expression levels compared to scrambled siRNA transfected cells. Modestly reduced titres of HAZV were also secreted from HSP70 depleted cells, compared to scrambled siRNA transfected cells. Complete and sustained depletion of HSP70 over long time periods could not be achieved however; it could be that a greater effect on HAZV replication could be observed with longer term HSP70 depletion, or depletion of HSC70 could have a greater effect on HAZV replication. Preliminary experiments using the HSP70 inhibitors VER and PIF also resulted in decreased HAZV replication in SW13 cells compared to DMSO controls, which suggested a positive role for HSP70/HSC70 within the HAZV replication cycle. However the precise role of HSP70 and/or HSC70 within the replication cycles of HAZV and CCHFV remains to be determined.

The secreted virions of enveloped viruses contain multiple host cell proteins in addition to viral proteins; virions contain all the molecular machinery needed to enter and exit host cells, and, given the simplicity of viral proteomes, and the complexity of these processes, it is thought virions contain cellular proteins to help perform these tasks (Cantin, Methot et al. 2005; Nuss, Kehn-Hall et al. 2014). Cellular proteins associated with virions can therefore act as essential or important host cell factors for viral infection, and represent novel targets for antiviral therapeutic development. HSP70 and/or HSC70 have been found associated with the extracellular virus particles of RSV, RV, HCV, HSV, RVFV and HIV-1, amongst others (Chertova, Chertov et al. 2006; Radhakrishnan, Yeo et al. 2010; Stegen, Yakova et al. 2013; Nuss, Kehn-Hall et al. 2014).

In this study, MS analysis of the fraction containing infectious HAZV identified HSC70 in this fraction, as well as all 4 viral structural proteins. The presence of HAZV N and HSC70 in the fraction containing infectious HAZV was verified by western blot analysis, and furthermore, HSC70/HSP70 immunoprecipitated with HAZV N from the fraction containing infectious HAZV. This indicated HSC70 might associate with HAZV N during multiple stages of the virus replication cycle, and may be packaged into, or associated with, secreted HAZV particles. In addition to HSC70, MS analysis identified 260 cellular proteins in the fraction containing infectious HAZV. Interestingly, a recent study carried out by (Nuss, Kehn-Hall et al. 2014) identified ~270 cellular proteins associated with secreted RVFV particles, in the first published proteomic analysis of purified bunyavirus particles. RVFV particle preparations were enriched in host cell chaperones, and many of the proteins identified in this study were common to both HAZV and RVFV particle preparations, including the cellular chaperones HSC70, HSPA5, HSP90, CCT6a (T-complex protein 1 subunit zeta) and CCT2 (T-complex protein 1 subunit beta), and the cytoskeletal proteins actin and tubulin, amongst others. Interestingly CCT6a and CCT2 were also both identified in the GFP-Trap MS analysis as being potential interacting partners of EGFP-CCHFV N (SILAC ratios of 4.2, and 5.54, 6.13 respectively). siRNA depletion of HSPA5, HSP90, CCT6a and CCT2 resulted in decreased RVFV replication compared to scrambled siRNA, whereas depletion of HSC70 resulted in increased RVFV replication compared to scrambled siRNA.

The heat shock protein 90 (HSP90) inhibitor 17-AAG, and the HSP70 inhibitors KNK437 (blocks induction of chaperones from HSP70, 72 and 105 families) and BAPTA-AM (a calcium chelator that decreases expression of BiP mRNA) greatly inhibited RVFV replication without any cytotoxic effect (Nuss, Kehn-Hall et al. 2014). This suggests HSP90, and some members of the HSP70 family play a positive role in RVFV replication, and could possibly represent promising therapeutic targets. It could be that HSPA5, HSP90, CCT6a, and CCT2 also play a role in HAZV (and CCHFV) replication, and these cellular chaperones may be recruited as part of a common replication strategy by these viruses. It must be noted however, that in both these studies the ratio of infectious virus particles:defective, non-infectious virus particles analysed by MS is unknown.

The identification of specific cellular proteins and pathways required by individual viruses, provides novel targets for host-directed antiviral development. Additionally, the identification of common cellular proteins recruited by multiple different viruses could potentially lead to the identification of host-directed antiviral therapeutics with broad-spectrum antiviral activity. For example HSP90, an ubiquitous cytoplasmic molecular chaperone, is involved in the replication cycle of multiple different RNA viruses; HSP90 is thought to stabilise viral RNA-dependent RNA polymerases. Several small molecule HSP90 inhibitors including geldanamycin, and its derivative 17-AAG, inhibit the replication cycle of vaccinia virus, IAV, HSV, VSV, several paramyxoviruses, LACV, RVFV and HCV (Nuss, Kehn-Hall et al. 2014). 17-AAG is currently in Phase 2 clinical trials as an anti-cancer therapeutic and it is hoped 17-AAG, as well as other HSP90 inhibitors could also be developed as broad-spectrum antivirals. Given the abundant recruitment of members of the HSP70 family by multiple different viruses, and the development of HSP70 inhibitors as anti-cancer therapeutics, it is hoped HSP70 inhibitors could also potentially be developed as antiviral agents. Interestingly a synergistic effect against cancerous cells can be observed when compounds targeting HSP90 and HSP70 are administered concurrently, a similar effect could also be observed when targeting virus replication. However the benefits of designing host-directed antivirals, such as the slower mutation rates of mammalian proteins reducing the rate at which virus escape mutants develop, is tempered by the potentially cytotoxic effects of targeting host proteins.

6.3 HAZV INDUCES APOPTOSIS IN SW13 CELLS

In this study HAZV infection of SW13 cells was shown to result in the induction of apoptosis 48 and 72 hrs after infection, which occurred concomitantly with the cleavage and activation of caspase-9, -3, -7, and cleavage of PARP. Viral control of apoptosis is often a critical factor determining the outcome of virus infection. Viral replication is dependent on the exploitation of existing cellular pathways and proteins; upon detection of viruses mammalian cells have developed strategies to limit viral spread, often through the elimination of these crucial proteins and pathways by controlled cell death, or apoptosis. Apoptosis is induced during the replication cycle of CCHFV, in addition to several other bunyaviruses. The induction of apoptosis for many of these viruses, including OROV and CCHFV is dependent on live virus replication and the synthesis of viral RNA or proteins, as UV inactivated virus fails to induce apoptosis (Acrani, Gomes et al. 2010; Rodrigues, Paranhos-Baccala et al. 2012). It is not known whether live virus replication is necessary for the induction of apoptosis observed during HAZV infection. This is in contrast to other viruses such as HIV-1, where the interaction of the virus glycoprotein gp120 with cell surface receptors is sufficient to induce apoptosis. It is thought the ability of gp120 to induce apoptosis in this manner could account for some of the bystander effects and pathogenesis observed during HIV-1 replication, such as depletion of CD4 positive cells (Anand and Ganju 2006).

It has been suggested that the ability of CCHFV to replicate to high titres and induce apoptosis in HUH-7 cells, may be related to CCHF disease pathogenesis, where the liver is one of the main target organs during infection, high viral loads in the blood are an effective predictor of disease outcome, and death of macrophages and other cells in a manner that is consistent with apoptosis can be observed in the liver and spleen. Apoptosis induction in CCHFV infected cells is thought to be linked to ER stress; this results in increased expression of the pro-apoptotic proteins PUMA, Noxa, and BAX, and induction of apoptosis is thought to occur through the mitochondrial pathway (Rodrigues, Paranhos-Baccala et al. 2012). Apoptosis is induced 48 and 72 hrs after infection, and results in the activation of caspase-3, which subsequently cleaves CCHFV N at a conserved DEVD motif on the apex of the arm domain (Karlberg, Tan et al. 2011).

HAZV infection of the IFNRα/β$^{-/-}$ mouse model results in a similar disease progression to CCHFV, including high viral loads in the blood, elevated serum levels of pro-inflammatory cytokines, and the liver was also one of the main targets for HAZV replication (Dowall, Findlay-Wilson et al. 2012). Although apoptosis has not been definitively demonstrated in the liver and spleen of HAZV infected IFNRα/β$^{-/-}$ mice, lymphocyte depletion in the spleen is observed that could very well be due to apoptosis. The ability of HAZV to induce apoptosis in cultured mammalian cells suggests that apoptosis might also possibly occur during HAZV infection of IFNRα/β$^{-/-}$ mice, and this might be a contributing factor towards disease pathogenesis. This study did not address the mechanism through which apoptosis occurred in HAZV infected cells; it could be that apoptosis induction was linked to ER stress and up regulation of pro-apoptotic proteins, such as BAX. The activation of caspase-9 suggested that apoptosis occurred via the mitochondrial pathway, however there is crosstalk between the extrinsic and intrinsic apoptotic pathways (for example caspase-8 cleaves Bid, converting it to tBid, which translocates to the mitochondria and aids the release of cytochrome C from the intermembrane space, resulting in the formation of an ‘apoptosome’ and activation of caspase-9) therefore either one or both could be activated during HAZV infection. There is evidence that CCHFV is possibly able to modulate both the intrinsic and extrinsic apoptotic pathways (Rodrigues, Paranhos-Baccala et al. 2012).

HAZV and CCHFV both induce apoptosis at relatively late time points after infection; during CCHFV replication apoptosis was detected 48 hrs after infection. CCHFV delays activation of the IFN response possibly through delaying translocation of IRF-3 to the nucleus; it is possible that both CCHFV and HAZV also possess mechanisms to delay the activation of apoptosis until sufficient progeny virions have been released to ensure productive infection. Interestingly the Hantaan virus N protein has been shown to modulate apoptosis; tumor necrosis factor (TNF) stimulated cells expressing Hantaan virus N sequester nuclear factor kappa B (NFκB) in the cytoplasm, and have reduced caspase-7 activity; similar mechanisms could occur during CCHFV and HAZV replication (Ontiveros, Li et al. 2010).

This study showed that HAZV N was cleaved during a 72 hr course of infection, and the abundance of the ~20 kDa cleavage product increased at 48 and 72 hrs after infection, concomitant with the induction of apoptosis. In contrast to cleavage of CCHFV N, which could be blocked by treatment of cells with the caspase-3 inhibitor Z-DEVD-FMK (Karlberg, Tan et al. 2011), cleavage of HAZV N was not blocked in the presence of this compound, but was instead inhibited when cells were treated with the inhibitor of executioner caspases, Z-FA-FMK. This suggested that HAZV N was cleaved either by one of the other executioner caspases, one of their downstream targets, or another enzyme (such as cathepsin B), which is also inhibited by Z-FA-FMK (Schotte, Declercq et al. 1999). The specific site at which HAZV N was cleaved was not identified, however it is possible cleavage occurred at the DQVD site, as the ~20 kDa HAZV N cleavage product is similar in size to that which could be produced by cleavage at the DQVD site. The cleavage of viral proteins during apoptosis has been widely reported in the literature; and the outcome for virus infection of this cleavage varies substantially (Richard and Tulasne 2012).

Preliminary experiments were carried out to assay the effect on HAZV replication of treatment of SW13 cells with Z-DEVD-FMK and Z-FA-FMK. However it is not known whether apoptosis was fully inhibited in this study, therefore it is hard to draw conclusions regarding the effect of apoptosis on HAZV replication. In contrast to CCHFV however, HAZV titres did not increase in the presence of Z-DEVD-FMK; this could either be due to the slight decrease in the viability of cells grown in the presence of Z-DEVD-FMK, or activation of caspase-3 was not detrimental to HAZV replication. It could be that, analogous to OROV infection, where inhibition of apoptosis does not affect the titre of secreted virions, inhibition of apoptosis might not affect HAZV replication (Acrani, Gomes et al. 2010; Karlberg, Tan et al. 2011).

Apoptosis is often considered to be an antiviral response to infection, which viruses have to avoid or inhibit in order to replicate successfully. However, viruses can also take advantage of apoptosis; ultimately apoptosis results in the disintegration of the cell into small membrane bound entities, or apoptotic bodies, which pinch off from the dying cell, and are then phagocytised by neighbouring cells without inducing an inflammatory response. The phagocytosis of virus containing apoptotic bodies could enable the dissemination of progeny virions without their release into the extracellular fluid; an action that would result in the induction of a host response (Hay and Kannourakis 2002). During Sindbis virus replication the induction of apoptosis is timed to coincide with completion of the replication cycle; and although the induction of apoptosis during DNA replication of an adenovirus vector was detrimental to virus infection, apoptosis induced after virion assembly aided virus dissemination, and virions could be visualised within, or associated with, apoptotic bodies by EM (Lewis, Wesselingh et al. 1996; Everett and McFadden 1999; Mi, Li et al. 2001; Hay and Kannourakis 2002). CCHFV and HAZV infection of SW13 cells both result in productive replication, therefore these viruses must have evolved a way of circumventing the detrimental effects of apoptosis; such as delaying activation of apoptosis or replicating undetected in cells for long enough to sustain a productive infection. It is thought that the fast mutation rate of RNA viruses ensures that viruses with replication strategies and protein motifs that are not beneficial to replication are quickly out competed. The conservation of the DEVD motif among CCHFV strains, and the late induction of apoptosis during both CCHFV and HAZV replication could somehow be beneficial to the virus. However, it should be borne in mind that humans are not the natural host for CCHFV or HAZV, and the outcome of infection for their tick vectors and animals hosts is very different to human infection, in the case of CCHFV.

REFERENCES

Acrani, G. O., R. Gomes, et al. (2010). "Apoptosis induced by Oropouche virus infection in HeLa cells is dependent on virus protein expression." Virus Res **149**(1): 56-63.

Aguilar, P. V., A. D. Barrett, et al. (2011). "Iquitos virus: a novel reassortant Orthobunyavirus associated with human illness in Peru." PLoS Negl Trop Dis **5**(9): e1315.

Akutsu, M., Y. Ye, et al. (2011). "Molecular basis for ubiquitin and ISG15 cross-reactivity in viral ovarian tumor domains." Proc Natl Acad Sci U S A **108**(6): 2228-2233.

Altamura, L. A., A. Bertolotti-Ciarlet, et al. (2007). "Identification of a novel C-terminal cleavage of Crimean-Congo hemorrhagic fever virus PreGN that leads to generation of an NSM protein." J Virol **81**(12): 6632-6642.

Anand, A. R. and R. K. Ganju (2006). "HIV-1 gp120-mediated apoptosis of T cells is regulated by the membrane tyrosine phosphatase CD45." J Biol Chem **281**(18): 12289-12299.

Andersson, C., S. Henriksson, et al. (2012). "In situ rolling circle amplification detection of Crimean Congo hemorrhagic fever virus (CCHFV) complementary and viral RNA." Virology **426**(2): 87-92.

Andersson, I., L. Bladh, et al. (2004). "Human MxA protein inhibits the replication of Crimean-Congo hemorrhagic fever virus." J Virol **78**(8): 4323-4329.

Andersson, I., H. Karlberg, et al. (2008). "Crimean-Congo hemorrhagic fever virus delays activation of the innate immune response." J Med Virol **80**(8): 1397-1404.

Andersson, I., A. Lundkvist, et al. (2006). "Type I interferon inhibits Crimean-Congo hemorrhagic fever virus in human target cells." J Med Virol **78**(2): 216-222.

Andersson, I., M. Simon, et al. (2004). "Role of actin filaments in targeting of Crimean Congo hemorrhagic fever virus nucleocapsid protein to perinuclear regions of mammalian cells." J Med Virol **72**(1): 83-93.

Appannanavar, S. B. and B. Mishra (2011). "An update on crimean congo hemorrhagic Fever." J Glob Infect Dis **3**(3): 285-292.

Ariza, A., S. J. Tanner, et al. (2013). "Nucleocapsid protein structures from orthobunyaviruses reveal insight into ribonucleoprotein architecture and RNA polymerization." Nucleic Acids Res **41**(11): 5912-5926.

Arranz, R., R. Coloma, et al. (2012). "The structure of native influenza virion ribonucleoproteins." Science **338**(6114): 1634-1637.

Atkinson, B., J. Chamberlain, et al. (2013). "Identification and analysis of Crimean-Congo hemorrhagic fever virus from human sera in Tajikistan." Int J Infect Dis **17**(11): e1031-1037.

Atkinson, B., J. Chamberlain, et al. (2012). "Development of a real-time RT-PCR assay for the detection of Crimean-Congo hemorrhagic fever virus." Vector Borne Zoonotic Dis **12**(9): 786-793.

Atkinson, B., J. Latham, et al. (2012). "Sequencing and phylogenetic characterisation of a fatal Crimean - Congo haemorrhagic fever case imported into the United Kingdom, October 2012." Euro Surveill **17**(48).

Barr, J. N., R. M. Elliott, et al. (2003). "Segment-specific terminal sequences of Bunyamwera bunyavirus regulate genome replication." Virology **311**(2): 326-338.

Barr, J. N., J. W. Rodgers, et al. (2006). "Identification of the Bunyamwera bunyavirus transcription termination signal." J Gen Virol **87**(Pt 1): 189-198.

Battye, T. G., L. Kontogiannis, et al. (2011). "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 271-281.

Begum, F., C. L. Wisseman, Jr., et al. (1970). "Tick-borne viruses of West Pakistan. II. Hazara virus, a new agent isolated from Ixodes redikorzevi ticks from the Kaghan Valley, W. Pakistan." Am J Epidemiol **92**(3): 192-194.

Begum, F., C. L. Wisseman, Jr., et al. (1970). "Tick-borne viruses of West Pakistan. I. Isolation and general characteristics." Am J Epidemiol **92**(3): 180-191.

Bente, D. A., J. B. Alimonti, et al. (2010). "Pathogenesis and immune response of Crimean-Congo hemorrhagic fever virus in a STAT-1 knockout mouse model." J Virol **84**(21): 11089-11100.

Bente, D. A., N. L. Forrester, et al. (2013). "Crimean-Congo hemorrhagic fever: history, epidemiology, pathogenesis, clinical syndrome and genetic diversity." Antiviral Res **100**(1): 159-189.

Bereczky, S., G. Lindegren, et al. (2010). "Crimean-Congo hemorrhagic fever virus infection is lethal for adult type I interferon receptor-knockout mice." J Gen Virol **91**(Pt 6): 1473-1477.

Bergeron, E., C. G. Albarino, et al. (2010). "Crimean-Congo hemorrhagic fever virus-encoded ovarian tumor protease activity is dispensable for virus RNA polymerase function." J Virol **84**(1): 216-226.

Bergeron, E., M. J. Vincent, et al. (2007). "Crimean-Congo hemorrhagic fever virus glycoprotein processing by the endoprotease SKI-1/S1P is critical for virus infectivity." J Virol **81**(23): 13271-13276.

Bertolotti-Ciarlet, A., J. Smith, et al. (2005). "Cellular localization and antigenic characterization of crimean-congo hemorrhagic fever virus glycoproteins." J Virol **79**(10): 6152-6161.

Best, S. M., J. F. Shelton, et al. (2003). "Caspase cleavage of the nonstructural protein NS1 mediates replication of Aleutian mink disease parvovirus." J Virol **77**(9): 5305-5312.

Blakqori, G., I. van Knippenberg, et al. (2009). "Bunyamwera orthobunyavirus S-segment untranslated regions mediate poly(A) tail-independent translation." J Virol **83**(8): 3637-3646.

Boshra, H., G. Lorenzo, et al. (2011). "Rift valley fever: recent insights into pathogenesis and prevention." J Virol **85**(13): 6098-6105.

Boulon, S., Y. Ahmad, et al. (2010). "Establishment of a protein frequency library and its application in the reliable identification of specific protein interaction partners." Molecular & cellular proteomics : MCP **9**(5): 861-879.

Bowen, M. D., S. G. Trappier, et al. (2001). "A reassortant bunyavirus isolated from acute hemorrhagic fever cases in Kenya and Somalia." Virology **291**(2): 185-190.

Bowick, G. C., A. M. Airo, et al. (2012). "Expression of interferon-induced antiviral genes is delayed in a STAT1 knockout mouse model of Crimean-Congo hemorrhagic fever." Virol J **9**: 122.

Bray, M. and S. Mahanty (2003). "Ebola hemorrhagic fever and septic shock." J Infect Dis **188**(11): 1613-1617.

Briese, T., B. Bird, et al. (2006). "Batai and Ngari viruses: M segment reassortment and association with severe febrile disease outbreaks in East Africa." J Virol **80**(11): 5627-5630.

Brown, G., H. W. Rixon, et al. (2005). "Evidence for an association between heat shock protein 70 and the respiratory syncytial virus polymerase complex within lipid-raft membranes during virus infection." Virology **338**(1): 69-80.

Budina-Kolomets, A., G. M. Balaburski, et al. (2014). "Comparison of the activity of three different HSP70 inhibitors on apoptosis, cell cycle arrest, autophagy inhibition, and HSP90 inhibition." Cancer Biol Ther **15**(2): 194-199.

Burney, M. I., A. Ghafoor, et al. (1980). "Nosocomial outbreak of viral hemorrhagic fever caused by Crimean Hemorrhagic fever-Congo virus in Pakistan, January 1976." Am J Trop Med Hyg **29**(5): 941-947.

Burt, F. J., P. A. Leman, et al. (1994). "Serodiagnosis of Crimean-Congo haemorrhagic fever." Epidemiol Infect **113**(3): 551-562.

Burt, F. J., J. T. Paweska, et al. (2009). "Genetic relationship in southern African Crimean-Congo haemorrhagic fever virus isolates: evidence for occurrence of reassortment." Epidemiol Infect **137**(9): 1302-1308.

Burt, F. J., D. C. Spencer, et al. (1996). "Investigation of tick-borne viruses as pathogens of humans in South Africa and evidence of Dugbe virus infection in a patient with prolonged thrombocytopenia." Epidemiol Infect **116**(3): 353-361.

Burt, F. J., R. Swanepoel, et al. (1997). "Immunohistochemical and in situ localization of Crimean-Congo hemorrhagic fever (CCHF) virus in human tissues and implications for CCHF pathogenesis." Arch Pathol Lab Med **121**(8): 839-846.

Buttigieg, K. R., S. D. Dowall, et al. (2014). "A Novel Vaccine against Crimean-Congo Haemorrhagic Fever Protects 100% of Animals against Lethal Challenge in a Mouse Model." PLoS One **9**(3): e91516.

Cantin, R., S. Methot, et al. (2005). "Plunder and stowaways: incorporation of cellular proteins by enveloped viruses." J Virol **79**(11): 6577-6587.

Carroll, S. A., B. H. Bird, et al. (2010). "Ancient common ancestry of Crimean-Congo hemorrhagic fever virus." Mol Phylogenet Evol **55**(3): 1103-1110.

Carter, S. D., J. N. Barr, et al. (2012). "Expression, purification and crystallization of the Crimean-Congo haemorrhagic fever virus nucleocapsid protein." Acta Crystallogr Sect F Struct Biol Cryst Commun **68**(Pt 5): 569-573.

Carter, S. D., R. Surtees, et al. (2012). "Structure, function, and evolution of the Crimean-Congo hemorrhagic fever virus nucleocapsid protein." J Virol **86**(20): 10914-10923.

Casals, J. (1969). "Antigenic similarity between the virus causing Crimean hemorrhagic fever and Congo virus." Proc Soc Exp Biol Med **131**(1): 233-236.

Causey, O. R., G. E. Kemp, et al. (1970). "Congo virus from domestic livestock, African hedgehog, and arthropods in Nigeria." Am J Trop Med Hyg **19**(5): 846-850.

Ceianu, C. S., R. I. Panculescu-Gatej, et al. (2012). "First serologic evidence for the circulation of Crimean-Congo hemorrhagic fever virus in Romania." Vector Borne Zoonotic Dis **12**(9): 718-721.

Cevik, M. A., A. Erbay, et al. (2007). "Viral load as a predictor of outcome in Crimean-Congo hemorrhagic fever." Clin Infect Dis **45**(7): e96-100.

Cevik, M. A., A. Erbay, et al. (2008). "Clinical and laboratory features of Crimean-Congo hemorrhagic fever: predictors of fatality." Int J Infect Dis **12**(4): 374-379.

Chamberlain, J., N. Cook, et al. (2005). "Co-evolutionary patterns of variation in small and large RNA segments of Crimean-Congo hemorrhagic fever virus." J Gen Virol **86**(Pt 12): 3337-3341.

Chambers, M. C., B. Maclean, et al. (2012). "A cross-platform toolkit for mass spectrometry and proteomics." Nat Biotechnol **30**(10): 918-920.

Chapman, J., E. Abbott, et al. (2007). "RSV604, a novel inhibitor of respiratory syncytial virus replication." Antimicrob Agents Chemother **51**(9): 3346-3353.

Chen, L. H. and M. E. Wilson (2004). "Transmission of dengue virus without a mosquito vector: nosocomial mucocutaneous transmission and other routes of transmission." Clin Infect Dis **39**(6): e56-60.

Chen, V. B., W. B. Arendall, 3rd, et al. (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallogr D Biol Crystallogr **66**(Pt 1): 12-21.

Cheng, E., A. Haque, et al. (2011). "Characterization of the Interaction between hantavirus nucleocapsid protein (N) and ribosomal protein S19 (RPS19)." J Biol Chem **286**(13): 11814-11824.

Chertova, E., O. Chertov, et al. (2006). "Proteomic and biochemical analysis of purified human immunodeficiency virus type 1 produced from infected monocyte-derived macrophages." J Virol **80**(18): 9039-9052.

Chumakov, M. P. (1974). "On 30 years of investigation of Crimean hemorrhagic fever." Sb Tr Inst Polio Virusn Encefalitov Akad Med Nauk SSSR(22): 5-18.

Chumakov, M. P., A. M. Butenko, et al. (1968). "[New data on the viral agent of Crimean hemorrhagic fever]." Vopr Virusol **13**(3): 377.

Clerx, J. P., J. Casals, et al. (1981). "Structural characteristics of nairoviruses (genus Nairovirus, Bunyaviridae)." J Gen Virol **55**(Pt 1): 165-178.

Colon-Ramos, D. A., P. M. Irusta, et al. (2003). "Inhibition of translation and induction of apoptosis by Bunyaviral nonstructural proteins bearing sequence similarity to reaper." Mol Biol Cell **14**(10): 4162-4172.

Cowtan, K. (2006). "The Buccaneer software for automated model building. 1. Tracing protein chains." Acta Crystallogr D Biol Crystallogr **62**(Pt 9): 1002-1011.

Cox, J. and M. Mann (2008). "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification." Nat Biotechnol **26**(12): 1367-1372.

Darwish, M. A., H. Hoogstraal, et al. (1983). "A sero-epidemiological survey for Bunyaviridae and certain other arboviruses in Pakistan." Trans R Soc Trop Med Hyg **77**(4): 446-450.

Davies, C. R., L. D. Jones, et al. (1987). "In vivo reassortment of Thogoto virus (a tick-borne influenza-like virus) following oral infection of Rhipicephalus appendiculatus ticks." J Gen Virol **68 (Pt 9)**: 2331-2338.

Dinarello, C. A. (1996). "Cytokines as mediators in the pathogenesis of septic shock." Curr Top Microbiol Immunol **216**: 133-165.

Dohm, D. J., T. M. Logan, et al. (1996). "Transmission of Crimean-Congo hemorrhagic fever virus by Hyalomma impeltatum (Acari:Ixodidae) after experimental infection." J Med Entomol **33**(5): 848-851.

Dong, H., P. Li, et al. (2013). "Crystal structure of Schmallenberg orthobunyavirus nucleoprotein-RNA complex reveals a novel RNA sequestration mechanism." RNA **19**(8): 1129-1136.

Dowall, S. D., S. Findlay-Wilson, et al. (2012). "Hazara virus infection is lethal for adult type I interferon receptor-knockout mice and may act as a surrogate for infection with the human-pathogenic Crimean-Congo hemorrhagic fever virus." J Gen Virol **93**(Pt 3): 560-564.

Duh, D., A. Saksida, et al. (2007). "Viral load as predictor of Crimean-Congo hemorrhagic fever outcome." Emerg Infect Dis **13**(11): 1769-1772.

Duh, D., A. Saksida, et al. (2006). "Novel one-step real-time RT-PCR assay for rapid and specific diagnosis of Crimean-Congo hemorrhagic fever encountered in the Balkans." J Virol Methods **133**(2): 175-179.

Eifan, S., E. Schnettler, et al. (2013). "Non-structural proteins of arthropod-borne bunyaviruses: roles and functions." Viruses **5**(10): 2447-2468.

Eifan, S. A. and R. M. Elliott (2009). "Mutational analysis of the Bunyamwera orthobunyavirus nucleocapsid protein gene." J Virol **83**(21): 11307-11317.

Ellenbecker, M., J. M. Lanchy, et al. (2012). "Identification of Rift Valley fever virus nucleocapsid protein-RNA binding inhibitors using a high-throughput screening assay." J Biomol Screen **17**(8): 1062-1070.

Elliott, R. M. (2014). "Orthobunyaviruses: recent genetic and structural insights." Nat Rev Microbiol **12**(10): 673-685.

Ellis, D. S., T. Southee, et al. (1981). "Congo/Crimean haemorrhagic fever virus from Iraq 1979: I. Morphology in BHK21 cells." Arch Virol **70**(3): 189-198.

Elmore, S. (2007). "Apoptosis: a review of programmed cell death." Toxicol Pathol **35**(4): 495-516.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-2132.

Ergonul, O. (2006). "Crimean-Congo haemorrhagic fever." Lancet Infect Dis **6**(4): 203-214.

Ergonul, O. (2008). "Treatment of Crimean-Congo hemorrhagic fever." Antiviral Res **78**(1): 125-131.

Ergonul, O., A. Celikbas, et al. (2004). "Characteristics of patients with Crimean-Congo hemorrhagic fever in a recent outbreak in Turkey and impact of oral ribavirin therapy." Clin Infect Dis **39**(2): 284-287.

Ergonul, O., S. Tuncbilek, et al. (2006). "Evaluation of serum levels of interleukin (IL)-6, IL-10, and tumor necrosis factor-alpha in patients with Crimean-Congo hemorrhagic fever." J Infect Dis **193**(7): 941-944.

Erickson, B. R., V. Deyde, et al. (2007). "N-linked glycosylation of Gn (but not Gc) is important for Crimean Congo hemorrhagic fever virus glycoprotein localization and transport." Virology **361**(2): 348-355.

Estrada-Pena, A., A. M. Palomar, et al. (2012). "Crimean-Congo hemorrhagic fever virus in ticks, Southwestern Europe, 2010." Emerg Infect Dis **18**(1): 179-180.

Estrada, D. F. and R. N. De Guzman (2011). "Structural characterization of the Crimean-Congo hemorrhagic fever virus Gn tail provides insight into virus assembly." J Biol Chem **286**(24): 21678-21686.

Evans, P. (2006). "Scaling and assessment of data quality." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 72-82.

Evans, P. R. (2011). "An introduction to data reduction: space-group determination, scaling and intensity statistics." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 282-292.

Everett, H. and G. McFadden (1999). "Apoptosis: an innate immune response to virus infection." Trends Microbiol **7**(4): 160-165.

Ferron, F., Z. Li, et al. (2011). "The hexamer structure of Rift Valley fever virus nucleoprotein suggests a mechanism for its assembly into ribonucleoprotein complexes." PLoS Pathog **7**(5): e1002030.

Flick, K., A. Katz, et al. (2004). "Functional analysis of the noncoding regions of the Uukuniemi virus (Bunyaviridae) RNA segments." J Virol **78**(21): 11726-11738.

Foulke, R. S., R. R. Rosato, et al. (1981). "Structural polypeptides of Hazara virus." J Gen Virol **53**(Pt 1): 169-172.

Frese, M., G. Kochs, et al. (1996). "Inhibition of bunyaviruses, phleboviruses, and hantaviruses by human MxA protein." J Virol **70**(2): 915-923.

Frias-Staheli, N., N. V. Giannakopoulos, et al. (2007). "Ovarian tumor domain-containing viral proteases evade ubiquitin- and ISG15-dependent innate immune responses." Cell Host Microbe **2**(6): 404-416.

Fujii, K., Y. Fujii, et al. (2005). "Importance of both the coding and the segment-specific noncoding regions of the influenza A virus NS segment for its efficient incorporation into virions." J Virol **79**(6): 3766-3774.

Fujii, Y., H. Goto, et al. (2003). "Selective incorporation of influenza virus RNA segments into virions." Proc Natl Acad Sci U S A **100**(4): 2002-2007.

Gao, J., S. Xiao, et al. (2014). "Inhibition of HSP70 reduces porcine reproductive and respiratory syndrome virus replication in vitro." BMC Microbiol **14**: 64.

Garcin, D., M. Lezzi, et al. (1995). "The 5' ends of Hantaan virus (Bunyaviridae) RNAs suggest a prime-and-realign mechanism for the initiation of RNA synthesis." J Virol **69**(9): 5754-5762.

Garrison, A. R., S. R. Radoshitzky, et al. (2013). "Crimean-Congo hemorrhagic fever virus utilizes a clathrin- and early endosome-dependent entry pathway." Virology **444**(1-2): 45-54.

Gauliard, N., A. Billecocq, et al. (2006). "Rift Valley fever virus noncoding regions of L, M and S segments regulate RNA synthesis." Virology **351**(1): 170-179.

Geisbert, T. W., L. E. Hensley, et al. (2003). "Treatment of Ebola virus infection with a recombinant inhibitor of factor VIIa/tissue factor: a study in rhesus monkeys." Lancet **362**(9400): 1953-1958.

Gerrard, S. R., L. Li, et al. (2004). "Ngari virus is a Bunyamwera virus reassortant that can be associated with large outbreaks of hemorrhagic fever in Africa." J Virol **78**(16): 8922-8926.

Goedhals, D., P. A. Bester, et al. (2014). "Next-generation sequencing of southern African Crimean-Congo haemorrhagic fever virus isolates reveals a high frequency of M segment reassortment." Epidemiol Infect: 1-11.

Gonzalez, O., V. Fontanes, et al. (2009). "The heat shock protein inhibitor Quercetin attenuates hepatitis C virus production." Hepatology **50**(6): 1756-1764.

Goubau, D., S. Deddouche, et al. (2013). "Cytosolic sensing of viruses." Immunity **38**(5): 855-869.

Gouet, P., X. Robert, et al. (2003). "ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins." Nucleic Acids Res **31**(13): 3320-3323.

Gould, E. A., S. Higgs, et al. (2006). "Potential arbovirus emergence and implications for the United Kingdom." Emerg Infect Dis **12**(4): 549-555.

Graham, F. L., J. Smiley, et al. (1977). "Characteristics of a human cell line transformed by DNA from human adenovirus type 5." J Gen Virol **36**(1): 59-74.

Green, T. J., R. Cox, et al. (2014). "Common mechanism for RNA encapsidation by negative-strand RNA viruses." J Virol **88**(7): 3766-3775.

Guo, Y., W. Wang, et al. (2012). "Crimean-Congo hemorrhagic fever virus nucleoprotein reveals endonuclease activity in bunyaviruses." Proc Natl Acad Sci U S A **109**(13): 5046-5051.

Habjan, M., I. Andersson, et al. (2008). "Processing of genome 5' termini as a strategy of negative-strand RNA viruses to avoid RIG-I-dependent interferon induction." PLoS One **3**(4): e2032.

Halder, U. C., P. Bagchi, et al. (2011). "Cell death regulation during influenza A virus infection by matrix (M1) protein: a model of viral control over the cellular survival pathway." Cell Death Dis **2**: e197.

Haque, A. and M. A. Mir (2010). "Interaction of hantavirus nucleocapsid protein with ribosomal protein S19." J Virol **84**(23): 12450-12453.

Hastie, K. M., C. R. Kimberlin, et al. (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**(6): 2396-2401.

Hastie, K. M., T. Liu, et al. (2011). "Crystal structure of the Lassa virus nucleoprotein-RNA complex reveals a gating mechanism for RNA binding." Proc Natl Acad Sci U S A **108**(48): 19365-19370.

Hay, S. and G. Kannourakis (2002). "A time to kill: viral manipulation of the cell death program." J Gen Virol **83**(Pt 7): 1547-1564.

Hewson, R., J. Chamberlain, et al. (2004). "Crimean-Congo haemorrhagic fever virus: sequence analysis of the small RNA segments from a collection of viruses world wide." Virus Res **102**(2): 185-189.

Hewson, R., A. Gmyl, et al. (2004). "Evidence of segment reassortment in Crimean-Congo haemorrhagic fever virus." J Gen Virol **85**(Pt 10): 3059-3070.

Hill, M. M., C. Adrain, et al. (2004). "Analysis of the composition, assembly kinetics and activity of native Apaf-1 apoptosomes." EMBO J **23**(10): 2134-2145.

Hollidge, B. S., S. R. Weiss, et al. (2011). "The role of interferon antagonist, non-structural proteins in the pathogenesis and emergence of arboviruses." Viruses **3**(6): 629-658.

Honig, J. E., J. C. Osborne, et al. (2004). "Crimean-Congo hemorrhagic fever virus genome L RNA segment and encoded protein." Virology **321**(1): 29-35.

Hoogstraal, H. (1979). "The epidemiology of tick-borne Crimean-Congo hemorrhagic fever in Asia, Europe, and Africa." J Med Entomol **15**(4): 307-417.

Hornok, S. and G. Horvath (2012). "First report of adult Hyalomma marginatum rufipes (vector of Crimean-Congo haemorrhagic fever virus) on cattle under a continental climate in Hungary." Parasit Vectors **5**: 170.

Ikegami, T., S. Won, et al. (2005). "Rift Valley fever virus NSs mRNA is transcribed from an incoming anti-viral-sense S RNA segment." J Virol **79**(18): 12106-12111.

Ishihama, Y., Y. Oda, et al. (2005). "Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein." Mol Cell Proteomics **4**(9): 1265-1272.

James, T. W., N. Frias-Staheli, et al. (2011). "Structural basis for the removal of ubiquitin and interferon-stimulated gene 15 by a viral ovarian tumor domain-containing protease." Proc Natl Acad Sci U S A **108**(6): 2222-2227.

Jameson, L. J. and J. M. Medlock (2011). "Tick surveillance in Great Britain." Vector Borne Zoonotic Dis **11**(4): 403-412.

Jiao, L., S. Ouyang, et al. (2013). "Structure of severe fever with thrombocytopenia syndrome virus nucleocapsid protein in complex with suramin reveals therapeutic potential." J Virol **87**(12): 6829-6839.

Jin, H. and R. M. Elliott (1993). "Non-viral sequences at the 5' ends of Dugbe nairovirus S mRNAs." J Gen Virol **74 (Pt 10)**: 2293-2297.

Jones, L. D., C. R. Davies, et al. (1987). "Reassortment of Thogoto virus (a tick-borne influenza-like virus) in a vertebrate host." J Gen Virol **68 (Pt 5)**: 1299-1306.

Jones, L. D., C. R. Davies, et al. (1987). "A novel mode of arbovirus transmission involving a nonviremic host." Science **237**(4816): 775-777.

Joubert, J. R., J. B. King, et al. (1985). "A nosocomial outbreak of Crimean-Congo haemorrhagic fever at Tygerberg Hospital. Part III. Clinical pathology and pathogenesis." S Afr Med J **68**(10): 722-728.

Kabsch, W. (2010). "Xds." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 125-132.

Kanerva, M., K. Melen, et al. (1996). "Inhibition of puumala and tula hantaviruses in Vero cells by MxA protein." Virology **224**(1): 55-62.

Kang, J. I., S. H. Park, et al. (1999). "Apoptosis is induced by hantaviruses in cultured cells." Virology **264**(1): 99-105.

Kao, R. Y., D. Yang, et al. (2010). "Identification of influenza A nucleoprotein as an antiviral target." Nat Biotechnol **28**(6): 600-605.

Karlberg, H., Y. J. Tan, et al. (2011). "Induction of caspase activation and cleavage of the viral nucleocapsid protein in different cell types during Crimean-Congo hemorrhagic fever virus infection." J Biol Chem **286**(5): 3227-3234.

Karti, S. S., Z. Odabasi, et al. (2004). "Crimean-Congo hemorrhagic fever in Turkey." Emerg Infect Dis **10**(8): 1379-1384.

Kaya, A., A. Engin, et al. (2011). "Crimean-Congo hemorrhagic fever disease due to tick bite with very long incubation periods." Int J Infect Dis **15**(7): e449-452.

Khachatoorian, R., E. Ganapathy, et al. (2014). "The NS5A-binding heat shock proteins HSC70 and HSP70 play distinct roles in the hepatitis C viral life cycle." Virology **454-455**: 118-127.

Kinsella, E., S. G. Martin, et al. (2004). "Sequence determination of the Crimean-Congo hemorrhagic fever virus L segment." Virology **321**(1): 23-28.

Kohl, A., R. F. Clayton, et al. (2003). "Bunyamwera virus nonstructural protein NSs counteracts interferon regulatory factor 3-mediated induction of early cell death." J Virol **77**(14): 7999-8008.

Kohl, A., E. F. Dunn, et al. (2004). "Complementarity, sequence and structural elements within the 3' and 5' non-coding regions of the Bunyamwera orthobunyavirus S segment determine promoter strength." J Gen Virol **85**(Pt 11): 3269-3278.

Koksal, I., G. Yilmaz, et al. (2010). "The efficacy of ribavirin in the treatment of Crimean-Congo hemorrhagic fever in Eastern Black Sea region in Turkey." J Clin Virol **47**(1): 65-68.

Kondoh, Y. and H. Osada (2013). "High-throughput screening identifies small molecule inhibitors of molecular chaperones." Curr Pharm Des **19**(3): 473-492.

Kuismanen, E., K. Hedman, et al. (1982). "Uukuniemi virus maturation: accumulation of virus particles and viral antigens in the Golgi complex." Mol Cell Biol **2**(11): 1444-1458.

Lahaye, X., A. Vidy, et al. (2012). "Hsp70 protein positively regulates rabies virus infection." J Virol **86**(9): 4743-4751.

Leibovitz, A., W. M. McCombs, 3rd, et al. (1973). "New human cancer cell culture lines. I. SW-13, small-cell carcinoma of the adrenal cortex." J Natl Cancer Inst **51**(2): 691-697.

Lenz, O., J. ter Meulen, et al. (2001). "The Lassa virus glycoprotein precursor GP-C is proteolytically processed by subtilase SKI-1/S1P." Proc Natl Acad Sci U S A **98**(22): 12701-12705.

Leslie, A. G. (2006). "The integration of macromolecular diffraction data." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 48-57.

Lewis, J., S. L. Wesselingh, et al. (1996). "Alphavirus-induced apoptosis in mouse brains correlates with neurovirulence." J Virol **70**(3): 1828-1835.

Liu, T., C. K. Daniels, et al. (2012). "Comprehensive review on the HSC70 functions, interactions with related molecules and involvement in clinical diseases and therapeutic potential." Pharmacol Ther **136**(3): 354-374.

Lopez-Hernandez, F. J., M. A. Ortiz, et al. (2003). "Z-FA-fmk inhibits effector caspases but not initiator caspases 8 and 10, and demonstrates that novel anticancer retinoid-related molecules induce apoptosis via the intrinsic pathway." Mol Cancer Ther **2**(3): 255-263.

Lowen, A. C., A. Boyd, et al. (2005). "Attenuation of bunyavirus replication by rearrangement of viral coding and noncoding sequences." J Virol **79**(11): 6940-6946.

Lowen, A. C. and R. M. Elliott (2005). "Mutational analyses of the nonconserved sequences in the Bunyamwera Orthobunyavirus S segment untranslated regions." J Virol **79**(20): 12861-12870.

Lukashev, A. N. (2005). "Evidence for recombination in Crimean-Congo hemorrhagic fever virus." J Gen Virol **86**(Pt 8): 2333-2338.

M. P. Chumakov, A. A. S. a. A. A. K. (1945). "A new tick-borne virus disease-Crimean hemorrhagic fever." *Crimean hemorrhagic fever*: 13-43.

Manigold, T. and P. Vial (2014). "Human hantavirus infections: epidemiology, clinical features, pathogenesis and immunology." Swiss Med Wkly **144**: w13937.

Mann, M. (2006). "Functional and quantitative proteomics using SILAC." Nat Rev Mol Cell Biol **7**(12): 952-958.

Manzoor, R., K. Kuroda, et al. (2014). "Heat shock protein 70 modulates influenza A virus polymerase activity." J Biol Chem **289**(11): 7599-7614.

Marczinke, B. I. and S. T. Nichol (2002). "Nairobi sheep disease virus, an important tick-borne pathogen of sheep and goats in Africa, is also present in Asia." Virology **303**(1): 146-151.

Markotic, A., L. Hensley, et al. (2003). "Hantaviruses induce cytopathic effects and apoptosis in continuous human embryonic kidney cells." J Gen Virol **84**(Pt 8): 2197-2202.

Marriott, A. C. and P. A. Nuttall (1992). "Comparison of the S RNA segments and nucleoprotein sequences of Crimean-Congo hemorrhagic fever, Hazara, and Dugbe viruses." Virology **189**(2): 795-799.

Martin, M. L., H. Lindsey-Regnery, et al. (1985). "Distinction between Bunyaviridae genera by surface structure and comparison with Hantaan virus using negative stain electron microscopy." Arch Virol **86**(1-2): 17-28.

Marzluff, W. F. (1992). "Histone 3' ends: essential and regulatory functions." Gene Expr **2**(2): 93-97.

Massey, A. J., D. S. Williamson, et al. (2010). "A novel, small molecule inhibitor of Hsc70/Hsp70 potentiates Hsp90 inhibitor induced apoptosis in HCT116 colon carcinoma cells." Cancer Chemother Pharmacol **66**(3): 535-545.

Matsuoka, Y., S. Y. Chen, et al. (1991). "Bunyavirus protein transport and assembly." Curr Top Microbiol Immunol **169**: 161-179.

Mayer, M. P. (2005). "Recruitment of Hsp70 chaperones: a crucial part of viral survival strategies." Rev Physiol Biochem Pharmacol **153**: 1-46.

McCoy, A. J., R. W. Grosse-Kunstleve, et al. (2007). "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.

McJunkin, J. E., E. C. de los Reyes, et al. (2001). "La Crosse encephalitis in children." N Engl J Med **344**(11): 801-807.

McLean, J. E., E. Datan, et al. (2009). "Lack of Bax prevents influenza A virus-induced apoptosis and causes diminished viral replication." J Virol **83**(16): 8233-8246.

Medzhitov, R. and C. A. Janeway, Jr. (1999). "Innate immune induction of the adaptive immune response." Cold Spring Harb Symp Quant Biol **64**: 429-435.

Mi, J., Z. Y. Li, et al. (2001). "Induced apoptosis supports spread of adenovirus vectors in tumors." Hum Gene Ther **12**(10): 1343-1352.

Mild, M., M. Simon, et al. (2010). "Towards an understanding of the migration of Crimean-Congo hemorrhagic fever virus." J Gen Virol **91**(Pt 1): 199-207.

Mir, M. A., W. A. Duran, et al. (2008). "Storage of cellular 5' mRNA caps in P bodies for viral cap-snatching." Proc Natl Acad Sci U S A **105**(49): 19294-19299.

Muller, R., O. Poch, et al. (1994). "Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases." J Gen Virol **75 (Pt 6)**: 1345-1352.

Murshudov, G. N., P. Skubak, et al. (2011). "REFMAC5 for the refinement of macromolecular crystal structures." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 355-367.

Nakabayashi, H., K. Taketa, et al. (1982). "Growth of human hepatoma cells lines with differentiated functions in chemically defined medium." Cancer Res **42**(9): 3858-3863.

Ng, A. K., M. K. Lam, et al. (2012). "Structural basis for RNA binding and homo-oligomer formation by influenza B virus nucleoprotein." J Virol **86**(12): 6758-6767.

Niu, F., N. Shaw, et al. (2013). "Structure of the Leanyer orthobunyavirus nucleoprotein-RNA complex reveals unique architecture for RNA encapsidation." Proc Natl Acad Sci U S A **110**(22): 9054-9059.

Novoa, R. R., G. Calderita, et al. (2005). "Key Golgi factors for structural and functional maturation of bunyamwera virus." J Virol **79**(17): 10852-10863.

Nuss, J. E., K. Kehn-Hall, et al. (2014). "Multi-Faceted Proteomic Characterization of Host Protein Complement of Rift Valley Fever Virus Virions and Identification of Specific Heat Shock Proteins, Including HSP90, as Important Viral Host Factors." PLoS One **9**(5): e93483.

Nuttall, P. A. and M. Labuda (2003). "Dynamics of infection in tick vectors and at the tick-host interface." Adv Virus Res **60**: 233-272.

Nuttall, P. A. and M. Labuda (2004). "Tick-host interactions: saliva-activated transmission." Parasitology **129 Suppl**: S177-189.

Nylandsted, J., K. Brand, et al. (2000). "Heat shock protein 70 is required for the survival of cancer cells." Ann N Y Acad Sci **926**: 122-125.

Obijeski, J. F., D. H. Bishop, et al. (1976). "Segmented genome and nucleocapsid of La Crosse virus." J Virol **20**(3): 664-675.

Oestereich, L., T. Rieger, et al. (2014). "Evaluation of antiviral efficacy of ribavirin, arbidol, and T-705 (favipiravir) in a mouse model for Crimean-Congo hemorrhagic fever." PLoS Negl Trop Dis **8**(5): e2804.

Oglesbee, M., S. Ringler, et al. (1990). "Interaction of canine distemper virus nucleocapsid variants with 70K heat-shock proteins." J Gen Virol **71 (Pt 7)**: 1585-1590.

Oglesbee, M. J., Z. Liu, et al. (1996). "The highly inducible member of the 70 kDa family of heat shock proteins increases canine distemper virus polymerase activity." J Gen Virol **77 (Pt 9)**: 2125-2135.

Oliveira, A. P., F. M. Simabuco, et al. (2013). "Human respiratory syncytial virus N, P and M protein interactions in HEK-293T cells." Virus Res **177**(1): 108-112.

Ong, S. E., B. Blagoev, et al. (2002). "Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics." Mol Cell Proteomics **1**(5): 376-386.

Onguru, P., S. Dagdas, et al. (2010). "Coagulopathy parameters in patients with Crimean-Congo hemorrhagic fever and its relation with mortality." J Clin Lab Anal **24**(3): 163-166.

Ontiveros, S. J., Q. Li, et al. (2010). "Modulation of apoptosis and immune signaling pathways by the Hantaan virus nucleocapsid protein." Virology **401**(2): 165-178.

Osborne, J. C. and R. M. Elliott (2000). "RNA binding properties of bunyamwera virus nucleocapsid protein and selective binding to an element in the 5' terminus of the negative-sense S segment." J Virol **74**(21): 9946-9952.

Overby, A. K., R. F. Pettersson, et al. (2007). "The glycoprotein cytoplasmic tail of Uukuniemi virus (Bunyaviridae) interacts with ribonucleoproteins and is critical for genome packaging." J Virol **81**(7): 3198-3205.

Overby, A. K., V. L. Popov, et al. (2007). "The cytoplasmic tails of Uukuniemi Virus (Bunyaviridae) G(N) and G(C) glycoproteins are important for intracellular targeting and the budding of virus-like particles." J Virol **81**(20): 11381-11391.

Papa, A., S. Bino, et al. (2006). "Cytokine levels in Crimean-Congo hemorrhagic fever." J Clin Virol **36**(4): 272-276.

Papa, A., I. Christova, et al. (2004). "Crimean-Congo hemorrhagic fever in Bulgaria." Emerg Infect Dis **10**(8): 1465-1467.

Papa, A., V. Dalla, et al. (2010). "Emergence of Crimean-Congo haemorrhagic fever in Greece." Clin Microbiol Infect **16**(7): 843-847.

Papa, A., E. Papadimitriou, et al. (2011). "The Bulgarian vaccine Crimean-Congo haemorrhagic fever virus strain." Scand J Infect Dis **43**(3): 225-229.

Pappu, H. R., R. A. Jones, et al. (2009). "Global status of tospovirus epidemics in diverse cropping systems: successes achieved and challenges ahead." Virus Res **141**(2): 219-236.

Parent, R., X. Qu, et al. (2009). "The heat shock cognate protein 70 is associated with hepatitis C virus particles and modulates virus infectivity." Hepatology **49**(6): 1798-1809.

Pekosz, A., J. Phillips, et al. (1996). "Induction of apoptosis by La Crosse virus infection and role of neuronal differentiation and human bcl-2 expression in its prevention." J Virol **70**(8): 5329-5335.

Pettersson, R. F. and C. H. von Bonsdorff (1975). "Ribonucleoproteins of Uukuniemi virus are circular." J Virol **15**(2): 386-392.

Peyrefitte, C. N., M. Perret, et al. (2010). "Differential activation profiles of Crimean-Congo hemorrhagic fever virus- and Dugbe virus-infected antigen-presenting cells." J Gen Virol **91**(Pt 1): 189-198.

Pinheiro, F. P., A. P. Travassos da Rosa, et al. (1981). "Oropouche virus. I. A review of clinical, epidemiological, and ecological findings." Am J Trop Med Hyg **30**(1): 149-160.

Postigo, A. and P. E. Ferrer (2009). "Viral inhibitors reveal overlapping themes in regulation of cell death and innate immunity." Microbes Infect **11**(13): 1071-1078.

Potterton, E., P. Briggs, et al. (2003). "A graphical user interface to the CCP4 program suite." Acta Crystallogr D Biol Crystallogr **59**(Pt 7): 1131-1137.

Qi, X., S. Lan, et al. (2010). "Cap binding and immune evasion revealed by Lassa nucleoprotein structure." Nature **468**(7325): 779-783.

Radhakrishnan, A., D. Yeo, et al. (2010). "Protein analysis of purified respiratory syncytial virus particles reveals an important role for heat shock protein 90 in virus particle assembly." Mol Cell Proteomics **9**(9): 1829-1848.

Raju, R. and D. Kolakofsky (1989). "The ends of La Crosse virus genome and antigenome RNAs within nucleocapsids are base paired." J Virol **63**(1): 122-128.

Ramalingam, D., S. Duclair, et al. (2011). "RNA aptamers directed to human immunodeficiency virus type 1 Gag polyprotein bind to the matrix and nucleocapsid domains and inhibit virus production." J Virol **85**(1): 305-314.

Ramanathan, H. N., D. H. Chung, et al. (2007). "Dynein-dependent transport of the hantaan virus nucleocapsid protein to the endoplasmic reticulum-Golgi intermediate compartment." J Virol **81**(16): 8634-8647.

Raymond, D. D., M. E. Piper, et al. (2012). "Phleboviruses encapsidate their genomes by sequestering RNA bases." Proc Natl Acad Sci U S A **109**(47): 19208-19213.

Raymond, D. D., M. E. Piper, et al. (2010). "Structure of the Rift Valley fever virus nucleocapsid protein reveals another architecture for RNA encapsidation." Proc Natl Acad Sci U S A **107**(26): 11769-11774.

Reguera, J., S. Cusack, et al. (2014). "Segmented negative strand RNA virus nucleoprotein structure." Curr Opin Virol **5C**: 7-15.

Reguera, J., H. Malet, et al. (2013). "Structural basis for encapsidation of genomic RNA by La Crosse Orthobunyavirus nucleoprotein." Proc Natl Acad Sci U S A **110**(18): 7246-7251.

Reguera, J., F. Weber, et al. (2010). "Bunyaviridae RNA polymerases (L-protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription." PLoS Pathog **6**(9): e1001101.

Ren, X., C. Xue, et al. (2012). "Proteomic analysis of purified Newcastle disease virus particles." Proteome Sci **10**(1): 32.

Ribeiro, D., J. W. Borst, et al. (2009). "Tomato spotted wilt virus nucleocapsid protein interacts with both viral glycoproteins Gn and Gc in planta." Virology **383**(1): 121-130.

Richard, A. and D. Tulasne (2012). "Caspase cleavage of viral proteins, another way for viruses to make the best of apoptosis." Cell Death Dis **3**: e277.

Rodrigues, R., G. Paranhos-Baccala, et al. (2012). "Crimean-Congo hemorrhagic fever virus-infected hepatocytes induce ER-stress and apoptosis crosstalk." PLoS One **7**(1): e29712.

Saelens, X., N. Festjens, et al. (2004). "Toxic proteins released from mitochondria in cell death." Oncogene **23**(16): 2861-2874.

Sagara, J. and A. Kawai (1992). "Identification of heat shock protein 70 in the rabies virion." Virology **190**(2): 845-848.

Saksida, A., D. Duh, et al. (2010). "Interacting roles of immune mechanisms and viral load in the pathogenesis of crimean-congo hemorrhagic fever." Clin Vaccine Immunol **17**(7): 1086-1093.

Sanchez, A. J., M. J. Vincent, et al. (2006). "Crimean-congo hemorrhagic fever virus glycoprotein precursor is cleaved by Furin-like and SKI-1 proteases to generate a novel 38-kilodalton glycoprotein." J Virol **80**(1): 514-525.

Sanchez, A. J., M. J. Vincent, et al. (2002). "Characterization of the glycoproteins of Crimean-Congo hemorrhagic fever virus." J Virol **76**(14): 7263-7275.

Sandes, E., C. Lodillinsky, et al. (2007). "Cathepsin B is involved in the apoptosis intrinsic pathway induced by Bacillus Calmette-Guerin in transitional cancer cell lines." Int J Mol Med **20**(6): 823-828.

Santoro, M. G. (1994). "Heat shock proteins and virus replication: hsp70s as mediators of the antiviral effects of prostaglandins." Experientia **50**(11-12): 1039-1047.

Schlecht, R., S. R. Scholz, et al. (2013). "Functional analysis of Hsp70 inhibitors." PLoS One **8**(11): e78443.

Schnittler, H. J. and H. Feldmann (2003). "Viral hemorrhagic fever--a vascular disease?" Thromb Haemost **89**(6): 967-972.

Schotte, P., W. Declercq, et al. (1999). "Non-specific effects of methyl ketone peptide inhibitors of caspases." FEBS Lett **442**(1): 117-121.

Schuler, M. and D. R. Green (2001). "Mechanisms of p53-dependent apoptosis." Biochem Soc Trans **29**(Pt 6): 684-688.

Semiatkovskaia, Z. V. and N. K. Sidtdykova (1950). "[Clinical aspect of infectious hemorrhagic fever]." Klin Med (Mosk) **28**(8): 69-71.

Shapiro, S. E. and Z. S. Barkagan (1960). "[On history of hemorrhagic fever in Middle Asia]." Vopr Virusol **5**: 245-246.

Shaw, M. L., K. L. Stone, et al. (2008). "Cellular proteins in influenza virus particles." PLoS Pathog **4**(6): e1000085.

Shepherd, A. J., R. Swanepoel, et al. (1989). "Antibody response in Crimean-Congo hemorrhagic fever." Rev Infect Dis **11 Suppl 4**: S801-806.

Simmons, G., R. J. Wool-Lewis, et al. (2002). "Ebola virus glycoproteins induce global surface protein down-modulation and loss of cell adherence." J Virol **76**(5): 2518-2528.

Simon, M., C. Johansson, et al. (2009). "Microtubule-dependent and microtubule-independent steps in Crimean-Congo hemorrhagic fever virus replication cycle." Virology **385**(2): 313-322.

Simon, M., C. Johansson, et al. (2009). "Crimean-Congo hemorrhagic fever virus entry and replication is clathrin-, pH- and cholesterol-dependent." J Gen Virol **90**(Pt 1): 210-215.

Simpson, D. I., E. M. Knight, et al. (1967). "Congo virus: a hitherto undescribed virus occurring in Africa. I. Human isolations--clinical notes." East Afr Med J **44**(2): 86-92.

Smirnova, S. E. (1979). "A comparative study of the Crimean hemorrhagic fever-Congo group of viruses." Arch Virol **62**(2): 137-143.

Smirnova, S. E., N. M. Shestopalova, et al. (1977). "Experimental Hazara Virus infection in mice." Acta Virol **21**(2): 128-132.

Smith, D. B. and K. S. Johnson (1988). "Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase." Gene **67**(1): 31-40.

Soares-Weiser, K., S. Thomas, et al. (2010). "Ribavirin for Crimean-Congo hemorrhagic fever: systematic review and meta-analysis." BMC Infect Dis **10**: 207.

Song, J. W., L. J. Baek, et al. (2007). "Thottapalayam virus, a prototype shrewborne hantavirus." Emerg Infect Dis **13**(7): 980-985.

Stegen, C., Y. Yakova, et al. (2013). "Analysis of virion-incorporated host proteins required for herpes simplex virus type 1 infection through a RNA interference screen." PLoS One **8**(1): e53276.

Sylvester, P. W. (2011). "Optimization of the tetrazolium dye (MTT) colorimetric assay for cellular growth and viability." Methods Mol Biol **716**: 157-168.

Takeuchi, O. and S. Akira (2010). "Pattern recognition receptors and inflammation." Cell **140**(6): 805-820.

Tasdelen Fisgin, N., L. Doganci, et al. (2010). "Initial high rate of misdiagnosis in Crimean Congo haemorrhagic fever patients in an endemic region of Turkey." Epidemiol Infect **138**(1): 139-144.

Thiede, B., S. Lamer, et al. (2000). "Analysis of missed cleavage sites, tryptophan oxidation and N-terminal pyroglutamylation after in-gel tryptic digestion." Rapid Commun Mass Spectrom **14**(6): 496-502.

Trinkle-Mulcahy, L., S. Boulon, et al. (2008). "Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes." The Journal of cell biology **183**(2): 223-239.

van Kasteren, P. B., B. A. Bailey-Elkin, et al. (2013). "Deubiquitinase function of arterivirus papain-like protease 2 suppresses the innate immune response in infected host cells." Proc Natl Acad Sci U S A **110**(9): E838-847.

van Knippenberg, I., M. Lamine, et al. (2005). "Tomato spotted wilt virus transcriptase in vitro displays a preference for cap donors with multiple base complementarity to the viral template." Virology **335**(1): 122-130.

Vasconcelos, H. B., M. R. Nunes, et al. (2011). "Molecular epidemiology of Oropouche virus, Brazil." Emerg Infect Dis **17**(5): 800-806.

Vincent, M. J., A. J. Sanchez, et al. (2003). "Crimean-Congo hemorrhagic fever virus glycoprotein proteolytic processing by subtilase SKI-1." J Virol **77**(16): 8640-8649.

Walter, C. T. and J. N. Barr (2011). "Recent advances in the molecular and cellular biology of bunyaviruses." J Gen Virol **92**(Pt 11): 2467-2484.

Wang, Y., S. Dutta, et al. (2012). "Structure of Crimean-Congo hemorrhagic fever virus nucleoprotein: superhelical homo-oligomers and the role of caspase-3 cleavage." J Virol **86**(22): 12294-12303.

Weber, F., A. Bridgen, et al. (2002). "Bunyamwera bunyavirus nonstructural protein NSs counteracts the induction of alpha/beta interferon." J Virol **76**(16): 7949-7955.

Weber, F. and A. Mirazimi (2008). "Interferon and cytokine responses to Crimean Congo hemorrhagic fever virus; an emerging and neglected viral zonoosis." Cytokine Growth Factor Rev **19**(5-6): 395-404.

Whitehouse, C. A. (2004). "Crimean-Congo hemorrhagic fever." Antiviral Res **64**(3): 145-160.

Winn, M. D., C. C. Ballard, et al. (2011). "Overview of the CCP4 suite and current developments." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 235-242.

Woodall, J. P., M. C. Williams, et al. (1967). "Congo virus: a hitherto undescribed virus occurring in Africa. II. Identification studies." East Afr Med J **44**(2): 93-98.

Xiao, X., Y. Feng, et al. (2011). "Identification of a putative Crimean-Congo hemorrhagic fever virus entry factor." Biochem Biophys Res Commun **411**(2): 253-258.

Ye, Q., R. M. Krug, et al. (2006). "The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA." Nature **444**(7122): 1078-1082.

Yesilbag, K., L. Aydin, et al. (2013). "Tick survey and detection of Crimean-Congo hemorrhagic fever virus in tick species from a non-endemic area, South Marmara region, Turkey." Exp Appl Acarol **60**(2): 253-261.

Young, P. R. and C. R. Howard (1983). "Fine structure analysis of Pichinde virus nucleocapsids." J Gen Virol **64 (Pt 4)**: 833-842.

Yu, L., L. Ye, et al. (2009). "HSP70 induced by Hantavirus infection interacts with viral nucleocapsid protein and its overexpression suppresses virus infection in Vero E6 cells." Am J Transl Res **1**(4): 367-380.

Zhang, Y., L. Li, et al. (2013). "Crystal structure of Junin virus nucleoprotein." J Gen Virol **94**(Pt 10): 2175-2183.

Zheng, W., J. Olson, et al. (2013). "The crystal structure and RNA-binding of an orthomyxovirus nucleoprotein." PLoS Pathog **9**(9): e1003624.

Zhou, H., Y. Sun, et al. (2013). "The nucleoprotein of severe fever with thrombocytopenia syndrome virus processes a stable hexameric ring to facilitate RNA encapsidation." Protein Cell **4**(6): 445-455.

Zhou, Q., J. F. Krebs, et al. (1998). "Interaction of the baculovirus anti-apoptotic protein p35 with caspases. Specificity, kinetics, and characterization of the caspase/p35 complex." Biochemistry **37**(30): 10757-10765.

Zivcec, M., D. Safronetz, et al. (2013). "Lethal Crimean-Congo hemorrhagic fever virus infection in interferon alpha/beta receptor knockout mice is associated with high viral loads, proinflammatory responses, and coagulopathy." J Infect Dis **207**(12): 1909-1921.

APPENDIX I CCHFV N AND HAZV N NUCLEOTIDE SEQUENCES

Start codon, stop codon

CCHFV N strain 10200 nucleotide sequence

ATGGAAAACAAGATCGAGGTGAATAACAAAGATGAGATGAACAGGTGGTTTGAA
GAGTTCAAAAAAGGAAATGGACTTGTGGACACCTTCACAAACTCCTATTCCTTTT
GCGAGAGTGTTCCCAATTTGGACAGGTTTGTGTTTCAGATGGCCAGTGCCACCG
ATGATGCACAGAAGGACTCCATCTACGCATCTGCTCTGGTGGAGGCAACAAAGT
TTTGTGCACCTATATATGAGTGCGCATGGGTTAGCTCCACTGGCATTGTAAAAAA
GGGACTTGAATGGTTCGAGAAAAATGCAGGAACCATTAAGTCCTGGGATGAAAG
TTATACTGAGCTAAAGGTCGACGTCCCGAAAATAGAGCAGCTTACCGGTTACCAA
CAAGCTGCCTTGAAGTGGAGAAAAGACATAGGTTTCCGTGTCAATGCCAACACA
GCAGCTCTGAGCAACAAAGTCCTCGCAGAATACAAAGTCCCTGGTGAGATTGTG
ATGTCTGTCAAAGAGATGCTGTCAGACATGATTAGGAGAAGGAACCTGATTCTAA
ACAGGGGTGGTGATGAGAACCCACGTGGCCCAGTGAGCCATGAGCATGTAGAC
TGGTGCAGGGAGTTTGTCAAAGGCAAATACATCATGGCCTTCAACCCACCATGG
GGGGACATCAACAAGTCAGGCCGTTCAGGAATAGCACTTGTTGCAACAGGCCTT
GCTAAGCTTGCAGAGACTGAAGGAAAGGGAATATTTGATGAAGCCAAAAAGACT
GTGGAGGCCCTCAACGGGTATCTGGACAAGCATAAGGACGAAGTTGATAGAGCA
AGCGCCGACAGCATGATAACAAACCTTCTTAAGCATATTGCCAAGGCACAGGAG
CTCTATAAAAATTCATCTGCACTTCGTGCACAAAGCGCACAGATTGACACTGCTT
TCAGCTCATACTATTGGCTTTACAAGGCTGGCGTGACTCCTGAAACCTTCCCGAC
GGTGTCACAGTTCCTCTTTGAGCTAGGGAAACAGCCAAGAGGTACCAAGAAAAT
GAAGAAGGCTCTTCTGAGCACCCCAATGAAGTGGGGGAAGAAGCTTTATGAGCT
CTTTGCCGATGATTCTTTCCAGCAGAACAGGATTTACATGCATCCTGCCGTGCTT
ACAGCTGGTAGAATCAGTGAAATGGGAGTCTGCTTTGGGACAATCCCTGTGGCC
AATCCTGATGATGCTGCCCAAGGATCTGGACACACTAAGTCTATTCTCAACCTCC
GTACCAACACTGAGACCAATAATCCGTGTGCCAAAACCATCGTCAAGCTATTTGA
AGTTCAAAAAACAGGGTTCAACATTCAGGACATGGACATAGTGGCCTCTGAGCA
CTTGCTACACCAATCCCTTGTTGGCAAGCAATCCCCATTCCAGAACGCCTACAAC
GTCAAGGGCAATGCCACCAGTGCTAACATCATTTAA

Appendix I

HAZV N strain JC280 nucleotide sequence

ATGGAAAACAAGATCGTGGCTAGCACCAAGGAAGAGTTCAACACGTGGTACAAA
CAGTTTGCTGAAAAGCACAAACTGAATAACAAATACACGGAATCTGCATCCTTCT
GTGCAGAAATTCCACAACTGGACACTTATAAGTACAAAATGGAACTGGCATCCAC
CGATAACGAACGCGACGCGATCTACAGCTCTGCTCTGATCGAGGCTACTCGTTT
CTGTGCTCCGATCATGGAATGCGCATGGGCATCTTGCACTGGTACCGTAAAACG
TGGTCTGGAGTGGTTTGATAAGAACAAAGATTCCGATACTGTTAAAGTGTGGGAT
GCGAACTACCAGAAACTGCGTACCGAAACTCCACCTGCAGAAGCGCTGCTGGCT
TATCAGAAAGCGGCACTGAACTGGCGTAAAGATGTCGGTTTCTCCATCGGTGAG
TACACCAGCATTCTGAAGAAAGCTGTTGCCGCAGAATATAAAGTTCCGGGTACC
GTTATTAACAACATCAAAGAAATGCTGAGCGACATGATTCGTCGTCGCAACCGTA
TTATCAACGGTGGTAGCGACGACGCACCGAAACGCGGTCCGGTTGGCCGTGAG
CATCTGGATTGGTGTCGTGAGTTCGCGAGCGGCAAGTTTCTGAACGCTTTCAAT
CCGCCGTGGGGTGAGATCAACAAAGCGGGCAAGAGCGGCTATCCGCTGCTGGC
GACTGGTCTGGCTAAGCTGGTTGAACTGGAAGGCAAAGACGTTATGGACAAAGC
GAAGGCATCCATCGCTCAGCTGGAAGGTTGGGTGAAAGAGAATAAAGATCAGGT
TGACCAGGACAAAGCGGAGGATCTGCTGAAAGGCGTACGCGAATCTTACAAAAC
CGCCCTGGCGCTGGCTAAACAGTCTAATGCATTCCGCGCGCAGGGTGCGCAGA
TCGATACTGTTTTCTCCTCCTACTACTGGCTGTGGAAAGCAGGCGTAACCCCGGT
AACCTTTCCGTCCGTGTCCCAGTTCCTGTTTGAACTGGGTAAGAACCCGAAGGG
CCAGAAGAAGATGCAGAAAGCCCTGATCAACACCCCGCTGAAATGGGGCAAACG
TCTGATCGAACTGTTCGCTGACAATGACTTCACCGAAAACCGTATCTATATGCAC
CCATGCGTTCTGACCTCTGGTCGTATGTCCGAGCTGGGTATTTCCTTCGGCGCG
GTGCCAGTTACGTCTCCGGATGATGCAGCGCAGGGCAGCGGCCATACTAAAGC
GGTGCTGAACTACAAAACCAAAACCGAAGTGGGCAACCCATGTGCGTGTATTAT
CTCCTCTCTGTTCGAGATCCAGAAAGCGGGTTACGACATTGAGAGCATGGACAT
TGTTGCGAGCGAACACCTGCTGCATCAGTCCCTGGTTGGCAAACGCTCTCCGTT
CCAGAACGCCTATCTGATTAAAGGTAACGCAACTAACATTAACATTATCTAA

APPENDIX II. PROTEINS IDENTIFIED AND QUANTIFIED IN THE EGFP-CCHFV N IP

MAP[1] dataset (white) and MAP[2] dataset (green).

Accession number	Protein Name	EGFP only	EGFP -N	Peptides (seq)	PEP	Protein function
IPI00000816.1	14-3-3 protein epsilon	0.55	1.81	2	4.48E-06	Adaptor protein involved in intracellular signalling.
		0.30	3.33	4	1.1827E-15	
IPI00008599.3	3-beta-hydroxysteroid-Delta(8),Delta(7)-isomerase	0.39	2.59	2	0.019156	Involved in steroid and cholesterol biosynthesis.
IPI00011253.3	40S ribosomal protein S3	0.26	3.86	6	2.0921E-14	Involved in protein biosynthesis.
IPI00554481.1	4F2 cell-surface antigen heavy chain	0.19	5.22	6	8.79E-27	Amino acid transport into cells at the plasma membrane. Required for the function of light chain amino-acid transporters.
		0.17	5.94	6	2.7026E-13	
IPI00784402.1	5-azacytidine-induced protein 1	0.09	10.57	4	2.1667E-13	A component of the cytoskeleton, localised to the centrosomes and intracellular membrane bound organelles.
IPI00784154.1	60 kDa heat shock protein, mitochondrial	0.42	2.4	13	7.7237E-55	Mitochondrial protein import and macromolecular assembly. May facilitate the correct folding of imported proteins.
		0.40	2.48	16	5.27E-74	
IPI00008530.1	60S acidic ribosomal protein P0	0.32	3.15	2	0.00010867	Involved in protein biosynthesis.
		0.26	3.79	2	5.35E-05	
IPI00376798.3	60S ribosomal protein L11	0.55	1.83	2	4.66E-14	Binds to 5S ribosomal RNA. Required for rRNA maturation and formation of the 60S ribosomal subunits.
		0.44	2.28	2	1.3345E-13	

IPI00215790.6	60S ribosomal protein L38	0.71	1.4	2	0.00026745	Belongs to the ribosomal protein L38e family. Involved in protein biogenesis.
		0.62	1.6	2	9.47E-10	
IPI00003362.2	78 kDa glucose-regulated protein	0.20	5.1	17	2.7815E-102	Plays a role in facilitating the assembly of multimeric protein complexes inside the ER.
		0.17	5.87	13	4.74E-108	
IPI00023006.1	Actin, alpha cardiac muscle 1	0.08	13	12	1.10E-34	Highly conserved cytoskeletal protein. Involved in various types of cell motility and are ubiquitously expressed in all eukaryotic cells.
		0.07	15.02	9	2.3388E-51	
IPI00021440.1	Actin, cytoplasmic 2	0.15	6.8	19	3.68E-73	
		0.06	16.46	17	1.5247E-146	
IPI00031131.4	Adipocyte plasma membrane-associated protein	0.18	5.64	4	4.8038E-08	Type II transmembrane glycoprotein, localised to the plasma membrane.
IPI00007188.5	ADP/ATP translocase 2	0.05	18.25	7	5.52E-64	A multipass transmembrane protein located on the inner mitochondrial membrane.
		0.05	18.5	8	2.3689E-69	
IPI00291467.7	ADP/ATP translocase 3	0.16	6.32	5	6.7723E-48	Inner mitochondrial membrane protein.
IPI00465248.5	Alpha-enolase	0.34	2.96	3	1.5824E-09	Multifunctional enzyme; involved in glycolysis, growth control, and hypoxia tolerance.
IPI00001453.2	Alpha-internexin	0.22	4.47	4	4.7225E-08	Structural constituent of the cytoskeleton.
IPI00163085.2	Angiomotin	0.09	10.63	26	8.7864E-137	Cell surface protein that plays a central role in tight junction maintenance.
		0.07	13.39	8	3.81E-29	

IPI00292953.6	Ankycorbin	0.19	5.38	6	2.195E-12	Indirectly associated with the cortical actin cytoskeleton structures.
IPI00329801.12	Annexin A5	0.24	4.18	3	3.60E-13	Involved in intracellular signalling: an anti-apoptotic, anticoagulant protein.
IPI00440493.2	ATP synthase subunit alpha, mitochondrial	0.13	7.5	12	8.7414E-53	Mitochondrial membrane ATP synthase (F1F0 ATP synthase or Complex V) produces ATP from ADP in the presence of a proton gradient across the inner mitochondrial membrane.
		0.13	7.75	13	1.33E-72	
IPI00029133.4	ATP synthase subunit b, mitochondrial	0.24	4.22	2	0.00086751	
		0.18	5.42	2	0.000039858	
IPI00303476.1	ATP synthase subunit beta, mitochondrial	0.13	7.87	19	4.4014E-115	
		0.12	8.26	20	3.86E-91	
IPI00478410.2	ATP synthase subunit gamma, mitochondrial	0.18	5.7	2	1.92E-16	
IPI00007611.1	ATP synthase subunit O, mitochondrial	0.18	5.64	2	2.2646E-11	
IPI00844578.1	ATP-dependent RNA helicase A	0.19	5.2	3	2.0152E-07	Unwinds double-stranded DNA and RNA in a 3' to 5' direction. Functions as a transcriptional activator.
IPI00295992.4	ATPase family AAA domain-containing protein 3A	0.15	6.78	3	2.35E-47	A novel anti-apoptotic factor. Hydrolyses ATP.
IPI00218200.7	B-cell receptor-associated protein 31	0.29	3.4	2	0.00033456	May play a role in anterograde transport of membrane proteins from the endoplasmic reticulum to the Golgi.
IPI00218019.1	Basigin	0.18	5.65	3	3.43E-08	Cell surface receptor involved in intracellular signalling.

		0.10	9.67	5	3.1937E-24	Involved in blood coagulation, and leukocyte migration.
IPI00301263.2	CAD protein	0.20	4.99	3	4.2939E-10	Essential for pyrimidine biosynthesis.
		0.11	9.37	3	4.57E-38	
IPI00290085.2	Cadherin-2	0.18	5.44	5	4.6633E-08	Calcium dependent cell adhesion proteins. Involved in cell junction assembly.
		0.13	7.49	6	2.62E-09	
IPI00007084.2	Calcium-binding mitochondrial carrier protein Aralar2	0.08	12.33	7	1.04E-12	Calcium-dependent mitochondrial aspartate and glutamate carrier. Located on the mitchondrial inner membrane.
IPI00020984.2	Calnexin	0.14	6.94	12	8.01E-59	Calcium-binding protein that interacts with newly synthesized glycoproteins in the endoplasmic reticulum.
		0.14	6.98	15	6.3345E-66	
IPI00215948.4	Catenin alpha-1	0.13	7.89	20	2.5896E-97	Involved in cell-cell adhesion. Found at cell-cell boundaries and probably at cell-matrix boundaries.
		0.12	8.38	10	1.45E-63	
IPI00017292.1	Catenin beta-1	0.11	9.16	11	2.4315E-107	Involved in the regulation of cell adhesion and signal transduction.
		0.06	16.86	6	2.55E-70	
IPI00182540.5	Catenin delta-1	0.18	5.7	3	2.64E-06	Belongs to a multiprotein cell-cell adhesion complex and helps regulate cell-cell adhesion.
		0.16	6.19	8	8.3676E-21	
IPI00909515.1	cDNA FLJ51495	0.24	4.12	2	6.42E-26	Function unknown. Evidence at transcript level.
IPI00910164.1	cDNA FLJ53381	0.14	7.02	2	0.00027342	Function unknown. Evidence at transcript level.
IPI00910142.1	cDNA FLJ54183	0.28	3.59	2	7.8515E-10	

IPI00911039.1	cDNA FLJ54303	0.04	22.65	27	3.491E-233	
IPI00797612.1	cDNA FLJ56382	0.22	4.48	5	6.9793E-10	
		0.08	12.67	3	6.60E-06	
IPI00798127.1	cDNA FLJ75516	0.06	16.89	3	5.2114E-17	
		0.06	17.4	2	1.73E-08	
IPI00871856.2	cDNA FLJ76863	0.17	6.01	6	2.9531E-29	
IPI00025087.2	Cellular tumor antigen p53	0.04	26.26	11	2.6606E-26	Transcription factor that can induce growth arrest or apoptosis.
IPI00844508.1	Cingulin	0.17	5.91	5	1.97E-06	Plays a role in the formation and regulation of the tight junction (TJ) paracellular permeability barrier.
		0.11	8.73	24	1.5854E-131	
IPI00024067.4	Clathrin heavy chain 1	0.20	5.07	5	2.96E-31	Involved in endocytosis and post-Golgi vesicle-mediated transport.
		0.19	5.26	10	6.7783E-37	
IPI00015833.1	Coiled-coil-helix-coiled-coil-helix domain-containing protein 3, mitochondrial	0.12	8.01	3	1.8502E-11	Located in the mitochondria. A myristoylated protein that acts as a scaffold.
IPI00014230.1	Complement component 1 Q subcomponent-binding protein, mitochondrial	0.22	4.63	3	8.1298E-24	Precise function is unknown.
		0.18	5.57	2	2.16E-23	
IPI00867509.1	Coronin-1C	0.11	8.77	7	2.3358E-93	A 55KDa actin binding protein.
IPI00004839.1	Crk-like protein	0.79	1.26	7	2.6825E-29	May mediate the transduction of intracellular signals.
		0.67	1.49	10	3.02E-37	
IPI00290142.5	CTP synthase 1	0.55	1.82	10	7.3185E-62	Involved in pyrimidine metabolism.
		0.54	1.86	13	8.29E-40	

IPI00013847.4	Cytochrome b-c1 complex subunit 1, mitochondrial	0.18	5.62	3	3.1523E-11	Component of the ubiquinol-cytochrome c reductase complex. Part of the mitochondrial respiratory chain.
IPI00305383.1	Cytochrome b-c1 complex subunit 2, mitochondrial	0.13	7.72	4	4.77E-05	
		0.13	7.76	9	3.2128E-43	
IPI00303954.3	Cytochrome b5 type B	0.13	7.41	3	1.31E-06	Involved in respiration. A mitochondrial membrane bound hemoprotein.
IPI00017510.3	Cytochrome c oxidase subunit 2	0.12	8.41	5	1.06E-08	The component of the respiratory chain that catalyzes the reduction of oxygen to water.
		0.06	15.41	4	8.774E-09	
IPI00178072.7	Cytospin-A	0.17	6.05	4	7.6775E-11	Involved in the cell cycle and cell division.
IPI00011200.5	D-3-phosphoglycerate dehydrogenase	0.31	3.21	3	8.86E-09	Involved in amino-acid biosynthesis.
		0.29	3.43	2	0.000008134	
IPI00028931.2	Desmoglein-2	0.11	9.16	9	5.54E-26	A component of intercellular desmosome junctions, involved in cell-cell adhesion.
		0.07	14.45	12	8.3794E-44	
IPI00013933.2	Desmoplakin	0.18	5.49	25	1.0673E-101	Found in desmosomes (a type of junctional complex involved in cell-cell adhesion).
		0.12	8.07	15	1.51E-125	
IPI00296337.2	DNA-dependent protein kinase catalytic subunit	0.30	3.38	2	1.317E-12	Serine/threonine-protein kinase that acts as a molecular sensor for DNA damage.
IPI00012535.1	DnaJ homolog subfamily A member 1	0.09	11.19	4	5.22E-06	Co-chaperone of Hsc70.
		0.08	12.41	4	2.8045E-08	
IPI00032406.1	DnaJ homolog subfamily A member 2	0.19	5.34	2	9.48E-07	Co-chaperone of Hsc70.
IPI00297084.7	Dolichyl-diphosphooligosaccharide--	0.23	4.27	3	1.0906E-20	Located in the endoplasmic reticulum and is a

	protein glycosyltransferase 48 kDa subunit	0.21	4.83	3	1.50E-12	component of the oligosaccharyltransferase (OST) complex, which is involved in the N-linked glycosylation of nascent polypeptide chains.
IPI00025874.2	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	0.22	4.47	11	1.264E-43	
		0.18	5.63	14	2.15E-44	
IPI00396485.3	Elongation factor 1-alpha 1	0.77	1.3	9	4.8239E-44	Involved in protein biosynthesis.
		0.65	1.54	11	1.15E-10	
IPI00186290.6	Elongation factor 2	0.53	1.89	4	2.2533E-10	Involved in protein biogenesis.
IPI00027107.5	Elongation factor Tu, mitochondrial	0.32	3.16	5	1.1891E-13	Involved in protein biosynthesis.
		0.27	3.69	2	0.044981	
IPI00027230.3	Endoplasmin	0.23	4.33	3	5.409E-09	Molecular chaperone.
IPI00386604.1	Eukaryotic initiation factor 4A-II	0.36	2.8	2	1.0505E-14	Involved in translation initiation.
IPI00022744.5	Exportin-2	0.40	2.51	6	2.5175E-17	Involved in export of proteins from the nucleus to the cytoplasm.
IPI00005969.3	F-actin-capping protein subunit alpha-1	0.33	3.05	2	0.00061655	Binds in a Ca2+-independent manner to the fast growing ends of actin filaments.
		0.19	5.27	2	0.00083775	
IPI00026781.2	Fatty acid synthase	0.36	2.78	3	0.000010311	Catalyzes the formation of long-chain fatty acids.
		0.36	2.81	3	1.22E-20	
IPI00333541.6	Filamin-A	0.33	3.04	3	8.7719E-07	Links actin filaments to membrane glycoproteins.
		0.15	6.64	2	1.05E-07	
IPI00027438.2	Flotillin-1	0.15	6.69	3	6.09E-19	Peripheral membrane protein involved in clathrin independent endocytosis.
		0.10	10.47	4	4.8171E-24	
IPI00795979.1	Flotillin-2	0.15	6.53	6	1.0811E-23	Participates in the formation of caveolae or caveolae-like vesicles.
		0.05	20.32	2	2.76E-09	
IPI00465121.1	Galphai2 protein	0.17	5.86	4	6.87E-17	Located on the plasma membrane

		0.08	12.46	5	4.1954E-31	and functions as a G-protein family member.
CON_GFP	GFP	10.80	0.09	17	1.175E-152	
		3.19	0.31	14	5.52E-95	
IPI00788737.1	Glyceraldehyde-3-phosphate dehydrogenase	0.25	3.95	2	8.01E-05	A key enzyme in glycolysis.
IPI00221232.9	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12	0.15	6.47	2	0.00011933	Involved in intracellular signalling: Guanine nucleotide-binding proteins (G proteins) function as modulators or transducers in various transmembrane signaling systems.
IPI00003348.3	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2	0.06	16.99	5	2.82E-09	
IPI00220578.3	Guanine nucleotide-binding protein G(k) subunit alpha	0.10	10.37	5	4.7018E-23	
		0.08	13.08	6	2.97E-63	
IPI00305551.3	Guanine nucleotide-binding protein subunit alpha-11	0.23	4.43	2	0.0010169	Acts as an activator of phospholipase C. Belongs to the G-alpha family.
		0.18	5.62	2	0.0054347	
IPI00012451.3	Guanine nucleotide-binding protein subunit beta-4	0.08	12.46	3	2.0849E-10	G protein involved in the cellular response to glucagon stimulus.
IPI00010440.5	HCLS1-associated protein X-1	0.13	7.71	2	0.0017254	Promotes cell survival. Involved in the clathrin-mediated endocytosis pathway.
IPI00304925.5	Heat shock 70 kDa protein 1	0.03	37.85	31	2.0709E-253	Hsp70s stabilize preexisting proteins against aggregation and mediate the folding of newly translated polypeptides in the cytosol as well as within organelles.
		0.02	40.11	27	5.98E-236	
IPI00002966.2	Heat shock 70 kDa protein 4	0.15	6.61	2	1.45E-29	Protein chaperone.
		0.12	8.61	8	1.8125E-34	

IPI00003865.1	Heat shock cognate 71 kDa protein	0.04	24.72	24	9.46E-159	Protein chaperone. Localized in cytoplasmic mRNP granules containing untranslated mRNAs.
		0.04	27.11	24	5.6703E-207	
IPI00514983.4	Heat shock protein 105 kDa	0.21	4.72	4	5.53E-15	Belongs to the heat shock protein 70 family.
		0.08	12.42	3	6.2195E-10	
IPI00382470.3	Heat shock protein HSP 90-alpha	0.14	7.38	8	3.64E-20	Molecular chaperone.
		0.11	9.17	14	2.4805E-52	
IPI00414676.6	Heat shock protein HSP 90-beta	0.11	9.09	13	2.8509E-68	Molecular chaperone.
		0.09	10.88	12	2.10E-52	
IPI00479191.2	Heterogeneous nuclear ribonucleoprotein H	0.35	2.87	3	1.16E-24	A component of the heterogeneous nuclear ribonucleoprotein (hnRNP) complexes.
		0.33	3.04	3	1.2262E-36	
IPI00216746.1	Heterogeneous nuclear ribonucleoprotein K	0.31	3.18	2	0.000045576	One of the major pre-mRNA-binding proteins.
		0.27	3.68	4	1.90E-24	
IPI00644079.2	Heterogeneous nuclear ribonucleoprotein U (Scaffold attachment factor A)	0.22	4.61	2	0.0057521	Binds to pre-mRNA. Localized in cytoplasmic mRNP granules containing untranslated mRNAs.
IPI00874030.3	Heterogeneous nuclear ribonucleoproteins A2/B1	0.17	5.81	4	1.59E-28	Involved with pre-mRNA processing.
IPI00477313.3	Heterogeneous nuclear ribonucleoproteins C1/C2	0.23	4.28	3	2.08E-23	Binds pre-mRNA and nucleates the assembly of 40S hnRNP particles.
IPI00005711.4	Histone deacetylase 6	0.58	1.73	4	2.2145E-16	Deacetylates lysine residues on the N-terminal part of the core histones.
		0.57	1.77	2	0.00063255	
IPI00902514.1	Histone H2A	0.30	3.31	2	0.00052103	Core component of nucleosome.
		0.15	6.72	2	3.92E-07	

IPI00001639.2	Importin subunit beta-1	0.27	3.71	2	1.71E-20	Functions in nuclear protein import.
IPI00793443.1	Importin-5	0.46	2.16	2	5.48E-09	
		0.33	3.02	2	4.7107E-06	
IPI00020729.1	Insulin receptor substrate 4	0.19	5.4	2	5.1293E-06	Involved in intracellular signalling.
IPI00554711.3	Junction plakoglobin	0.13	7.59	4	2.5381E-14	Common junctional plaque protein. Interacts with calveolin.
IPI00299554.3	Kinesin-like protein KIF14	0.36	2.76	8	1.9519E-24	Plays an essential role in cytokinesis.
IPI00219217.3	L-lactate dehydrogenase B chain	0.39	2.58	2	0.00011347	Involved in glycolysis.
		0.31	3.2	2	0.0054615	
IPI00017592.1	LETM1 and EF-hand domain-containing protein 1, mitochondrial	0.17	5.89	2	1.4935E-16	Crucial for the maintenance of mitochondrial tubular networks.
IPI00783271.1	Leucine-rich PPR motif-containing protein, mitochondrial	0.30	3.37	4	1.1446E-10	May play a role in RNA metabolism in both nuclei and mitochondria.
IPI00216172.2	Lysosomal-associated membrane protein 2C	0.18	5.66	2	1.80E-09	Function of this LAMP2 isoform is unknown. LAMP2 may help protect lysosomal membranes from autodigestion.
IPI00293276.10	Macrophage migration inhibitory factor	0.78	1.29	2	9.0766E-16	Pro-inflammatory cytokine.
		0.61	1.64	2	3.30E-20	
IPI00639812.1	Microsomal glutathione S-transferase 3 variant	0.10	10.08	2	1.65E-10	Precise function unknown - ervidence only at the transcript level.
IPI00418497.1	Mitochondrial import inner membrane translocase subunit TIM50	0.17	5.99	4	3.89E-10	Essential component of the TIM23 complex, that mediates the translocation of proteins across the mitochondrial inner membrane.

IPI00024976.5	Mitochondrial import receptor subunit TOM22 homolog	0.08	12.17	3	3.69E-30	Central receptor component of the translocase of the outer membrane of mitochondria (TOM complex).
		0.08	12.38	2	6.0602E-26	
IPI00015602.1	Mitochondrial import receptor subunit TOM70	0.17	5.82	2	2.57E-07	Central receptor component of the translocase of the outer membrane of mitochondria (TOM complex).
		0.10	10.34	2	1.431E-07	
IPI00009960.6	Mitochondrial inner membrane protein	0.10	10.37	6	1.674E-19	Controls cristae morphology.
IPI00844172.1	Myosin	0.15	6.53	4	1.99E-66	Evidence only at transcript level. Function unknown.
		0.13	7.86	14	3.1085E-77	
IPI00335168.9	Myosin light polypeptide 6	0.07	13.39	4	2.28E-08	Regulatory light chain of myosin..
IPI00295457.6	Myosin phosphatase Rho-interacting protein	0.15	6.55	14	1.4312E-46	Targets myosin phosphatase to the actin cytoskeleton.
		0.11	8.83	4	4.52E-37	
IPI00604523.1	Myosin regulatory light chain MRCL3 variant	0.19	5.39	2	3.3658E-06	A variant of the regulatory light chain of myosin.
		0.18	5.44	3	4.57E-22	
IPI00790503.3	Myosin-10	0.11	9.47	61	0	Cellular myosin that appears to play a role in cytokinesis.
		0.08	11.96	43	0	
IPI00019502.3	Myosin-9	0.15	6.82	42	0	Cellular myosin that appears to play a role in cytokinesis.
		0.12	8.14	34	0	
IPI00376344.3	Myosin-Ib	0.08	12.34	3	8.0095E-09	Motor protein associated with the actin cytoskeleton.
IPI00743335.1	Myosin-Ic	0.27	3.69	2	0.00058806	Actin-based motor molecule with ATPase activity.
IPI00337541.3	NAD(P) transhydrogenase, mitochondrial	1.53	0.65	2	0.000096694	An oxidoreductase enzyme located in the mitochondrial inner membrane.
IPI00025796.3	NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial	0.22	4.45	3	1.06E-19	Core subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase:

IPI00010845.3	NADH dehydrogenase [ubiquinone] iron-sulfur protein 8, mitochondrial	0.38	2.66	2	0.00022751	Complex I, which transfers electrons from NADH to the respiratory chain.
IPI00217507.5	Neurofilament medium polypeptide	0.15	6.81	12	7.2389E-53	A component of the cytoskeleton.
		0.03	39.77	5	2.40E-20	
IPI00019472.4	Neutral amino acid transporter B(0)	0.26	3.9	3	3.0873E-08	Transports neutral amino acids across the plasma membrane.
IPI00465432.3	Nodal modulator 2	0.31	3.26	2	5.65E-05	May antagonize Nodal signalling.
IPI00783781.1	Nuclear pore complex protein Nup205	0.38	2.65	2	1.7656E-06	Involved in regulating the import and export of proteins into the nucleus.
IPI00549248.4	Nucleophosmin	0.22	4.53	4	1.78E-29	Involved in ribosome biogenesis, centrosome duplication, and protein chaperoning.
IPI00001654.5	Pericentriolar material 1 protein	0.19	5.19	11	5.8684E-33	Required for centrosome assembly and function.
		0.09	11.76	4	6.76E-18	
IPI00000874.1	Peroxiredoxin-1	0.59	1.7	8	9.3036E-15	Involved in redox regulation of the cell.
		0.52	1.92	7	2.02E-12	
IPI00014898.3	Plectin-1	0.11	9.5	33	2.27E-264	Interlinks intermediate filaments with microtubules.
		0.09	10.53	97	0	
IPI00008524.1	Polyadenylate-binding protein 1	0.21	4.7	5	6.6144E-21	Binds the poly(A) tail of mRNA. Helps initiate translation.
		0.16	6.41	4	1.26E-07	
IPI00384280.5	Prenylcysteine oxidase 1	0.24	4.14	3	1.29E-29	Involved in the degradation of prenylated proteins.
		0.18	5.56	3	1.2038E-10	
IPI00017617.1	Probable ATP-dependent RNA helicase DDX5	0.28	3.6	2	0.0046811	RNA-dependent ATPase activity.
IPI00017334.1	Prohibitin	0.09	10.82	8	1.07E-42	Inhibits DNA synthesis.
		0.08	11.96	9	2.8807E-72	

IPI00027252.6	Prohibitin-2	0.09	11.13	9	3.4011E-41	Acts as a mediator of transcriptional repression via recruitment of histone deacetylases.
		0.07	13.9	6	9.93E-62	
IPI00385250.1	Protease, serine, 3	0.03	37.01	2	1.3265E-07	A serine type endopeptidase.
IPI00299571.5	Protein disulfide-isomerase A6	0.65	1.55	5	3.9058E-31	May function as a chaperone that inhibits aggregation of misfolded proteins.
		0.54	1.84	4	9.12E-26	
IPI00298547.3	Protein DJ-1	0.54	1.85	2	0.4226	Protects cells against oxidative stress and cell death.
IPI00183002.6	Protein phosphatase 1 regulatory subunit 12A	0.22	4.52	6	1.3646E-19	Regulates myosin phosphatase activity.
IPI00794221.1	Putative uncharacterized protein DBN1	0.13	7.85	9	1.6611E-37	N/A
		0.04	23.74	2	3.23E-36	
IPI00418169.3	Putative uncharacterized protein DKFZp686P03159	0.25	4.07	2	2.2855E-07	N/A
		0.24	4.12	3	8.30E-05	
IPI00792352.1	Putative uncharacterized protein RAN	0.45	2.21	2	0.0057561	N/A
		0.28	3.56	2	0.000077029	
IPI00872387.1	Putative uncharacterized protein RPL7P23	0.30	3.37	2	7.87E-08	N/A
		0.21	4.84	2	0.000013341	
IPI00879810.1	Putative uncharacterized protein SPTAN1	0.11	9.28	44	6.4285E-210	N/A
		0.10	10.05	37	0	
IPI00016513.5	Ras-related protein Rab-10	0.18	5.45	2	9.5794E-12	May be involved in vesicular trafficking and neurotransmitter release.
		0.14	7.08	2	2.85E-13	
IPI00429190.4	Ras-related protein Rab-11A	0.18	5.42	2	4.47E-05	Modulates endosomal trafficking.
		0.17	5.94	5	1.3516E-08	

IPI00008964.3	Ras-related protein Rab-1B	0.18	5.6	3	1.835E-08	Regulates vesicular transport between the ER and successive Golgi compartments.
IPI00031169.1	Ras-related protein Rab-2A	0.31	3.2	2	0.0025402	Transports proteins from the ER to the Golgi complex.
IPI00016342.1	Ras-related protein Rab-7a	0.22	4.61	3	7.03E-14	Involved in late endocytic transport.
		0.19	5.29	3	2.902E-18	
IPI00019345.1	Ras-related protein Rap-1A	0.21	4.72	2	0.00014642	Lipid anchored in the cell membrane. Belongs to the small GTPase superfamily.
		0.17	5.83	2	6.7779E-21	
IPI00028635.4	Ribophorin II	0.26	3.82	2	4.7453E-15	Involved in N-linked glycosylation of nascent polypeptides.
		0.17	6.06	2	1.69E-08	
IPI00450975.1	RPS16 protein	0.20	4.93	2	5.98E-07	Evidence only at transcript level. Function unknown.
IPI00009104.7	RuvB-like 2	0.15	6.64	2	0.003193	Possesses single-stranded DNA-stimulated ATPase and ATP-dependent DNA helicase (5' to 3') activity.
IPI00219078.5	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	0.27	3.68	2	3.0435E-06	Catalyzes the hydrolysis of ATP coupled with the translocation of calcium from the cytosol to the sarcoplasmic reticulum lumen.
IPI00902512.1	Serine/threonine protein phosphatase	0.14	7.37	2	2.07E-09	Evidence at transcript level.
IPI00009368.4	Sideroflexin-1	0.08	11.97	2	4.48E-07	An ion transporter located in the mitochondrial membrane.
		0.06	16.31	2	2.4121E-06	
IPI00295098.3	Signal recognition particle receptor subunit beta	0.26	3.91	2	1.91E-16	Helps ensures the correct targeting of nascent secretory proteins to the ER membrane system.
		0.22	4.59	3	6.269E-24	

IPI00006482.1	Sodium/potassium-transporting ATPase subunit alpha-1	0.10	9.55	20	4.5674E-157	Enzyme involved in the exchange of Na+ and K+ ions across the plasma membrane.
		0.09	10.63	11	7.90E-58	
IPI00008167.1	Sodium/potassium-transporting ATPase subunit beta-3	0.18	5.7	3	1.4291E-11	Involved in the exchange of Na+ and K+ ions across the plasma membrane.
IPI00871932.1	Spectrin beta chain, brain 1	0.13	7.82	39	3.4409E-242	May be involved in the movement of the cytoskeleton at the membrane.
		0.12	8.14	25	1.45E-130	
IPI00479946.3	STIP1 protein	0.10	9.69	6	1.72E-49	Protein function unknown. Evidence at transcript level.
IPI00334190.4	Stomatin-like protein 2	0.10	10.15	3	7.33E-13	Peripheral plasma membrane protein associated with the cytoskeleton.
IPI00007765.5	Stress-70 protein, mitochondrial	0.15	6.84	12	2.24E-72	Protein chaperone of the HSP70 family.
		0.13	7.75	21	2.6842E-76	
IPI00290566.1	T-complex protein 1 subunit alpha	0.21	4.85	2	2.0364E-06	The T-complex is located in the cytoplasm - it is associated with the cytoskeleton, especially at centrosomes. The T-complex functions as a molecular chaperone; it assists the folding of proteins upon ATP hydrolysis and plays a role in 'de novo' post-translational protein folding.
		0.17	5.81	4	1.00E-18	
IPI00297779.7	T-complex protein 1 subunit beta	0.24	4.24	5	3.69E-20	
IPI00010720.1	T-complex protein 1 subunit epsilon	0.21	4.75	3	2.1314E-08	
IPI00018465.1	T-complex protein 1 subunit eta	0.13	7.69	4	1.30E-05	
IPI00553185.2	T-complex protein 1 subunit gamma	0.18	5.65	3	2.2641E-11	
		0.13	7.66	4	2.02E-28	
IPI00784090.2	T-complex protein 1 subunit theta	0.14	7.19	8	6.45E-24	
		0.13	7.7	2	5.2692E-09	
IPI00027626.3	T-complex protein 1 subunit zeta	0.18	5.54	3	1.3074E-25	
		0.16	6.13	5	3.81E-18	
IPI00873222.2	T-complex protein 1, delta subunit	0.15	6.78	3	1.99E-25	
		0.12	8.05	3	8.3691E-08	

IPI00216298.6	Thioredoxin	0.78	1.28	3	3.2122E-18	Involved in various redox reactions through the reversible oxidation of its active center dithiol to a disulfide.
		0.53	1.88	3	2.51E-12	
IPI00395887.4	Thioredoxin-related transmembrane protein 1	0.18	5.6	2	0.013369	Involved in electron transport, and ER to Golgi vesicle-mediated transport.
IPI00216219.3	Tight junction protein ZO-1	0.17	5.8	15	5.82E-70	Peripheral plasma membrane protein involved in the formation of tight junctions between cells.
		0.07	14.68	38	6.9093E-188	
IPI00019385.3	Translocon-associated protein subunit delta	0.24	4.13	2	0.000081258	Located in the ER and involved in intracellular protein transport.
		0.15	6.74	2	1.74E-24	
IPI00028055.4	Transmembrane emp24 domain-containing protein 10	0.32	3.12	2	1.5648E-09	Involved in vesicular protein trafficking and in golgi-ER transport.
IPI00031522.2	Trifunctional enzyme subunit alpha, mitochondrial	0.17	5.78	6	4.57E-27	Involved in fatty acid metabolism.
		0.14	7.19	9	3.1728E-37	
IPI00022793.5	Trifunctional enzyme subunit beta, mitochondrial	0.34	2.92	3	3.5042E-13	Involved in fatty acid metabolism
IPI00218319.3	Tropomyosin alpha-3 chain	0.11	9.2	7	1.373E-34	Binds to actin filaments in muscle and non-muscle cells.
		0.11	9.43	4	7.25E-07	
IPI00180675.4	Tubulin alpha-1A chain	0.12	8.63	16	1.46E-89	The major constituent of microtubules.
IPI00387144.4	Tubulin alpha 6 variant	0.12	8.5	16	2.12E-69	
IPI00011654.2	Tubulin beta chain	0.20	5.12	18	5.5184E-133	Tubulin is the major constituent of microtubules.
		0.18	5.51	16	5.88E-96	
IPI00007752.1	Tubulin beta-2C chain	0.29	3.4	14	2.02E-69	
		0.26	3.84	15	8.2464E-97	
IPI00418471.6	Vimentin	0.04	25.44	33	9.9185E-201	Forms intermediate

		0.03	28.83	31	2.20E-225	filaments. Structural constituent of the cyoskeleton.
IPI00216308.5	Voltage-dependent anion-selective channel protein 1	0.17	5.95	6	9.0807E-34	Forms a channel through the mitochondrial outer membrane. May participate in the formation of the permeability transition pore complex (PTPC).
		0.07	13.84	7	2.91E-15	
IPI00024145.1	Voltage-dependent anion-selective channel protein 2	0.06	15.67	4	2.26E-13	A porin found in the mitochondrial outer membrane, that allows diffusion of small hydrophilic molecules.
IPI00216026.2	Voltage-dependent anion-selective channel protein 2	0.06	16.03	7	1.1548E-19	
newprotein999	N-protein CCHFV	0.05	19.79	26	0	

APPENDIX III MS ANALYSIS OF THE FRACTION CONTAINING INFECTIOUS HAZV

Proteins identified and quantified in the MS analysis of the fraction containing infectious HAZV.

Accession number	Protein name	MW Da	Peptides	emPAI	Function
P62258	14-3-3 protein epsilon	29437	2	0.24	Adaptor protein involved in intracellular signalling.
P27348	14-3-3 protein theta	28143	2	0.25	
P63104	14-3-3 protein zeta/delta	28010	2	0.25	
P62333	26S protease regulatory subunit 10B	44541	2	0.15	The 26S protease is involved in the ATP-dependent degradation of ubiquitinated proteins. The regulatory (or ATPase) complex confers ATP dependency and substrate specificity to the 26S complex.
P62191	26S protease regulatory subunit 4	49436	5	0.38	
P17980	26S protease regulatory subunit 6A	49569	3	0.21	
P62195	26S protease regulatory subunit 8	45879	4	0.32	
O00231	26S proteasome non-ATPase regulatory subunit 11	47830	2	0.14	Component of the lid subcomplex of the 26S proteasome.
O00232	26S proteasome non-ATPase regulatory subunit 12	53381	2	0.13	Acts as a regulatory subunit of the 26S proteasome which is involved in the ATP-dependent degradation of ubiquitinated proteins
Q9UNM6	26S proteasome non-ATPase regulatory subunit 13	43314	2	0.16	
Q13200	26S proteasome non-ATPase regulatory subunit 2	100988	6	0.21	
P51665	26S proteasome non-ATPase regulatory subunit 7	37171	3	0.29	
P62249	40S ribosomal protein S16	16660	2	0.45	Involved in protein biogenesis. Component of the small ribosomal subunit. Mature ribosomes consist of a small (40S) and a large (60S) subunit. The 40S subunit contains about 33 different
P62269	40S ribosomal protein S18	17819	2	0.41	
P15880	40S ribosomal protein S2	31701	5	0.64	
P23396	40S ribosomal protein S3	26953	5	0.79	
P61247	40S ribosomal protein	30265	5	0.68	

	S3a				proteins and 1 molecule of RNA (18S). The 60S subunit contains about 49 different proteins and 3 molecules of RNA (28S, 5.8S and 5S).
P62701	40S ribosomal protein S4, X isoform	29918	4	0.52	
P46782	40S ribosomal protein S5	23144	4	0.71	
P62241	40S ribosomal protein S8	24586	3	0.47	
P46781	40S ribosomal protein S9	22746	2	0.32	
P08865	40S ribosomal protein SA	33058	4	0.47	
P05388	60S acidic ribosomal protein P0	34534	6	0.73	Involved in protein biogenesis. Component of the large ribosomal subunit. The 60S subunit contains about 49 different proteins and 3 molecules of RNA (28S, 5.8S and 5S).
P62906	60S ribosomal protein L10a	25098	2	0.28	
P46776	60S ribosomal protein L27a	16776	2	0.44	
P36578	60S ribosomal protein L4	48064	4	0.3	
P46777	60S ribosomal protein L5	34680	5	0.58	
P62424	60S ribosomal protein L7a	30259	3	0.37	
P11021	78 kDa glucose-regulated protein	72513	2	0.09	Protein chaperone. Facilitates the assembly of multimeric protein complexes inside the ER.
P60709	Actin, cytoplasmic 1	42163	22	3.52	Major component of the actin cytoskeleton
O15145	Actin-related protein 2/3 complex subunit 3	20872	2	0.35	Regulates actin polymerization
P61158	Actin-related protein 3	47908	3	0.22	Regulates actin polymerization
P07741	Adenine phosphoribosyltransferase	19878	2	0.37	Involved in purine metabolism
P23526	Adenosylhomocysteinase	48366	8	0.69	Involved in amino acid biosynthesis
Q01518	Adenylyl cyclase-associated protein 1	52436	4	0.28	Peripheral membrane protein that directly regulates actin filament dynamics
P18085	ADP-ribosylation factor 4	20723	2	0.35	Involved in protein trafficking; may

					modulate vesicle budding and uncoating within the Golgi apparatus
O00468	Agrin	225357	3	0.04	Glycoprotein that plays a central role in the formation and the maintenance of the neuromuscular junction (NMJ)
P01023	Alpha-2-macroglobulin	164724	5	0.1	Protease inhibitor
P61163	Alpha-centractin	42812	3	0.25	Involved in microtubule based vesicle motility.
P06733	Alpha-enolase	47592	13	1.39	Enzyme involved in glycolysis
P02771	Alpha-fetoprotein	70569	4	0.2	Secreted protein. Binds copper and nickel
P04083	Annexin A1	39029	14	2.12	Promotes membrane fusion and is involved in exocytosis.
P07355	Annexin A2	38919	6	0.63	May be involved with exocytosis.
Q10567	AP-1 complex subunit beta-1	105593	3	0.1	Subunit of clathrin-associated AP 1 complex that plays a role in protein sorting in the late-Golgi/trans-Golgi network (TGN) and/or endosomes.
P63010	AP-2 complex subunit beta	105509	4	0.13	The adaptor protein complex 2 (AP-2) is involved in clathrin-dependent endocytosis
Q96CW1	AP-2 complex subunit mu	50076	3	0.21	
P08243	Asparagine synthetase [glutamine-hydrolyzing]	65010	3	0.16	Involved in amino acid biosynthesis
P14868	Aspartate--tRNA ligase, cytoplasmic	57610	4	0.25	Involved in protein biosynthesis
P06576	ATP synthase subunit beta, mitochondrial	56636	3	0.18	Mitochondrial protein involved in respiration
P53396	ATP-citrate synthase	121785	7	0.2	Involved in de novo lipid synthesis
Q08211	ATP-dependent RNA helicase A	142292	6	0.15	Unwinds double-stranded DNA and RNA in a 3' to 5'

					direction.
O00571	ATP-dependent RNA helicase DDX3X	73708	4	0.19	Multifunctional ATP-dependent RNA helicase. Involved in IRF-3 signalling
Q562R1	Beta-actin-like protein 2	42429	8	0.69	Associated with the actin cytoskeleton
P13929	Beta-enolase	47410	5	0.4	Involved in glycolysis
P16278	Beta-galactosidase	76594	2	0.09	Cleaves beta-linked terminal galactosyl residues from gangliosides, glycoproteins, and glycosaminoglycans.
P07814	Bifunctional glutamate/proline--tRNA ligase	172191	5	0.1	Involved in protein biosynthesis
P31939	Bifunctional purine biosynthesis protein PURH	65200	7	0.41	Involved in purine biosynthesis
P11586	C-1-tetrahydrofolate synthase, cytoplasmic	102291	12	0.46	Involved in amino acid biosynthesis
P27708	CAD protein	245278	5	0.07	Involved in pyrimidine biosynthesis
Q00610	Clathrin heavy chain 1	193371	54	1.47	The major protein of the polyhedral coat of coated pits and vesicles involved in endocytosis and protein sorting in the Golgi network.
P53675	Clathrin heavy chain 2	189131	16	0.31	
P12259	Coagulation factor V	252797	4	0.05	Central regulator of hemostasis
P53621	Coatomer subunit alpha	139908	5	0.12	The coatomer is a cytosolic protein complex that helps mediate protein transport from the ER, via the Golgi up to the trans Golgi network. Coatomer complex is required for budding from Golgi membranes, and is essential for the retrograde Golgi-to-ER transport of dilysine-tagged proteins.
P53618	Coatomer subunit beta	108325	3	0.09	
P48444	Coatomer subunit delta	57741	2	0.12	
Q9Y678	Coatomer subunit gamma-1	99078	2	0.07	

P12109	Collagen alpha-1(VI) chain	109713	2	0.06	Secreted protein involved in cell adhesion.
P12107	Collagen alpha-1(XI) chain	181692	23	0.45	
P05997	Collagen alpha-2(V) chain	145901	6	0.14	
P01024	Complement C3	188680	13	0.21	Secreted protein involved in the complement pathway
P0C0L4	Complement C4-A	194372	8	0.14	
P29279	Connective tissue growth factor	40400	13	1.57	Secreted protein involved in cell adhesion.
Q14204	Cytoplasmic dynein 1 heavy chain 1	534920	16	0.1	Acts as a motor for the intracellular retrograde motility of vesicles and organelles along microtubules.
Q7L576	Cytoplasmic FMR1-interacting protein 1	146853	18	0.49	Involved in translation
Q96F07	Cytoplasmic FMR1-interacting protein 2	150409	10	0.24	Involved in the p53/TP53-dependent induction of apoptosis.
P15924	Desmoplakin	334132	3	0.03	Found in desmosomes.
P78527	DNA-dependent protein kinase catalytic subunit	473860	5	0.03	Molecular sensor for DNA damage
Q9UBS4	DnaJ homolog subfamily B member 11	40885	10	1.18	Co-chaperone for HSPA5
Q5T4S7	E3 ubiquitin-protein ligase UBR4	580658	2	0.01	Involved in the ubiqutin conjugation pathway
Q05639	Elongation factor 1-alpha 2	50891	9	0.76	Involved in protein biosynthesis
P24534	Elongation factor 1-beta	25030	2	0.29	
P29692	Elongation factor 1-delta	31328	3	0.35	
P26641	Elongation factor 1-gamma	50540	6	0.46	
P13639	Elongation factor 2	96357	34	1.91	
P14625	Endoplasmin	92807	3	0.11	Protein chaperone
P60842	Eukaryotic initiation factor 4A-I	46464	4	0.32	Involved in protein biosynthesis
P05198	Eukaryotic translation initiation factor 2 subunit 1	36485	6	0.68	

P41091	Eukaryotic translation initiation factor 2 subunit 3	51758	6	0.45	
Q14152	Eukaryotic translation initiation factor 3 subunit A	166978	11	0.24	
P55884	Eukaryotic translation initiation factor 3 subunit B	92934	5	0.19	
Q99613	Eukaryotic translation initiation factor 3 subunit C	106073	3	0.1	
O00303	Eukaryotic translation initiation factor 3 subunit F	37765	4	0.4	
Q13347	Eukaryotic translation initiation factor 3 subunit I	36989	3	0.29	
Q9Y262	Eukaryotic translation initiation factor 3 subunit L	67023	8	0.47	
P55010	Eukaryotic translation initiation factor 5	49759	2	0.14	
P15311	Ezrin	69595	2	0.1	Probably involved in connections of major cytoskeletal structures to the plasma membrane.
Q08945	FACT complex subunit SSRP1	81478	2	0.08	Involved in DNA and mRNA synthesis
P49327	Fatty acid synthase	275988	33	0.47	Involved in fatty acid biosynthesis
P02751	Fibronectin	266163	10	0.13	Involved in cell adhesion, cell motility, opsonization, and maintenance of cell shape.
P21333	Filamin-A	283412	24	0.32	Links actin filaments to membrane glycoproteins.
Q14315	Filamin-C	293518	2	0.02	Actin cytoskeletal reorganisation
P09104	Gamma-enolase	47692	4	0.31	Involved in glycolysis
P04406	Glyceraldehyde-3-phosphate dehydrogenase	36312	13	2.11	A key enzyme in glycolysis.

A6XIP3	Glycoprotein	161105	40	1.23	HAZV glycoprotein polypeptide
P62826	GTP-binding nuclear protein Ran	24690	4	0.66	GTP-binding protein involved in nucleocytoplasmic transport
P62873	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	38262	2	0.18	A modulator or transducer in various transmembrane signaling systems.
P63244	Guanine nucleotide-binding protein subunit beta-2-like 1	35622	9	1.23	Involved in the recruitment, assembly and/or regulation of a variety of signaling molecules.
P34931	Heat shock 70 kDa protein 1-like	70841	3	0.15	Hsp70s stabilize preexistent proteins against aggregation and mediate the folding of newly translated polypeptides in the cytosol as well as within organelles.
P17066	Heat shock 70 kDa protein 6	71551	4	0.2	
P11142	Heat shock cognate 71 kDa protein	71193	15	0.97	
P07900	Heat shock protein HSP 90-alpha	85117	11	0.46	Protein chaperone.
P08238	Heat shock protein HSP 90-beta	83665	21	1.16	Protein chaperone.
O60812	Heterogeneous nuclear ribonucleoprotein C-like 1	32291	2	0.22	May play a role in nucleosome assembly
P31943	Heterogeneous nuclear ribonucleoprotein H	49595	2	0.14	A component of the heterogeneous nuclear ribonucleoprotein (hnRNP) complexes which provide the substrate for the processing events that pre-mRNAs undergo before becoming functional, translatable mRNAs in the cytoplasm. Mediates pre-mRNA alternative splicing regulation.
P61978	Heterogeneous nuclear ribonucleoprotein K	51341	5	0.36	
O60506	Heterogeneous nuclear ribonucleoprotein Q	69899	3	0.15	
O43390	Heterogeneous nuclear ribonucleoprotein R	71295	2	0.09	
Q00839	Heterogeneous nuclear ribonucleoprotein U	91380	2	0.07	
P07910	Heterogeneous nuclear ribonucleoproteins C1/C2	33818	3	0.32	
O14929	Histone acetyltransferase type B catalytic subunit	49991	2	0.14	Acetylates soluble but not nucleosomal histone H4

P0C0S8	Histone H2A type 1	14194	3	0.91	Core component of nucleosome. Nucleosomes wrap and compact DNA into chromatin.
P06899	Histone H2B type 1-J	14007	4	1.39	
Q5QNW6	Histone H2B type 2-F	14023	3	0.92	
Q71DI3	Histone H3.2	15547	5	1.69	
P62805	Histone H4	11471	9	9.69	
Q86YZ3	Hornerin	283251	2	0.02	Involved in keratinization.
Q9Y4L1	Hypoxia up-regulated protein 1	111605	2	0.06	Protein chaperone.
O00505	Importin subunit alpha-3	58399	2	0.12	Involved in nuclear import of proteins.
Q14974	Importin subunit beta-1	98531	3	0.1	
P12268	Inosine-5~-monophosphate dehydrogenase 2	56337	4	0.25	Involved in purine biosynthesis
Q9Y6M1	Insulin-like growth factor 2 mRNA-binding protein 2	66306	2	0.1	Involved in mRNA transport.
P19827	Inter-alpha-trypsin inhibitor heavy chain H1	101893	2	0.07	Secreted protein that may act as a carrier of hyaluronan in serum. Protease inhibitor.
P19823	Inter-alpha-trypsin inhibitor heavy chain H2	106964	4	0.13	
Q06033	Inter-alpha-trypsin inhibitor heavy chain H3	100183	3	0.1	
Q12905	Interleukin enhancer-binding factor 2	43374	3	0.25	Involved in transcription regulation.
P41252	Isoleucine--tRNA ligase, cytoplasmic	145829	3	0.07	Involved in protein biosynthesis
P13645	Keratin, type I cytoskeletal 10	59131	38	6.82	Components of the cytoskeleton. Keratins form intermediate filaments, which are composed of a heterotetramer of two type I and two type II keratins. There are two types of cytoskeletal and microfibrillar keratin: I (acidic; 40-55 kDa) and II (neutral to basic; 56-70 kDa). Different species of keratin have various different roles including; involvement in signalling various
P13646	Keratin, type I cytoskeletal 13	50011	5	0.38	
P02533	Keratin, type I cytoskeletal 14	51983	25	3.65	
P08779	Keratin, type I cytoskeletal 16	51689	31	5.78	
Q04695	Keratin, type I cytoskeletal 17	48472	17	2.06	
P08727	Keratin, type I cytoskeletal 19	44190	6	0.54	
P35900	Keratin, type I cytoskeletal 20	48625	2	0.14	
Q7Z3Z0	Keratin, type I cytoskeletal 25	49969	3	0.21	
Q7Z3Y8	Keratin, type I	50530	3	0.21	

	cytoskeletal 27				pathways, involved in the organization of myofibers. Together with KRT8, these keratins help to link the contractile apparatus to dystrophin at the costameres of striated muscle. Some species play a significant role in maintaining keratin filament organization in intestinal epithelia. When phosphorylated, these keratins play a role in the secretion of mucin in the small intestine.
P04264	Keratin, type II cytoskeletal 1	66281	49	8.26	
Q7Z794	Keratin, type II cytoskeletal 1b	62260	4	0.23	
Q01546	Keratin, type II cytoskeletal 2 oral	66481	6	0.34	
P12035	Keratin, type II cytoskeletal 3	64660	9	0.56	
P19013	Keratin, type II cytoskeletal 4	57760	3	0.18	
P13647	Keratin, type II cytoskeletal 5	62679	15	1.04	
P02538	Keratin, type II cytoskeletal 6A	60404	24	2.39	
P04259	Keratin, type II cytoskeletal 6B	60426	23	2.21	
P48668	Keratin, type II cytoskeletal 6C	60384	24	2.39	
P08729	Keratin, type II cytoskeletal 7	51522	4	0.28	
Q3SY84	Keratin, type II cytoskeletal 71	57880	4	0.25	
Q14CN4	Keratin, type II cytoskeletal 72	56581	6	0.4	
Q7RTS7	Keratin, type II cytoskeletal 74	58340	4	0.25	
O95678	Keratin, type II cytoskeletal 75	59920	8	0.45	
Q5XKE5	Keratin, type II cytoskeletal 79	58196	7	0.39	
P05787	Keratin, type II cytoskeletal 8	53782	4	0.27	
Q6KB66	Keratin, type II cytoskeletal 80	51118	2	0.13	
O60282	Kinesin heavy chain isoform 5C	110108	2	0.06	Microtubule-dependent motor involved in organelle transport.
P33176	Kinesin-1 heavy chain	110469	3	0.09	
P00338	L-lactate dehydrogenase A chain	37061	2	0.19	Involved in glycolysis
P07195	L-lactate dehydrogenase B chain	37011	2	0.19	Involved in glycolysis

P26038	Moesin	68003	2	0.1	Probably involved in connections of major cytoskeletal structures to the plasma membrane.
P22234	Multifunctional protein ADE2	47901	4	0.3	Involved in purine biosynthesis
P35580	Myosin-10	229938	2	0.03	Cellular myosin that appears to play a role in cytokinesis, cell shape and other functions such as secretion and capping
Q7Z406	Myosin-14	228812	2	0.03	
P35579	Myosin-9	227757	19	0.31	
Q9Y2A7	Nck-associated protein 1	130129	8	0.22	Regulates actin reorganisation.
Q14697	Neutral alpha-glucosidase AB	107374	3	0.09	Involved in glycan metabolism in ER and golgi
M4PWE6	Nucleocapsid protein	54639	56	16.61	HAZV N
Q9NR30	Nucleolar RNA helicase 2	87915	2	0.08	Can unwind double-stranded RNA
P19338	Nucleolin	76736	6	0.29	The major nucleolar protein
P06748	Nucleophosmin	32837	7	0.96	Involved in ribosome biogenesis, centrosome duplication, and protein chaperoning.
P55209	Nucleosome assembly protein 1-like 1	45742	2	0.15	May be involved in modulating chromatin formation
P32119	Peroxiredoxin-2	22160	2	0.32	Involved in redox regulation of the cell.
Q9Y285	Phenylalanine--tRNA ligase alpha subunit	57696	3	0.18	Involved in protein biosynthesis
Q9NSD9	Phenylalanine--tRNA ligase beta subunit	66812	6	0.33	Involved in protein biosynthesis
P36955	Pigment epithelium-derived factor	46565	5	0.41	Secreted protease inhibitor
P00747	Plasminogen	93358	4	0.15	Involved in blood coagulation
P43034	Platelet-activating factor acetylhydrolase IB subunit alpha	47289	2	0.14	Involved in dynein mediated microtubule-based transport

P09874	Poly [ADP-ribose] polymerase 1	113922	3	0.09	Involved in DNA repair.
Q15366	Poly(rC)-binding protein 2	39066	2	0.18	Negatively regulates cellular antiviral responses mediated by MAVS signaling.
P11940	Polyadenylate-binding protein 1	70965	2	0.09	Involved in mRNA processing.
P0CG47	Polyubiquitin-B	25914	3	0.44	Involved in protein signalling and degradation
Q6S8J3	POTE ankyrin domain family member E	122993	11	0.3	Cytoplasmic. Co-localises with actin filaments
P0CG39	POTE ankyrin domain family member J	118851	4	0.11	
P20742	Pregnancy zone protein	165353	3	0.06	Secreted protease inhibitor
P02545	Prelamin-A/C	74491	3	0.14	Component of the nuclear lamina
Q9UMS4	Pre-mRNA-processing factor 19	55714	4	0.26	Plays a role in DNA double-strand break (DSB) repair
Q92841	Probable ATP-dependent RNA helicase DDX17	81017	3	0.13	Involved in transcriptional regulation
Q93008	Probable ubiquitin carboxyl-terminal hydrolase FAF-X	295626	4	0.04	De-ubiquitinates target proteins
Q02809	Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1	84179	5	0.21	Forms hydroxylysine residues in collagens
Q9UQ80	Proliferation-associated protein 2G4	44212	2	0.16	Involved in transcription regulation
Q5VYK3	Proteasome-associated protein ECM29 homolog	206096	2	0.03	May couple the proteasome to different compartments
Q99873	Protein arginine N-methyltransferase 1	42170	5	0.46	Methylates amino acid residues of specific proteins
O60678	Protein arginine N-methyltransferase 3	60633	2	0.11	
O14744	Protein arginine N-methyltransferase 5	73433	2	0.09	
P00734	Prothrombin	71586	3	0.14	Involved in haemostasis

Q5VTE0	Putative elongation factor 1-alpha-like 3	50606	14	1.42	N/A
Q58FF6	Putative heat shock protein HSP 90-beta 4	58966	2	0.11	N/A
Q58FF7	Putative heat shock protein HSP 90-beta-3	68735	3	0.15	N/A
O43143	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15	91784	2	0.07	N/A
Q9H853	Putative tubulin-like protein alpha-4B	27930	3	0.4	N/A
P14618	Pyruvate kinase isozymes M1/M2	58581	22	2.15	Involved in glycolysis
P46940	Ras GTPase-activating-like protein IQGAP1	189872	6	0.11	Could be involved in signal transduction
P61026	Ras-related protein Rab-10	22866	3	0.51	The small GTPases Rab are key regulators of intracellular membrane trafficking, from the formation of transport vesicles to their fusion with membranes. Some regulate transport of vesicles from the ER through successive Golgi compartments.
P59190	Ras-related protein Rab-15	24771	2	0.29	
Q9H0U4	Ras-related protein Rab-1B	22439	3	0.52	
Q15286	Ras-related protein Rab-35	23407	2	0.31	
P00352	Retinal dehydrogenase 1	55565	12	0.99	Oxidoreductase enzyme.
Q13464	Rho-associated protein kinase 1	159213	2	0.04	A key regulator of the actin cytoskeleton and cell polarity
Q9Y265	RuvB-like 1	50649	5	0.37	Involved in transcription regulation
Q14563	Semaphorin-3A	90027	5	0.2	Involved in cell differentiation.
O95025	Semaphorin-3D	90846	19	0.89	
Q9NVA2	Septin-11	49763	2	0.14	Involved in actin cytoskeleton organisation
Q15019	Septin-2	41800	6	0.58	
Q07955	Serine/arginine-rich splicing factor 1	27953	3	0.4	Involved in mRNA processing
P30153	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A	66176	2	0.1	Involved in chromosome segregation

	alpha isoform				
P62714	Serine/threonine-protein phosphatase 2A catalytic subunit beta isoform	36234	3	0.3	Modulates the activity of phosphorylase B kinase casein kinase 2, mitogen-stimulated S6 kinase, and MAP-2 kinase
P62136	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit	38340	4	0.39	Protein phosphatase that associates with over 200 regulatory proteins to form highly specific holoenzymes which dephosphorylate hundreds of biological targets.
Q9Y3F4	Serine-threonine kinase receptor-associated protein	38867	2	0.18	Involved in mRNA processing
P49591	Serine--tRNA ligase, cytoplasmic	59364	3	0.18	Involved in protein biosynthesis
P10768	S-formylglutathione hydrolase	32067	2	0.22	Involved in the detoxification of formaldehyde.
Q99519	Sialidase-1	46063	2	0.15	Removes of sialic acid moities from glycoproteins and glycolipids
P63162	Small nuclear ribonucleoprotein-associated protein N	24880	3	0.46	Involved in RNA processing.
Q13813	Spectrin alpha chain, non-erythrocytic 1	285274	4	0.05	Involved in secretion
O75533	Splicing factor 3B subunit 1	146590	2	0.05	Involved in mRNA processing
Q15393	Splicing factor 3B subunit 3	136686	9	0.24	
P23246	Splicing factor, proline- and glutamine-rich	76327	5	0.23	
Q7KZF4	Staphylococcal nuclease domain-containing protein 1	102729	3	0.1	Involved in transcription regulation
P31948	Stress-induced-phosphoprotein 1	63338	2	0.11	Mediates the association of the molecular chaperones HSC70 and HSP90

Q9Y490	Talin-1	271877	15	0.2	Involved in connections of major cytoskeletal structures to the plasma membrane.
Q9Y4G6	Talin-2	273892	4	0.05	
P17987	T-complex protein 1 subunit alpha	60930	25	2.72	Molecular chaperone; assists the folding of proteins upon ATP hydrolysis. As part of the BBS/CCT complex may play a role in the assembly of BBSome, a complex involved in ciliogenesis regulating transport vesicles to the cilia.
P78371	T-complex protein 1 subunit beta	57905	26	3.2	
P50991	T-complex protein 1 subunit delta	58512	17	1.54	
P48643	T-complex protein 1 subunit epsilon	60200	12	0.89	
Q99832	T-complex protein 1 subunit eta	59953	21	1.91	
P49368	T-complex protein 1 subunit gamma	61177	19	1.7	
P50990	T-complex protein 1 subunit theta	60264	24	2.58	
P40227	T-complex protein 1 subunit zeta	58555	19	1.82	
Q92526	T-complex protein 1 subunit zeta-2	58466	6	0.39	
P07996	Thrombospondin-1	133402	14	0.4	Adhesive glycoprotein that mediates cell-to-cell and cell-to-matrix interactions
P10646	Tissue factor pathway inhibitor	36186	2	0.19	Involved in haemostasis
P02786	Transferrin receptor protein 1	85385	2	0.08	Involved in cellular uptake of iron
P55072	Transitional endoplasmic reticulum ATPase	90061	2	0.07	Involved in the transfer of membranes from the ER to the Golgi apparatus
Q92616	Translational activator GCN1	295078	2	0.02	Involved in translation regulation
Q92973	Transportin-1	103882	2	0.06	Involved in nuclear protein import
Q71U36	Tubulin alpha-1A chain	50899	20	2.5	The major constituent of microtubules
P68363	Tubulin alpha-1B chain	50915	22	2.97	
P68366	Tubulin alpha-4A chain	50745	18	2.11	
P07437	Tubulin beta chain	50206	25	3.61	

Q9H4B7	Tubulin beta-1 chain	50976	6	0.37	
Q13509	Tubulin beta-3 chain	50967	16	1.56	
P68371	Tubulin beta-4B chain	50366	20	2.33	
O75643	U5 small nuclear ribonucleoprotein 200 kDa helicase	246117	8	0.11	Involved in mRNA processing
P22314	Ubiquitin-like modifier-activating enzyme 1	118969	13	0.42	Involved in ubiquitin conjugation pathway
Q96QK1	Vacuolar protein sorting-associated protein 35	92558	3	0.11	Component of the retromer complex
P13010	X-ray repair cross-complementing protein 5	83333	7	0.31	Involved in DNA repair
P12956	X-ray repair cross-complementing protein 6	70195	7	0.38	