Unravelling the replication biology of Providence virus in a cell culture-based model system

A thesis submitted in the fulfilment of the requirement for the degree of

Master of Science

(Microbiology)

At

Rhodes University

By

Rachel Anne Jarvie

ORCiD: 0000-0001-8913-9452

December 2019

Abstract

There has been an increase in the number of viral outbreaks in the last decade; the majority of these are attributed to insect-human or animal-human transfer. Despite this awareness, there is limited understanding of the replication biology of the viruses causing the outbreaks and there are few model systems that are available to study RNA virus replication and viral persistence.

In this study, we describe a Providence (PrV)-based model system to study virus replication biology. PrV is a single-stranded RNA virus that can cross Kingdom boundaries; it is capable of establishing a productive infection in insect and mammalian cell culture and it is also capable of replicating in plants. Only one other virus has been reported to infect a similar host range - the Nodavirus, Flock House virus (FHV).

First, we performed a bioinformatic analysis of the PrV genome and validated the tools that were currently available to work with this model system in mammalian cells. Our data indicate that PrV infection of human cervical cancer (HeLa) cells results in the production of p130, p104/p40 and VCAP, albeit at low levels. While PrV replication in insect cells is associated with the Golgi apparatus and secretory vesicles, in HeLa cells, PrV replication is associated with the mitochondria. It is interesting to note that FHV replication factories are located on the outer mitochondrial membrane.

In an attempt to study PrV virus replication *in vitro,* we adapted the BioID system reported by Roux *et al*. (2012). Here a promiscuous biotin ligase enzyme (BirA) was fused to a protein of interest and the expression of the fusion protein in mammalian cells resulted in the proximity-based biotinylation of proteins associated with the protein of interest. Using p40 as the protein of interest, we studied the fusion protein (BirA-p40) in transiently transfected HeLa cells and in a stable cell line, using western blot analysis and confocal microscopy. We faced challenges comparing the data collected using the two antibody-based detection techniques and the lack of BirA-p40 detection when using western analysis was attributed to the associated of p40 with detergent resistant membranes. BirA-p40 was subsequently expressed using *in vitro* coupled transcription/translation reactions, in the presence of excess biotin. While BirA-p40 was robustly expressed under these conditions, biotinylation of BirA-p40 was not detected. We attributed this to the conditions used in the experiments and given additional time, we would extend the duration of biotinylation, *in vitro.*

PrV replication in mammalian cells was detectable using confocal microscopy however the levels of fluorescence were relatively low. The knowledge that p40 was associated with detergent resistant membranes led us to question the impact of detergent treatment of live cells on the detection of PrV replication. PrV-infected HeLa cells were treated with detergents with varying biochemical characteristics and the impact of these treatments on the detection of PrV replication were evaluated. We observed that linear and non-ionic detergents, namely NP-40 and Triton X-100, were most effective at enhancing the detection of viral replication in PrV-infected HeLa cells. Our data confirm that detergent treatment results in enhanced detection, and not enhanced PrV replication, in HeLa cells. Using the stable BirA-p40 expressing HeLa cell line, we showed that the protein is associated with membranes *in vitro*, and that the enhanced expression of BirA-p40 results in the formation of greater volumes of detergent-resistant membranes. In addition, detergent treatment of unfixed PrV-infected HeLa cells revealed the presence of the PrV p40 protein in the nucleoli of the cells. This is the first report of PrV proteins, which are translated in the cytosol of the mammalian cells, occurring in the nucleus.

Our study has resulted in a deeper understanding of PrV replication in mammalian cell lines. A ‘simple RNA virus’ with only three predicted open reading frames has exhibited high levels of complexity within its elegant simplicity. This study has also highlighted the challenges associated with studying RNA virus replication biology *in vitro.* Looking forward, the identification of detergent-based enhancement for the detection of PrV replication provides the opportunity to perform more targeted PrV replication studies. The PrV-based model system can also be applied to the identification and analysis of potential broad-spectrum antiviral drugs *in vitro*. The latter application is particularly relevant considering the increase in the number of viral outbreaks over the last decade

Chapter one – Literature review

- Discussion of emerging viruses and viral outbreaks, acute and persistent viral infections, antiviral drugs and vaccines that are currently available, the systems in place to study new viruses as well as model systems.
- In summary, there are very few systems or tools in place to study viruses.

Chapter two – Methods and Materials

Chapter three – Providence virus as a model system

- Providence virus (PrV) is capable of replicating in insect, plant and mammalian systems.
- Bioinformatic tools were used to analyse the PrV genome and predict potential proteins produced.
- PrV-specific antibodies with immunoprecipitation reactions and mass spectrometry were used to assess the bioinformatic predictions.
- PrV site of replication in mammalian cells was investigated using confocal microscopy and found to be mitochondria associated.
- PrV was presented as a model system to study viral replication and persistence.

Chapter four – Development of an *in vitro* labelling system

- Modified biotin ligase (BirA) labels proteins in a proximity dependent manner.
- PrV replication accessory protein, p40, was fused to BirA to study the proteins that are associated with the PrV replication complex.
- A stable cell line expressing BirA-p40 was developed but there was difficulty in detection of the protein likely due to membrane association.
- The protein was expressed using a coupled *in vitro* transcription/translation system but was unable to efficiently biotinylate *in vitro*.

Chapter five – Enhancing the detection of PrV replication

- The previous chapters highlighted the challenges of detecting membrane-bound proteins.
- Treatment of live, PrV-infected mammalian cells was found to enhance the detection of viral RNA and p40 protein when analysed by confocal microscopy.
- The viral replication is not enhanced, the detergent treatment increases the availability of the epitopes and increases antibody binding.
- This system could be used as a virus screening tool.

Chapter six – Final discussion and conclusion

- Preliminary data generated provides insight into the potential of PrV-based applications.
- The challenges of working with viruses, antibodies and the variation between applications was highlighted.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Viruses:

BDV – Borna disease virus

CaMV – Carnation mottle virus

DpTV – *Dendrolimus punctatus* virus

EBV – Epstein-Barr virus

FHV – Flock house virus

GRV – Groundnut rosette virus

HaSV – *Helicoverpa armigera* stunt virus

HBV – Hepatitis B virus

HCMV – Human cytomegalovirus

HCV – Hepatitis C virus

HIV – Human immunodeficiency virus

HPV – Human papillomavirus

HSV – Herpes simplex virus

NβV – *Nudaurelia capensis* beta virus

NωV - *Nudaurelia capensis* omega virus

PaV – Pariacoto virus

PFBV – Pelargonium flower break virus

PrV – Providence virus

RSV – Respiratory syncytial virus

TBEV – Tick-borne encephalitis virus

TBST – Tomato bushy stunt virus

VZV – Varicella zoster virus

General:

AF – Alexa Fluor

BCCP – Biotin carboxyl carrier protein

BirA – Biotin ligase

bioAMP – Biotinyl-adenylate

BSA – Bovine serum albumin

CMC – Critical micelle concentration

DAPI - 4′,6-diamidino-2-phenylindole

dddH_2O – Triple distilled water

DRC – Democratic Republic of the Congo

dsRNA – Double-stranded RNA

E. coli - *Escherichia coli*

ER – Endoplasmic reticulum

FB33 - *Helicoverpa zea* fat body cell line

HEK293 – Embryonic kidney 293 cell line

HeLa – Human cervical cancer cell line

HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HLB – Hydrophilic:lipophilic balance

IP – Immunoprecipitation

IRES – Internal ribosome entry site

LA – Luria agar

MCF-7 – Breast cancer cell line

MG8 – *Helicoverpa zea* midgut cell line

MLS – Mitochondrial localisation signal

ORF – Open reading frame

PBS – Phosphate buffered saline

PEP – Posterior error possibility

RdRp – RNA dependent RNA polymerase

RTS – Read through stop

Se-1 – *Spodoptera exigua* embryonic cell line

sgRNA – Subgenomic RNA

TBST – Tris-buffered saline Tween 20

TEM – Transmission electron microscopy

TnT – Transcription/Translation reaction

VLP – Virus-like particle

WHO – World Health Organisation

Acknowledgements

I would like to express my deepest gratitude to the following people:

Firstly, to my supervisor Mez Jiwaji. Thank you for your guidance and continuous support throughout this project. It was a privilege to work with you and thank you for all that you have taught me. You have been so much more than a supervisor and I will miss our 3 hour long meetings discussing new data over coffee at Homeground.

To Professor Rosemary Dorrington who showed enthusiasm and interest in this research. Your knowledge and experience have been immensely valuable and they have resulted in the completion of this project.

To all the people in Lab 325 and 417, thanks for the camaraderie over the years and in particular Val, Mich and Tarryn, for making it all happen behind the scenes.

To Dr Mare Vlok, thank you for sharing your experience and knowledge on mass spectrometry analysis and for your continuous help and generosity.

To Prof Caroline Knox, for your enthusiasm and encouragement throughout my undergraduate and postgraduate studies.

To Dr Earl Prinsloo for sharing your expertise and humour with me.

To my parents, Antony and Penny who supported me through this endeavour with constant interest. Thank you for proofreading my thesis, an important task considering my track record. And to my sister Erin, thank you for coming to visit in Grahamstown and helping with bizarre lab tasks that Lab 325 roped you into.

To dear Simon, thank for your love and support over these two years and for trying to understand what my project was about.

To the Menagerie, William, Marcel and Eric. Thank you for providing a supportive base from which to work and providing much needed light relief when times were tough.

To my cat Luther who lay curled up next to me every day as I wrote this thesis.

To Providence virus; you have taught me to be patient and to always expect the unexpected.

I would like to acknowledge the National Research Foundation and Rhodes University for research funding. I state that all views expressed in this document are my own and should not be attributed to either of the funding bodies.

Chapter 1: Review of literature

1.1 Emerging viruses and viral outbreaks

With the increase in urbanisation, international travel and climate change there is increased exposure to microbiota, some of which are emerging viral pathogens (Garcia-Sastre and Mena, 2013; Geoghegan and Holmes, 2017). The emergence of a new virus in an infection-free area, where the susceptible population have no pre-existing immunity, can result in the rapid spread of disease (Garcia-Sastre and Mena, 2013). Many new emerging viral outbreaks are caused by zoonotic viruses (for example, Ebola virus, Lassa virus, Zika virus and Dengue virus).

Arthropod vectors

The globalisation of arbovirus-induced diseases (arthropod-borne viruses) has been highlighted as a concerning situation. The increase in arbovirus outbreaks is exacerbated by the increased resistance of vectors to pesticides, the poor implementation of vector control stragies and the resistance pathogens to the few, available antiviral drugs (Benelli and Mehlhorn, 2016). In addition, population growth and expansion into tropical areas, which are "hot spots" for vector-borne diseases, contribute to the increased observation of arbovirus-linked outbreaks (Devaux, 2012; Gould *et al*., 2017). Arbovirus outbreaks require the following factors: an adequate population of reservoir and susceptible hosts, an insect population to act as vectors and appropriate climate conditions for viral transmission (Devaux, 2012). Control of arboviruses requires control of the virus vector.

Emerging/re-emerging arboviruses such as Dengue virus (genus- *Flavivirus*; family- *Flaviviridae*), Zika virus (genus- *Flavivirus*; family- *Flaviviridae*) and Chikungunya virus (genus- *Alphavirus*; family- *Togaviridae*) pose major threats to public health. The challenge with these viruses is that the symptoms presented by Dengue, Zika and Chikungunya infections are often clinically indistinguishable, which may result in misdiagnosis (Devaux, 2012; Patterson *et al*., 2016). These three viruses are all transmitted by *Aedes* mosquitoes; Zika virus can be sexually transferred between humans thereafter (Kim *et al.,* 2018; Gould *et al*., 2017; Shan *et al*., 2016). Only 50 % of Dengue virus infections are symptomatic and the clinical presentation of the disease varies greatly. In addition, the course of the disease is unpredictable (Patterson *et al*., 2016). Dengue virus has 4 serotypes, serotype 2 is considered to be the most virulent strain (Khandia *et al*., 2018; Patterson *et al*., 2016; Rothman and Ennis, 2016). A particular challenge with Dengue virus is that antibody-dependent enhancement and subsequent exposure to a different Dengue virus serotype results in a second infection leading to a more severe illness (Khandia *et al*., 2018; Patterson *et al*., 2016; Rothman and Ennis, 2016). The sequence

similarity between the Zika and Dengue virus envelope and non-structural proteins is high (Priyamvada *et al*., 2016). As a result of this similarity, immunological cross-reactivity between Zika and Dengue virus was demonstrated by Priyamvada *et al*. (2016). This highlights the potential of antibody-dependent enhancement of a Zika virus infection (Khandia *et al*., 2018; Priyamvada *et al*., 2016). This is highly concerning, considering that some areas are endemic for both Zika and Dengue virus (Patterson *et al*., 2016).

Until recently, Zika virus was not a cause for concern; individuals infected with Zika virus were either asymptomatic or presented with a mild fever (Patterson *et al*., 2016). A Zika virus infection is now has a more serious health concern due to the link between viral infection and foetal microcephaly and the possibility of virus transmission through sexual contact (Bhatnagar *et al*., 2017; Kim *et al*., 2018; Patterson *et al*., 2016). Chikungunya, like Zika, was not considered a serious health concern as there were few localised outbreaks but since 2004 there have been a number of large-scale outbreaks (Patterson *et al*., 2016). The initial symptoms of a Chikungunya infection are very similar to Dengue and Zika infections (Patterson *et al*., 2016) bar the fact that the majority of those infected with Chikungunya virus are symptomatic.

Zoonotic outbreaks

Many emerging viruses are zoonotic and bats are reported to be important reservoirs for emerging viruses including filoviruses, lyssaviruses, SARS coronaviruses and paramyxoviruses, in particular Hendra and Nipah viruses (Calisher *et al*., 2006; Leech and Baker, 2017; Luis *et al*., 2015). What is most interesting is that the bats do not show signs of viral infection when harbouring viral pathogens, but they are able to transmit infective viruses to susceptible hosts (Leech and Baker, 2017). This is attributed to the constitutively activated interferon response, even in the absence of infection (Leech and Baker, 2017). Bats, like boy scouts, appear to have immune systems that are always prepared.

There are several hypotheses as to why bats are well suited as viral reservoirs. Bats have a relatively long life span when compared to their body size, this factor is thought to facilitate viral persistence (Luis *et al.*, 2015). In addition, the fact that bats live in gregarious colonies and fly may contribute to the increased transmission of zoonotic pathogens (Luis *et al*., 2015).

Despite more than 40 years of research and continuous viral outbreaks, the reservoirs of *Zaire ebolavirus*remain uncertain (Goldstein *et al*., 2018). It is commonly believed that bats are the reservoir for Ebola virus, but because the virus particle, or a full genome sequence, have not been isolated from this host there remains some uncertainty (Goldstein *et al*., 2018; Yang *et*

al., 2019). *Zaire ebolavirus* (genus- *Ebolavirus*, family- *Filoviridae*) is a zoonotic virus that causes severe haemorrhagic fever with a high mortality rate (Gatherer, 2014). The largest *Zaire ebolavirus* (Ebola virus) epidemic recorded to date began in Guinea in February 2014 and spread through Sierra Leone, Liberia and Nigeria (Gire *et al*., 2014). A small number of cases were also reported in Mali. The initial infection is believed to have been via a zoonotic transmission event which then spread through human-to-human contact (Bah *et al*., 2015; Gire *et al*., 2014). As of 1 August 2018, the Democratic Republic of the Congo (DRC) is recorded to be struggling with the second largest Ebola virus epidemic on record (WHO, 2019).

Novel filoviruses

Novel filoviruses have recently been identified in bats, namely Lloviu virus (Kemenesi *et al*., 2018; Negredo *et al*., 2011), *Bombali ebolavirus* (Bombali virus; Goldstein *et al*., 2018) and Měnglà virus (Yang *et al.,* 2019). The sequence of Lloviu virus was first isolated from deceased bats, *Miniopterus schreibersii*, in Spain (Negredo *et al*., 2011) and then several years later in deceased *M. schreibersii* bats in Hungary (Kemenesi *et al*., 2018). Lloviu virus is genetically distinct from other marburgviruses and ebolaviruses and has since been classified into its own genus, *Cuevavirus*, within the *Filoviridae* family (Amarasinghe *et al*., 2019; Negredo *et al*., 2011). There is a possible connection between Lloviu virus infection and the mortality of bats, which suggests that *M. schreibersii* is not the reservoir for this virus (Kemenesi *et al*., 2018; Negredo *et al*., 2011). Negredo *et al*. (2011) found evidence of Lloviu RNA sequences in bat lung and spleen tissue which suggested that Lloviu virus was replicating within the bat. The detection of Lloviu virus in samples from both Spain and Hungary suggests that the range of this virus has been expanded but further investigation is required to determine the consequences of this finding (Kemenesi *et al*., 2018; Negredo *et al*., 2011).

In 2018, Goldstein *et al*. described the complete genome of a new ebolavirus, Bombali virus which was identified in free-tailed bats from Sierra Leone. Phylogenetic analysis of the genome sequence showed that Bombali virus is sufficiently distinct and represents a new species of *Ebolavirus* (Goldstein *et al*., 2018). The authors demonstrated that the viral glycoprotein was able to mediate viral entry into human cells. It is important to note that binding and entry are not the only determinants of host susceptibility, but these factors do represent the first critical step in viral spill-over (Goldstein *et al*., 2018). Although the pathogenic potential of Bombali virus is unknown, the virus's ability to enter human cells suggests that it could pose a threat as a potential zoonotic virus (Goldstein *et al*., 2018).

Měnglà virus was isolated from *Rousettus* bats in China and has since been classified into a new genus, *Dianlovirus* in the family *Filoviridae* (Yang *et al*., 2019). Měnglà virus has a broad cell tropism that is comparable to other filoviruses, in particular, Ebola and Marburg viruses (Yang *et al*., 2019). This cellular tropism demonstrates the elevated potential of interspecies transmission. The identification of new filoviruses from bats from a wide range of locations provides evidence that these animals harbour genetically diverse filoviruses (Yang *et al*., 2019); there is a high likelihood that further viral diversity will be identified in bat species.

Lassa virus

Lassa virus (genus- *Arenavirus*; family- *Arenaviridae*) is a zoonotic virus that can cause haemorrhagic fever and high mortality rates (Andersen *et al*., 2015). Infection with Lassa virus can lead to an acute fever with symptoms similar to those infected with Ebola virus (Andersen *et al*., 2015). In other cases, infection with Lassa virus can remain undetected due to the lack of clinical symptoms (CEPI, 2019). Lassa virus is endemic to West Africa, primarily Sierra Leone, Guinea, Liberia and Nigeria (Andersen *et al.*, 2015; Siddle *et al*., 2018). The virus is maintained in rodent reservoirs, most often in *Mastomys natalensis.* Contact with the rodent excreta permits the infection of human hosts (Andersen *et al*., 2015; Siddle *et al*., 2018). Transmission between humans has been reported to occur but at low frequency (Andersen *et al.*, 2015; Siddle *et al*., 2018).

Nigeria has recently experienced an unusual increase in the number of Lassa fever cases (Siddle *et al*., 2018). This increase has been linked to an increase in cross-species transmission as a result of the increasing rodent population as opposed to an increase in the rate of human-to-human transmission (Siddle *et al*., 2018). The rise in the number of viral outbreaks has put an increasing burden on the health care services of the affected countries and there is currently no vaccine that is available to prevent or reduce the impact of Lassa virus infection (Andersen *et al.*, 2015; CEPI, 2019; Siddle *et al*., 2018). Ribavirin is the only drug that has been approved for the treatment of Lassa fever and it is reported to be effective only in some cases (Debing *et al*., 2013; Hadi *et al*., 2010). Lassa virus has been identified as a priority disease for vaccine production by the Coalition for Epidemic Preparedness Innovations (CEPI, 2019).

Host-switching

Emergence of infectious diseases is reported to occur when a pathogen switches from its native host to a novel species (Longdon *et al*., 2014). Host-switching can result in one of three outcomes: 1- a spillover event that results in dead end infections; 2- short, stuttering transmission chains or 3- a successful host shift that results in infection and sustained

transmission (Longdon *et al*., 2014; Parrish *et al*., 2008). Following a host-shift event, selection favours the viral variants with mutations that allow the pathogen to enter the host cell with greater efficiency and to optimise the fitness of the virus in the new host. This includes the virus' improved use of the host cellular machinery; the enhanced ability to evade immune responses or to actively suppress the host's immune responses; or enhancing virulence and transmission (Longdon *et al*., 2014; Parrish *et al*., 2008). While host-shifts are more likely to occur between species that are closely related, there are reports of viruses that can transfer between hosts that span large phylogenetic distances. As an example, the tobacco ringspot virus is a plant pathogen that can infect and replicate in honeybees (Lian *et al*., 2014). This highlights the possibility of virus host-switching and RNA viruses, due to the nature of their error-prone replication mechanisms, are a cause for particular concern.

There has been increased interest in virus emergence patterns and whether we can predict the next viral outbreak. Factors such as taxonomic relatedness of the original host species to the new host species, geographical overlap and host range have to be considered when predicting whether a virus is capable of crossing the species barrier (Woolhouse *et al*., 2012). In addition, the abilitity of a virus to cross the species barrier is not the only contributor to the success of the virus. The availability of susceptible hosts plays a large role in the transmissibility of the virus (Geoghegan and Holmes, 2017).

In 2016, there was the development of an initiative under the Global Virome Project which aimed to identify and characterise 99 % of zoonotic viruses with epidemic potential (Daszak *et al*., 2016). Using metagenomic surveys of viruses in vertebrate populations to better predict, prevent and respond to future viral threats. However, the true scale of the virosphere can only be imagined. And so, it would be more beneficial to predict the impact a new epidemic would have on a population and establish strategies to combat the outbreak than to try and predict, where, when and what virus could cause the next epidemic (Geoghegan and Holmes, 2017). The link between climate change and emerging diseases has be raised recently in a book titled, *The Stockhlom Paradigm: Climate change and emerging diseases*. The authors (Brooks *et al*., 2019) discuss how the effects of climate change have increased humans exposure to to previously unknown pathogens (Cable, 2020). The authors provide a possible solution to combat the rise in the emerging diseases: DAMA (document, assess, monitor and act). With the aim of increasing the efforts toward the discovery and documenting pathogens in the environment that have the potential of causing the next epidemic (Cable, 2020).

1.2 Acute vs persistent viral infections

Most virus infections induce characteristic symptoms of acute disease and are described as self-limiting infections (Kane and Golovkina, 2010; Randall and Griffin, 2017). During acute infections, the virus replicates rapidly within the host cells and virus particles are shed into the environment to infect the next susceptible host (Randall and Griffin, 2017). The virus is either cleared by the host's immune system or, alternatively, infection with the virus results in the death of the host (Kane and Golovkina, 2010; Randall and Griffin, 2017). Provided there is a continuous supply of susceptible hosts, the virus is maintained within a population and the transmission lifecycle of the virus is sustained (Randall and Griffin, 2017).

A subset of viruses can establish persistent infections within the host (Kane and Golovkina, 2010). Persistent viral infections are characterised as virus infections that are not cleared by the immune system but instead remain within specific cells in the host (Boldogh *et al*., 1996). Persistent infections can be described as either latent or chronic infections. Latent infections occur when there are no virus particles produced in between episodes of reoccurrence, for example, during infection with Herpes simplex virus (Boldogh *et al*., 1996; Goodrum *et al*., 2012; Kane and Golovkina, 2010). Chronic infections are caused by viruses that continuously produce infectious virus particles after the initial infection; this infection requires the virus to employ immune evasion strategies (Boldogh *et al*., 1996; Kane and Golovkina, 2010; Zuniga *et al*., 2015). Hepatis C virus (HCV) is an example of a virus that establishes a chronic persistent infection and so results in the development of chronic liver disease and, in some cases, cancer (Bartenschlager *et al*., 2018).

For a persistent viral infection to occur, a number of requirements need to be met. These include the infection of a subset of host cells that are suitable for the long-term maintenance of the viral genome; the modulation of viral gene expression during infection; the subversion of the host cell's antiviral and apoptotic pathways; and the avoidance of detection and clearance by the host's immune system (Kane and Golovkina, 2010; Randall and Griffin, 2017). The ability of a virus to establish a persistent infection is greatly influenced by the competence of the host's immune system (Randall and Griffin, 2017). Hosts with compromised immune systems or immunodeficiencies are susceptible to the development of persistent viral infections or progressive infections induced by both attenuated and wild-type viruses. All viruses need to circumvent the interferon response to some degree to establish a persistent infection. Viruses accomplish this by either hiding or by modifying their genomes. Alternatively, they produce proteins that act as antagonists of the interferon response system (Randall and Griffin, 2017; Randall and Goodbourn, 2008).

Persistent DNA viruses

DNA viruses establish persistent infections by either integrating the viral genome into the host cell genome for example, Hepatitis B virus (HBV) (Tang *et al*., 2018); or by regulated association, for example, Herpes simplex virus (HSV) (Roizman *et al*., 2011). These viruses alternate between an acute infection, where symptoms are displayed; and latency, where no replication occurs and no virus particles are produced (Goodrum *et al*., 2012). These characteristics present a great challenge for the treatment of persistent DNA viruses.

Goodrum *et al*. (2012) suggested that viral persistence as a coexisting strategy comes at the cost of moderating viral replication and therefore reduced pathogenesis. This hypothesis appears to be correct for human cytomegalovirus (HCMV); for as long as the HCMV infection remains undetected in immunocompetent host cells, there is no detectable pathology (Goodrum *et al*., 2012; Irwin *et al*., 2016). HCMV is reported to generate latent virus reservoirs in haematopoietic, endothelial and epithelial cells (Goodrum *et al*., 2012).

HBV (genus- *Orthohepadnavirus*; family- *Hepadnaviridae*) is a particularly interesting DNA virus with respect to the replication strategy employed. Genome replication involves the transcription of DNA to RNA intermediates which are then reverse transcribed into DNA (Irwin *et al*., 2016). HBV is hypothesized to exist as a quasi-species; the increase in genome diversity can be explained by the lack of proofreading during reverse transcription (Irwin *et al*., 2016). HBV integrates into the host's genome and persists indefinitely within the nucleus of the long-lived hepatocytes (Tang *et al*., 2018). These cells act as viral replication reservoirs (Tang *et al*., 2018). A chronic HBV infection can result in liver cancer, with HBV infections accounting for 50 % of all hepatocellular carcinoma cases (Parkin, 2006; Tang *et al*., 2018). There exists the concern that only one third of adults infected with HBV develop symptoms during the acute phase of infection (Tang *et al*., 2018). Large numbers of asymptomatic individuals are therefore untreated and may develop chronic HBV and liver-associated diseases (Tang *et al.*, 2018). The development of a HBV vaccine has resulted in the reduction of new HBV infections, particularly in America (Liu *et al*., 2016; Nelson *et al*., 2016).

Persistent RNA viruses

RNA viruses have evolved mechanisms that permit the continued infection of a diverse range of hosts. The high mutation frequency of RNA viruses, which results in the ongoing selection of antigenic variants; this tactic is employed by Influenza virus (Randall and Griffin, 2017). Viruses such as Respiratory Syncytial virus (RSV) infect mucosal cells; in these host cells, it is difficult to induce long-lasting protective immunity and repeated infection with the same virus is possible (Heylen *et al*., 2017; Randall and Griffin, 2017). Viruses can infect multiple species,

which increases the number of susceptible hosts (Randall and Griffin, 2017). HCV and Borna disease virus (BDV) have both evolved mechanisms that permit them to establish persistent infections in human hosts; these hosts can often act as reservoirs and enable the transmission of disease (Randall and Griffin, 2017). Acute RNA viruses such as Zika and Ebola virus have been reported to establish persistent infections within the host for a few months after initial infection. These viral reservoirs may represent sources of virus that are required for a subsequent viral outbreak (Harrower *et al.,* 2016; Heeney, 2015; Randall and Griffin, 2017). Despite the importance of persistent viral infections, the mechanisms involved in persistence and in virus-induced chronic diseases are not well understood (Randall and Griffin, 2017).

There are a number of RNA viruses that are known to establish persistent infections in humans. This literature review will focus on HCV as it is a well-documented persistent virus. In addition, we will discuss Ebola and Zika virus, both of which are emerging, acute viruses that have been reported to establish persistent infections.

HCV

Infections with HCV (genus- *Hepacivirus*; family- *Flaviviridae*) are a major cause of acute liver disease. Approximately 80 % of infected individuals are able to clear the infection while 20 % of those infected develop chronic liver disease and, possibly, liver cancer (Bartenschlager *et al*., 2018; Burke and Cox, 2010). HCV is one of seven oncogenic viruses that infect humans; of these seven viruses, HCV is the only positive-sense RNA virus (Bartenschlager *et al*., 2018). Highly effective antiviral drugs are available to cure the majority of HCV infections (Bartenschlager *et al*., 2018; Pawlotsky, 2016), however individuals treated with anti-HCV antivirals remain susceptible to reinfection with HCV and the re-establishment of a persistent HCV infection (Grebely *et al*., 2017; Midgard *et al.*, 2016).

HCV has developed a number of mechanisms to evade detection by the virus-specific immune responses within the host cells thereby enabling it to establish persistent infections (Barathan *et al*., 2018). Mechanisms include mutation of the viral genome; viral replication in immunologically privileged sites; and the production of antagonistic proteins that contribute to HCV persistence (Barathan *et al*., 2018). HCV is an extremely heterogenous virus; 7 major genotypes and more than 85 subtypes have been identified. This heterogeneity is attributed to the error-prone nature of RNA polymerase-based viral replication (Irwin *et al*., 2016). The frequent mutations generate a variant population that circulates within a single individual as a HCV quasi-species (Bartenschlager *et al*., 2018; Bukh, 2016). The numerous HCV variants within a single infection result in varying levels of viral persistence as well as differing susceptibility to antiviral drugs (Barathan *et al*., 2018; Irwin *et al*., 2016). This presents a great

challenge in the development of appropriate vaccines as well as effective antiviral drugs (Bartenschlager *et al*., 2018; Bukh, 2016; Irwin *et al*., 2016).

Ebola virus

Infection with Ebola virus results in the onset of severe, acute disease symptoms including a haemorrhagic fever; those infected with Ebola virus have a high mortality rate (Hoenen *et al.,* 2019). Bats are thought to be the natural hosts of filoviruses and outbreaks are attributed to either direct transmission from bats to humans or through an intermediate, such as a non-human primate (Hoenen *et al.*, 2019; Schuh *et al*., 2017). Ebola virus can also be transmitted through human-to-human contact with infected people or through infected bodily fluids (Hoenen *et al*., 2019). In survivors of Ebola virus disease, Ebola virus RNA has been detected in patients' months after recovery (Whitmer *et al*., 2018). The Ebola virus persists in immune privileged sites and in semen, aqueous humor in the eye and in urine (Whitmer *et al*., 2018). Recent research suggests that Ebola virus performs long-term maintenance on the viral genome which includes active transcription and replication (Whitmer *et al*., 2018). As viral persistence progresses, the level of viral replication decreases. Ebola virus has been detected almost a year after the initial and acute infection (Whitmer *et al*., 2018). The mechanisms involved in the transition from an acute viral infection to a persistent infection are not fully understood. This information does however highlight that Ebola virus survivors may be potential Ebola virus reservoirs and serve as sources of virus in a new Ebola virus outbreak (Whitmer *et al*., 2018). It is important to note that there has been no report of a survivor of an Ebola virus infection acting as the source of a new outbreak, to date.

Zika virus

Symptoms of Zika virus infection are comparable to the symptoms of infection with Dengue virus however recently, infection with Zika virus has also resulted in congenital defects in infants (Bhatnagar *et al*., 2017; Patterson *et al*., 2016). The mechanism of intra-uterine transmission and viral pathogenesis that results in microcephaly and other congenital abnormalities are not completely understood (Bhatnagar *et al*., 2017). Although Zika virus is classified as an acute virus, which is cleared from the host within a couple weeks, recent reports have demonstrated that Zika RNA can be detected in the semen and placenta for months after recovery from the symptoms of Zika infection (Atkinson *et al*., 2017; Bhatnagar *et al*., 2017). Zika virus has also been reported to cause ocular manifestations (Singh *et al*., 2018). Most arboviruses establish lifelong persistent infections in arthropod hosts yet cause acute infections in human hosts. It is important to note that the ability of these viruses to establish persistent infections in vertebrate hosts may be underestimated (Kuno, 2001).

There is much to be learnt about the mechanisms of viral persistence but the continuing development of new technologies, for example next generation sequencing, provide increasing opportunities to study viral persistence both *in vivo* and *in vitro*.

1.3 Antiviral drugs and vaccines

Antiviral drugs

Antiviral drugs need to effectively neutralize the threat of the virus however there is also a clear need to decrease the rate at which mutations that confer viral resistance to the drug occur (Irwin *et al*., 2016). This is achieved by developing new drugs or combinations of drugs that require several mutations for antiviral drug resistance to be conferred (Irwin *et al*., 2016).

Therapeutic approaches can be divided into two groups; the approaches can either directly target viral components or processes, such as the viral polymerase. Alternatively, antiviral drugs can indirectly target the virus by interfering with host mechanisms used by the virus to achieve viral replication. This would include impairing host factor activities or interfering with the interaction of the virus with the host factors (Hoenen *et al*., 2019). These mechanisms can also include stimulating the host's immune responses or ameliorating the disease process without interfering with the virus (Hoenen *et al*., 2019; Zumla *et al*., 2016). Direct antivirals are more favourable as the drug will affect the virus and not the host. These antivirals are however more susceptible to viral resistance (Hoenen *et al*., 2019).

Since the first antiviral drug was approved in 1963 there have been approximately 90 antiviral drugs formally approved for the treatment of 9 human diseases (De Clercq and Li, 2016). Of the 9 human diseases, 5 are DNA viruses (HBV, HCMV, HSV, Human Papillomavirus (HPV) and Varicella Zoster virus (VZV)), 3 are RNA viruses (HCV, RSV and Influenza) and one retrovirus namely, Human Immunodeficiency virus (HIV; De Clercq and Li, 2016). Of these antiviral drugs, 11 are approved for the treatment of more than one virus; Ribavirin is effective against HCV, RSV and Influenza. In some cases, these approved antiviral drugs can be used for off-label treatments, for example Ribavirin can be used for the treatment of Lassa virus disease (Debing *et al*., 2013; De Clercq and Li, 2016).

There are only 3 RNA viruses that have antiviral drugs approved for their treatment and the majority of the available antiviral drugs are targeted towards HCV (De Clercq and Li, 2016; Hoenen *et al*., 2019). Recently, highly effective direct-acting antiviral drugs have been approved for the treatment of HCV (Hoenen *et al*., 2019; Pawlotsky, 2016). These direct-acting antivirals took 25 years to develop and secure approval for therapeutic use (Pawlotsky, 2016).

This highlights the delay between identification of an effective drug compound and its application to treatment of virus-induced disease.

Vaccines

Vaccines are the most effective strategy that can be employed to prevent viral infections and no infectious viral disease has been controlled to date without a functional vaccine. One relevant example is the control of the smallpox virus (Bartenschlager *et al*., 2018; Breman *et al*., 1980; Liu *et al*., 2016). Most vaccines use immunologically relevant viral antigens rather than the whole virus. Individual antigens however are often immunogenically poor and the immune response is sub-standard outside the context of a viral infection (Garcia-Sastre and Mena, 2013). For this reason, there is a push to develop vaccines that are highly immunogenic as well as safe to use. There are several alternative strategies that are employed including the development of virus-like particles (VLPs) and replication competent viral vectors (Garcia-Sastre and Mena, 2013).

VLPs are generated from viral proteins that self-assemble into structures that resemble native viral particles (Liu *et al*., 2016). These VLPs are safe to use as the particles do not contain a viral genome and therefore are unable to establish a productive viral infection (Liu *et al*., 2016; Garcia-Sastre and Mena, 2013). VLPs have been used to develop an effective vaccine against HBV and HPV (Liu *et al*., 2016). Unfortunately, there are currently no VLP vaccines for emerging viruses. Challenges include a lack of knowledge about new emerging viruses, the purification technology required for VLPs is limited, and the time it takes to develop and test new vaccines is significant. All these factors contribute to the challenge of developing new and effective VLP-based vaccines (Liu *et al*., 2016; Garcia-Sastre and Mena, 2013).

Replication competent viral vectors have been used for decades for protein expression and for vaccination (Garcia-Sastre and Mena, 2013). Adenovirus, vaccina virus and herpes viruses have been used as vector platforms to develop vaccines (Parks *et al*., 2013). The replication competent viral vectors can be manipulated to enhance viral safety and immunogenicity. Virulence factors can be removed, envelope proteins can be altered to change viral tropism and non-essential genes can be removed to increase the coding capacity of the vector (Garcia-Sastre and Mena, 2013). This allows the expression of the viral antigens in the context of a viral infection (Garcia-Sastre and Mena, 2013; Parks *et al*., 2013). However, the use of replication competent viral vectors does have disadvantages: the antigens from the vector may cause competition with the antigen of interest (immuno-dominance); the presence of pre-existing immunity against the vector may result in a reduction in efficacy; and in some cases, the vector pathogenesis can raise safety concerns (Garcia-Sastre and Mena, 2013).

Replication competent viral vectors have however shown promise, and have been used to generate vaccines, for example against HIV (Parks, 2017; Parks *et al*., 2013).

There are 15 viruses that have approved vaccines: human adenovirus, HBV, VZV, HPV, smallpox, rotavirus, yellow fever virus, Japanese encephalitis virus, Hepatitis E virus, poliovirus, Influenza virus, rabies virus, rubella virus, measles and mumps (De Clercq and Li, 2016). This a miniscule number of antiviral vaccines especially when compared to the number of known human-infecting viruses. This highlights the desperate need to prioritise the development of more antiviral vaccines.

Very recently (11 November 2019), an Ebola vaccine (Ervebo) that was developed by Merck was approved by the European Medicines Agency for production (Callaway, 2019). The vaccine has already been administered to individuals during the current Ebola virus outbreak in the DRC and is reported to be highly effective at preventing infection with Ebola virus (Callaway, 2019). The WHO announced that this Ebola vaccine is "prequalified" as it meets the agency's standards for quality, safety and efficacy. This highlights two important factors; firstly, drugs can move rapidly from discovery to market if the situation requires it and secondly, unapproved drugs can be utilized during an outbreak if the severity of the situation requires it. Due to the recent release, there is no data to evaluate the duration of protection and whether additional boosters will be required to extend immunity (Callaway, 2019). A second Ebola vaccine, produced by Johnson and Johnson, is currently being administered to individuals in the DRC as part of a major clinical trial (Mazumdar, 2019). The Johnson and Johnson vaccine is being tested as a pre-exposure vaccine that will complement the vaccine being produced by Merck, which is administered to individuals exposed to Ebola (Mazumdar, 2019). The weakness of the Johnson and Johnson vaccine is that it requires two doses, administered 56 days apart (Mazumdar, 2019). A lack of compliance with the vaccination schedule means there will be a collection of individuals who are partially protected because they received only one of the two doses. This may result in a lack of confidence in the vaccine or incomplete efficacy, both of which are undesirable.

Challenges in the development of antiviral drugs and vaccines

Antiviral drug development

Current antiviral therapies, particularly those used to treat HIV and HBV, are effective at reducing viral replication but they are incapable of eliminating the virus (Blair and Cox, 2016). And so, new antiviral therapies must be developed and must continue to evolve, overcoming the challenges of the resistant viral populations and also the high cost of drug development and production (Irwin *et al*., 2016; Lipsitch *et al*., 2012).

Antiviral drug resistance

If an antiviral treatment is robust and viral fitness is impaired, replication of the viral genome is unsuccessful (Irwin *et al*., 2016). If, however, an antiviral treatment is not perfect and some viral genomes are replicated, the selective pressure imposed by the treatment may result in the rapid adaptation of the virus and the generation of drug-resistant viral strains (Blair and Cox, 2016; Irwin *et al*., 2016). The development of resistant viral strains is exacerbated by the large circulating virus populations as well as the high mutation frequency associated with RNA viruses in particular (Blair and Cox, 2016; Irwin *et al*., 2016). This phenomenon has pushed the development of new classes of antiviral drugs and also resulted in treatment with a combination of drugs (Irwin *et al*., 2016).

Viruses, such as HIV, are treated with a combination of antiviral drugs (combination antiretroviral therapy) in order to increase the genetic barrier and effectively control viral replication (Blair and Cox, 2016; Irwin *et al*., 2016). In contrast, HBV is commonly treated with a single reverse transcriptase inhibitor. It is possible that this may be the reason why there are HBV strains that are resistant to this drug. HBV resistance to this first-line drug requires the use of a second-line reverse transcriptase inhibitor. While effective in the short-term, this method of drug-switching within a class of drugs can encourage the selection of resistance mutations (Irwin *et al*., 2016). This highlights the importance of combination drugs as well as the use of the correct drug when treating a viral disease.

Vaccine development

Despite the importance of vaccines, the challenges associated with their development has resulted in the availability of a few vaccines for viruses. Conventional vaccination strategies have proved effective in some cases but a lack of required viral immunogenicity and a failure to reach the required level of safety and cross-protection across the different strains has hampered vaccine development (Garcia-Sastre and Mena, 2013; Liu *et al*., 2016). For example, Influenza vaccines are effective only against some viral strains; this is because viral strains differ between seasons and pandemics (Liu *et al*., 2016). In some cases, the administration of a vaccine exacerbates the viral disease as described with Dengue virus (de Silva and Harris, 2018; Garcia-Sastre and Mena, 2013).

1.4 Systems to study virus replication biology

Replicons

Replicons are systems used to study viral replication; the structural genes required to form viral particles are removed and only the genes required for replication are encoded (Xie *et al*., 2016). This provides a safe system that can be used to study viral replication without the threat of infectious particle formation. The replicon system permits the study of replication events, including viral translation and RNA synthesis in the absence of viral entry and virion assembly. This approach also permits the study of mechanisms involved in antiviral inhibition and intracellular immune evasion (Xie *et al*., 2016).

Replicon systems have been applied to the study of many flaviviruses including Dengue virus (Ng *et al*., 2007), yellow fever virus (Jones *et al*., 2005), HCV (Uprichard, 2010) and Zika virus (Xie *et al*., 2016). Yang *et al*. (2019) used a chimeric mini-genome system to study Mengla virus and to demonstrate that the replication complex is functional. Manhart *et al*. (2018) used the chimeric mini-genome system to demonstrate that Lloviu virus replication is more similar to that of ebolaviruses than marburgviruses. The replicon-based system is therefore one system to study the replication of novel viruses without the need for extensive prior knowledge about the virus and its replication biology.

Mathematical modelling

Mathematical modelling of viruses and virus infections has become a useful tool in the analysis of aspects of the viral lifecycle (Lessler and Cummings, 2016; Mohammadi *et al*., 2015). There are two broad categories of dynamic modelling approaches, mechanistic and statistical (Mohammadi *et al*., 2015). Mechanistic approaches analyse the changes within a set system and make hypotheses about the biological mechanisms involved during viral infections (Lessler and Cummings, 2016; Mohammadi *et al*., 2015). Mechanistic models were initially designed to study the long-term dynamics involved in viral latency (Mohammadi *et al*., 2015). They have been used to model HIV infection and the effect of drug treatment on HIV infection, making predictions about the complex changes that occur during the process (Mohammadi *et al*., 2015). More recently, models have been used to predict the outcome of an HIV/HPV co-infection (Verma *et al*., 2017).

Statistical models make use of transcriptomic and proteomic data sets to analyse changes in the cellular state at a specific point in time (Mohammadi *et al*., 2015). They provide a holistic view of the cellular changes that occur during viral infections by using repeated high-throughput measurements to analyse the dynamics within the cell (Mohammadi *et al*., 2015). These data facilitate the cause and consequence observations that occur as a result of viral infection and viral replication, viral latency and virus reactivation (Mohammadi *et al*., 2015). Statictical models have been used to analyse the effect of respiratory viral infections (de

Lamballerie *et al*., 2019) and the host response to HIV infections (Golumbeanu *et al*., 2019). Attentions are now focusing on analysing viral infections at a single-cell level (Rato *et al*., 2017).

Mathematical models are used on a global scale to assess the risks associated with the spread of infectious diseases and the emergence of epidemics (Walters *et al*., 2018). These models can be used to predict the probability that a disease will occur in a particular country, the expected number of cases within a specific timeframe as well as the effect of interventions employed in the treatment of the diseases (Walters *et al*., 2018).

1.5 Model systems

Model systems are incredibly valuable for the the study of virus-host interactions, progression of viral infection and viral replication biology. The aim of a virus model system is to make it applicable to other viruses that are more difficult to study, whether it be for biological reasons or safety-related reasons. There are only a few examples of robust model systems that can be used to study virus replication biology including those based on Poliovirus, tick-borne encephalitis virus and HCV.

Poliovirus model system

Poliovirus has been eradicated in many parts of the world because of a rigorous vaccine administration program. Poliovirus does however provide a useful model to study RNA viruses. Poliovirus has been used to demonstrate that RNA viruses benefit from the use of an error-prone replication strategy (Pfeiffer and Kirkegaard, 2005). Based on this study, other RNA viruses have also been shown to benefit from error-prone replication, including Chikungunya virus (Coffey *et al*., 2011). Bird and Kirkegaard (2015) used poliovirus in an elegant study to demonstrate the non-lytic spread of viruses between cells; here viral RNA contained in vesicles permitted the transfer of infective RNA between virus-infected and uninfected cells.

Flavivirus model systems

Tick-borne encephalitis virus (TBEV; genus- *Flavivirus*, family- *Flaviviridae*) infection can cause severe encephalitis and neuron damage in a number of cases (Yau *et al*., 2019). Yau *et al*. (2019) used TBEV to design an inducible, cell-based model system to study flavivirus replication complex formation, in the absence of viral replication. They used viral subgenomes as replicons, these contained only the non-structural proteins involved in the replication complex formation (Yau *et al*., 2019). Yau *et al*. (2019) proposed that the platform be utilized to investigate and identify the intracellular inhibitors that target the TBEV replication complex *in vivo*.

HCV as a model system

HCV has been used as model system to study virus-host lipid interactions to better understand HCV virus replication mechanisms (Bartenschlager *et al*., 2018). HCV hijacks the cellular lipid metabolism in the host cell and expands the ER membranes to generate membranous replication factories. This process requires extensive reprogramming of the steps in cellular lipid biosynthesis as well as in the transport pathways (Bartenschlager *et al*., 2018). The heterogenous HCV genome also provides the opportunity to pursue fundamental studies that focus on RNA virus evolution (Bartenschlager *et al*., 2018). The ability of HCV to establish a persistent infection in hepatocytes permits this to serve as a model system to study viral persistence (Burke and Cox, 2010). The challenge with using HCV as a model system is that the virus is incapable of replicating under cell culture conditions; the study of HCV in culture relies on adapted variants and recombinant *in vivo* and *in vitro* systems (Bartenschlager *et al*., 2018; Bukh, 2016). The development of a complete *in vivo* HCV model system in which the virus can persistently replicate is critical for meaningful research into virus replication mechanisms.

1.6 Motivation, aims and objectives

Motivation

There is currently no available model system to study RNA virus persistence and replication biology in mammalian cells. This provides an opportunity to develop a cell culture-based viral model system to study persistent viral infection and replication biology. Providence virus (PrV) is a small, positive-sense RNA virus that can persistently infect mammalian cell culture lines. Here, we use this opportunity to develop PrV as a model system to investigate mechanisms involved in viral persistence and RNA virus replication biology.

Aim

To use a PrV-based model system in combination with an *in vitro* labelling system, to study persistent virus replication biology in mammalian cells.

Objectives

- To investigate PrV as a model system by performing a bioinformatic analysis of the PrV genome and its putative translation productions. To validate the bioinformatic analysis using confocal microscopy, western analyses and mass spectrometry.
- To develop an *in vitro* labelling system using a promiscuous biotin ligase enzyme fused to a PrV replication-associated protein. To validate this *in vitro* labelling system within PrV-infected mammalian cells.
- To study persistent replication biology, using PrV as a model system.

Chapter 2: Methods and Materials

2.1 Bioinformatic analysis

Genome and gene analysis:

The PrV genome (NCBI accession number NC_014126.1) was submitted to GeneMarkS (http://opal.biology.gatech.edu/GeneMark/genemarks.cgi; accessed 4/12/2018) and SoftBerry (http://www.softberry.com/berry.phtml?topic=virus&group=programs&subgroup=gfindv; accessed 4/12/2018) to predict genes that may be present in the PrV genome. The results of the analyses were viewed in SnapGene® Viewer 4.2.6. The predicted protein sequences were submitted for BLAST analysis using the blastp algorithm and the non-redundant protein database.

Subcellular localisation of p40 and p104:

The amino acid sequences of p40 (YP_003620398.1) and p104 (YP_003620397.1) were submitted to the following online subcellular localisation sites: UCL Bioinformatics (http://bioinf.cs.ucl.ac.uk; accessed 7/12/2018) , Predict Protein (https://open.predictprotein.org; accessed 23/10/2019), CELLO2GO (http://cello.life.nctu.edu.tw/cello2go/; accessed 23/10/2019).

2.2 Cell culture

Human cervical cancer (HeLa), breast cancer (MCF-7) and embryonic kidney 293 (HEK293) cell lines were maintained in Dulbecco's Modified Eagle Medium (DMEM; Gibco®) supplemented with 10 % foetal calf serum (Thermo Fisher, Cat. # 1049904) and 1% Pen/Strep Fungizone (Thermo Fisher, Cat. # 15240062). The cells were grown in vented flasks at 37 °C in an atmosphere containing 5 % CO_2.

2.3 Plasmid purification and transient transfections

Plasmid purification and confirmation:

Plasmids (pBirA, pBirA-p40 and pBirA-CNK) were generated in a pcDNA 3.1 vector backbone. The sequence of the BirA gene was informed by the sequences published by Roux *et al*. (2012) and was first reported by Moodley (2019).

Competent *E. coli* DH5α were transformed with pBirA and pBirA-p40 and the cells were plated on Luria agar (LA) containing 100 µg/µl Ampicillin. The LA plates were incubated overnight at 37 °C. Colonies were picked and inoculated into Luria broth containing 100 µg/ml ampicillin.

The plasmids were purified using the PureYield™ Plasmid MidiPrep kit (Promega, Cat. # PRA6742) according to the manufacturer's instructions.

The identity of the plasmids was confirmed using restriction enzyme digests with the enzymes *Bam* HI and *Xba* I. The reactions were set up as described in a total volume of 20 µl and were then incubated at 37 °C for 2 hours; 10X restriction enzyme buffer (Buffer E), 10 µg/µl acetylated BSA, 200 ng plasmid DNA, 10 U/µl restriction enzymes (*Bam* HI and *Xba* I). The restricted DNA was separated on a 1 % (w/v) agarose gel containing SYBR Safe (1:20 000).

Transient transfections:

HeLa cells were plated on glass coverslips in a 24 well plate and permitted to settle overnight. When the cells were between 50 and 60 % confluent, they were transfected with 1 µg of pBirA or pBirA-p40 using Xfect (Separations, Cat. # 631318) according to the manufacturer's instructions. The transfected cells were incubated for 24 hrs and then prepared for confocal microscopy as described below. For protein isolation, HeLa cells were plated in a 6 well plate and permitted to settle overnight. These cells were transfected with 6 µg pBirA or pBirA-p40 DNA when the cells were between 50 and 60 % confluent. The proteins were permitted to express for 24 hrs and were then collected for western blot analysis.

2.4 Generation of a stable cell line expressing BirA or BirA-p40

HeLa cells were transfected with pBirA and pBirA-p40 and were grown in non-selective medium for 48 hrs. The non-selective growth medium was replaced with growth medium containing 1 mg/ml G418 sulphate (Geneticin; Inqaba Biotech, Cat. # E859). The medium was changed every second day until a population of Geneticin-resistant cells were growing in the cell culture. These cells were maintained in 0.75 mg/ml Geneticin and 50 µM biotin (dissolved in DMSO; Sigma-Aldrich, Cat. # B469-1G).

2.5 Western blot

Cells were lysed in Cell Lytic M (Sigma Aldrich, Cat. # C2978) containing cOmplete Protease Inhibitor Cocktail Tablets (Sigma Aldrich, Cat. # P8340). The cell debris was removed using centrifugation (2000 rpm for 2 minutes) unless otherwise stated. The cell lysate was mixed with 4X SDS sample buffer, incubated at room temperature for 5 minutes and then boiled for a further 5 minutes. The proteins were separated on a 10 % SDS acrylamide gel and then transferred onto Immobilon-P PVDF membranes (Merck Millipore, Cat. # IPVH00010) using the Trans-Blot Turbo Transfer system (BioRad). The membranes were blocked overnight at room temperature in 5 % BLOTTO (5 % non-fat milk powder in Tris buffered saline containing 1 % Tween 20 (TBST)). The membranes were rinsed in TBST and then incubated in primary

antibody, diluted in 1 % BSA (bovine serum albumin, fraction V) in TBST, overnight at 4 °C. The primary antibodies were used at a dilution of 1:1000, unless otherwise stated, and include anti-p17 (rabbit), anti-p104 (rabbit), anti-VCAP (rabbit), anti-biotin (rabbit, Abcam, Cat. # ab53494), anti-p113 (mouse), anti-myc clone 4A6 (mouse, Merck, Cat. # 05-724), anti-actin (I-19) (goat, 1:500; Santa Cruz Biotechnologies, Cat. # sc-1616) and IgG-biotin-linked anti-p40 (expanded PrV antibodies in Appendix A5). The membranes were washed twice in TBST for a period of 1 minute and then incubated in the secondary antibodies for 90 minutes at room temperature. We used goat anti-mouse HRP (1:10 000; Advansta, Cat. # R-05071-500), goat anti-rabbit HRP (1:20 000; Advansta, Cat. # R-05072-500), donkey anti-goat HRP (1:10 000; Advansta, Cat. # R-05077-500), streptavidin HRP (1:10 000; Thermo Fisher, Cat. #SA10001). The membranes were washed four times for 10 minutes before visualisation using the WesternBright ECL HRP substrate (Advansta, Cat. # K-12045-D50) on a ChemiDoc XRS+ (BioRad). Images were analysed using Image Lab 5.2.1 software.

2.6 Immunoprecipitation (IP) reactions

PrV-infected HeLa cells were lysed with Cell Lytic M containing protease inhibitors. The lysate was mixed with primary antibodies and incubated overnight at 4 °C, with rotation (18 rpm). Pierce Protein A/G magnetic beads (Thermo Fisher, Cat. # 88803) were washed twice with TBS containing 0.05 % Tween 20 (wash buffer); the washed beads were mixed with the antibody-lysate mixture and then incubated overnight at 4 °C, with rotation. The beads were collected using a magnet and the depleted lysate was removed. The beads were washed twice with wash buffer and then stored in Cell Lytic M containing protease inhibitors. The samples were analysed using western blot analysis as described above.

2.7 Determining the location of PrV translation

PrV-infected HeLa cells were treated with 20 µg/ml chloramphenicol for 48 hours or with 10 µg/ml cycloheximide for 24 hours. Treated cells were lysed, and the supernatant was used in IP reactions with IgG-biotin anti-p40 as described above. The immunoprecipitated proteins were separated and analysed using SDS-PAGE and western blots. The membranes were probed with IgG-biotin anti-p40 and rabbit anti-p104 primary antibodies and streptavidin HRP and goat anti-rabbit HRP secondary antibodies were used.

2.8 *In vitro* transcription/translation reactions

RNase-free plasmid DNA was prepared using the QiaPrep Spin MiniPrep kit (Qiagen, Cat. # 27104). The plasmid DNA was used in coupled transcription/translation reactions using the TnT T7 Quick Coupled Transcription/Translation System (Promega, Cat. # L1170). The reactions were set up as follows: 40 µl TnT Quick Master Mix (per reaction), 1 mM Methionine,

1 µg/µl plasmid DNA (pBirA, pBirA-p40, pBirA-CNK) and 5 µM biotin. The negative control contained no plasmid DNA. The total reaction volume was 50 µl. The reactions were incubated at 30 °C for 90 min. The proteins were separated using 7.5 % SDS-PAGE and analysed using western blots.

IP reactions of proteins expressed in the TnT reactions:

10 µl of the TnT reaction supplemented with pBirA-p40 was mixed with 300 µL Cell Lytic M containing 2X protease inhibitor. 1 µg/ml of anti-myc (mouse) antibodies was added to the TnT reaction. The samples were incubated overnight at 4 °C. A/G coated magnetic beads were washed with wash buffer and then mixed with the TnT-antibody mixture. The precipitation of proteins was permitted to occur overnight at 4 °C. The beads were collected using a magnet and the depleted lysate was removed. PrV-infected HeLa cells were lysed in Cell Lytic M containing 2X protease inhibitors; this mixture was subjected to shear forces by passing it through an 18 gauge needle. The TnT IP was incubated with the HeLa cell lysate overnight at 4 °C. The beads were collected, washed twice and then resuspended in Cell Lytic M + protease inhibitors. The proteins were separated using 7.5 % SDS-PAGE and analysed using western blots.

2.9 Confocal microscopy

HeLa cells were plated on glass coverslips and permitted to settle overnight. When required, PrV-infected HeLa cells were stained with 250 nM MitoTracker Red FM (Invitrogen, Cat. # M22425) for 4 hours prior to the preparation of cells for confocal microscopy. The cells were washed with phosphate buffered saline (PBS), fixed with 4 % paraformaldehyde and then permeabilised with permeablisation buffer (PBS containing 5 % bovine serum, 10 % sucrose and 1 % Triton X-100). The cells were then incubated with antibodies required in the experiment. Permeabilised HeLa cells were incubated with primary (1:500) and secondary (1:1000) antibodies for 90 min, respectively. Antibodies used in the preparation of cells for confocal microscopy included anti-VCAP (rabbit), IgG-biotin anti-p40, anti-dsRNA (mouse, English & Scientific Consulting, Cat. # 10020500), anti-myc (mouse), anti-ATP5B (C-20) (goat, Santa Cruz Biotechnology, Cat. # sc-16690), goat anti-rabbit Alexa Fluor (AF) 633 (Thermo Fisher, Cat. # A21070), goat anti-mouse AF 546 (Thermo Fisher, Cat. # A11003), streptavidin AF 488 (Thermo Fisher, Cat. # S11223), donkey anti-goat AF 633 (Thermo Fisher, Cat. # A21082) and donkey anti-mouse AF 546 (Thermo Fisher, Cat. # A11036). The probed cells were washed three times with permeabilization buffer, the second wash contained 1 µg/ml 4′,6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich, Cat. # D8417-1MG). The coverslips were mounted onto glass slides using DAKO fluorescent mounting medium (Diagnostech, Cat. # S302380). The samples were visualised using Zeiss LSM780 laser scanning confocal

microscope using the x63, 0.75 NA objective and the images were analysed using Zen Blue 2012 software.

2.10 Detergent treatment for the enhanced detection of PrV replication

Detergent treatment:

PrV-infected HeLa, MCF-7 and HEK293 cells were treated with detergents diluted in S buffer (130 mM sucrose, 50 mM potassium chloride, 50 mM potassium acetate, 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)), pH 7.4, at a concentration of 0.1 %. The following detergents were used: SDS (Sigma, Cat. # L3771; CAS No. 151-21-3), CTAB (Merck, Cat. # 8.14119.0100; CAS No. 57-09-0), Tween 20 (Sigma, Cat. # P9416; CAS No 9005-64-5), Digitonin (Sigma, Cat. # D141; CAS No. 11025-24-1), Saponin (Sigma, Cat. # 47036; CAS No. 8074-15-2), NP-40 substitute (Roche, Cat. # 11754599001; CAS No. 9036-19-5) and Triton X-100 (Sigma, Cat. # T8787; CAS No. 9002-93-1). The cells were treated with detergent for 15 minutes at 37 °C and then permitted to recover for 15 minutes at 37 °C in fresh growth medium prior to preparation for confocal microscopy, cell viability assay, RT-PCR, western blot or transmission electron microscopy (TEM) analysis.

Cell viability assay:

Subsequent to the detergent treatment, the viability of the HeLa cells was measured using a resazurin-based *in vitro* toxicology assay kit (Sigma, Cat. # R6892). This assay uses the dye resazurin, which is dark blue in colour when in the oxidized form. When resazurin was added to metabolically active cells, it was reduced from the blue form to the fluorescent red form. The level of dye conversation was used to quantify the change in viability. Cell viability was evaluated after 12 h on the SpectraMax M series microplate reader (Molecular Devices, San Jose, CA, USA). Samples were excited at 560 nm and the change in fluorescence was measured at 590 nm.

RT-PCR:

PrV-infected HeLa cells were plated in a 6 well plate and transfected, using Xfect (as described above), with 6 µg pEGFP-N1. The cells were treated with either S buffer, Triton X-100 or Tween-20 (as described above). Total RNA was extracted using RNA Shield™ Purification Kit (Zymo Research, Cat. # R1100) and random hexamers that were provided in the kit were used to generate cDNA. The cDNA was PCR-amplified for 30 cycles with AccuPOL DNA Polymerase (Ampliqon, Cat. # A211102) using the following primers: JRS79 (CGA GGT TAC CAC AAC CTG C) and JRS80 (GAT GCC CTC GGC AAC C); EGFP_F (AAG GGC GAG GAG CTG TTC ACC G) and EGFP_R (CGG CGG CGG TCA CGA ACT C); Enolase_F (ACT GCC TGC TCC TGA AGG TC) and Enolase_R (ATA ATG ATT AGA TCA AGG TT). The PCR

conditions specified by the manufacturer were followed using an annealing temperature of 55 °C.

SDS-PAGE and western blot:

The detergent-treated, pEGFP-N1 transfected HeLa cells were separated on a 10 % SDS acrylamide gel and analysed using western blot analysis. The membranes were probed with the following antibodies: IgG-biotin anti-p40 with streptavidin HRP and anti-GFP (mouse) with goat anti-mouse HRP.

Transmission electron microscopy (TEM):

PrV-infected HeLa cells were treated with S buffer containing 0.1 % Tween 20 or Triton X-100 as described above. The cells were washed in PBS, fixed in 2.5 % glutaraldehyde in 0.1 M phosphate buffer for 2 hours and then washed twice with 0.1 M phosphate buffer. The cells were stained with 1 % osmium tetroxide in 0.1 M phosphate buffer for 90 minutes, then washed twice with 0.1 M phosphate buffer. The stained samples were dehydrated in increasing concentrations of ethanol (30 %, 50 %, 70 %, 80 %, 90 % and absolute) for 15 minutes. The ethanol was replaced with propylene oxide in two consecutive 30 minute treatments. The propylene oxide was replaced with increasing resin:propylene oxide containing solutions (25 %, 50 %, 75 % and 100 %). The cells were embedded in absolute resin during a 36 hour incubation at 60 °C. Sections of the embedded cells were cut on an RMC products microtome and these 100 nM sections were stained with 5 % uranyl acetate for 20 minutes. The stained sections were visualised on a Zeiss Libra 120.

Labelling with quantum dots:

The sections prepared for TEM were blocked with 1 % glycine for 15 minutes and then with 4 % BSA in TBST for a further 30 minutes. The antibodies (IgG-biotin anti-p40 and AF 488 Streptavidin, 10 nm colloidal gold conjugate – Cat. # A32361) were diluted 3:100. The sections were incubated with the antibodies overnight at 4 °C in a saturation chamber. After which, the sections were rinsed in 4 % BSA in TBST for 10 minutes, twice and then with TBST for 10 minutes, twice. Finally, the sections were washed with 3 ml of dddH_2O and permitted to dry overnight before visualisation.

2.11 Mass spectrometry and data analysis

Sample preparation:

Proteins were immunoprecipitated from HeLa cell lysates using anti-PrV antibodies (anti-p17, anti-p113, anti-p40, anti-p104 and anti-VCAP). The immunoprecipitated proteins were permitted to electrophorese into a 10 % acrylamide gel. The gel was stained with Coomassie

blue and the whole band, containing the total cell lysatewas excised and cut into 1 mm X 1 mm size fragments. The gel fragments were dehydrated using acetonitrile and then vacuum dried using the CentriVap (Labcono). The dehydrated gel pieces were processed by Dr Mare Vlok at the Central Analytical Facilities (CAF; Stellenbosch, South Africa). Dr Vlok performed LC-MS/MS using a Fusion Mass Spectrometer and a 60min gradient. The PrV peptide sequences identified were summarised and included in Appendix A7.

Data analysis:

Dr Vlok provided raw MS data files as well as processed data (Appendix A1). With respect to processing, Dr Vlok imported the data onto the Proteome Discover v1.4 and analysed it using both Sequest and Amanda algorithms. The sequences were probed against the Uniprot *Homo sapiens* reviewed database, the Uniprot viral database and cRAP contaminant database. Peptide validation was performed using the Target-Decoy PSM validator node. The resulting files were imported into Scaffold 1.4.4 and validated and identified peptides were analysed using the PeptideProphet and ProteinProphet algorithms in Scaffold. The raw data were also analysed using denovogui or searchgui and peptideshaker using the same parameters specified by Dr Vlok. PrV protein sequences as well as the sequences for proteins predicted by GeneMarkS and Softberry were used to create a library to permit the identification of any PrV sequences in the mass spectrometry data.

Chapter 3: Identification of a model system for the study of virus replication *in vivo*: Providence virus as a model system

3.1 Introduction

Providence virus

Providence virus (PrV) is a positive-sense, single-stranded RNA virus (Pringle *et al*., 2003). The viral particles are non-enveloped and have a *T* = 4 icosahedral symmetry which resulted in the inclusion of PrV into the *Tetraviridae* family (Pringle *et al*., 2003). PrV was first isolated from a *Helicoverpa zea* midgut (MG8) cell line that was being used as a host to study Pariacoto virus (PaV; *Nodaviridae*). This was interesting because prior to this discovery no tetraviruses had been identified that were capable of replication in tissue culture cell lines (Pringle *et al*., 2003). Preparation of RNA from MG8 cells infected with PaV resulted in the detection of the PaV genomic RNAs as well as two contaminating bands of 6.4 and 2.5 kb. The MG8 cell line was persistently infected with PrV and it is important to note that despite this persistent infection, PaV was able to infect the MG8 cells and to establish PaV replication. Pringle *et al*. (2003) also identified the two structural proteins of 60 and 7.4 kDa; these proteins are responsible for the PrV capsid. The viral capsid proteins were most similar to the structural proteins of the omegatetraviruses, and particularly *Helicoverpa armigera* stunt virus (HaSV). Omegatetraviruses have a bipartite genome where RNA 1 produces the RNA dependent RNA polymerase and RNA 2 produces the structural proteins (Hanzlik and Gordon, 1997). Betatetraviruses such as *Nudaurelia capensis* beta virus (NβV) have monopartite genomes and produce the structural proteins via subgenomic RNA (Hanzilk and Gordon, 1997). PrV, like the betatetraviruses, has a monopartite genome of 6.4 kb and produces the structural proteins from a 2.5 kb subgenomic RNA (sgRNA); both RNAs are packaged into the PrV capsid (Pringle *et al*., 2003). The discovery of PrV, and particularly its ability to replicate in cell culture, provided a unique opportunity to study the replication biology of Tetraviruses (Pringle *et al*., 2003).

Tetraviruses were reported to have a stringent host range; they were thought to exclusively infect and replicate in hosts belonging to the Order Lepidoptera (Dorrington *et al*., 2011; Hanzlik and Gordon., 1997; Gordon *et al*., 1999). Pringle *et al*. (2003) demonstrated that PrV was capable of establishing infection and replicating in a fat body cell line derived from *H. zea* (FB33) and an embryonic cell line derived from *Spodoptera exigua* larve (Se-1) in addition to the MG8 cell line. The host range of PrV has been dramatically widened in recent studies by our research group to include mammalian cell culture lines and cowpea plants (Jiwaji *et al*., 2019; Jiwaji *et al*., 2016). PrV can infect and replicate in human cervical cancer (HeLa), breast

cancer (MCF-7) and embryonic kidney (HEK293) cell culture lines (Jiwaji *et al*., 2019; Jiwaji *et al*., 2016, unpublished data). In addition, PrV purified from MG8 cells could infect and replicate in *Vigna unguicalata* (cowpea plants; Jiwaji *et al*., 2019). Virus isolated from cowpea plants was capable of infecting and replicating in both HeLa and MCF-7 cell lines (Jiwaji *et al*., 2019). It is interesting to note that members of the *Nodaviridae* can replicate in similar hosts however they are not capable of establishing infection from virus particles. In those experiments, researchers added genetic material in the form of RNA or DNA to establish Nodavirus replication in yeast, plant and animal cells (Ball and Johnson, 1999; Price *et al*., 2000; Miller *et al*., 2001; Selling *et al*., 1990). PrV is therefore unique as it this the first virus to be shown capable of crossing the Kingdom boundaries between plants, invertebrates and vertebrates (Jiwaji *et al*., 2019).

In 2016, Kemenesi *et al*. published the full sequence of PrV; this virus had been isolated from the guano of a female western barbastelle bat (*Barbastella barbastellus*) bat in Hungary. The source of the PrV is most interesting. The western barbastelle bat is insectivorous and it is possible that the bat may have consumed a PrV-infected insect. More interestingly, it is possible that the bat itself was infected. Speculation about the latter possibility is tantalizing because of the implications of such a discovery. The bats are reported as hosts of numerous medically relevant viruses including the families *Rhabdoviridae* (Rabies virus), *Filoviridae* (Ebola and Marburg viruses), *Coronaviridae* (SARS-CoV and MERS-CoV), *Paramyxoviridae* (Nipah and Hendra viruses), *Orthomyxoviridae* (Influenza), *Bunyaviridae* (Hantaan virus) and *Reoviridae* (mammalian othroreovirus) (Allocati *et al.,* 2016). Consideration of a PrV infection in a bat is therefore of great interest.

PrV, when added to cell-free transcription and translation (TnT) reactions, was capable of establishing virus replication (Jiwaji *et al*., 2016). This was evidenced by the generation of the full length and the sgRNA and also by the production of the PrV capsid protein. The process was most efficient in cell-free systems derived from insect cells and to a lesser degree those from plant cells. The least efficient was the mammalian system derived from rabbit reticulocytes (Jiwaji *et al*., 2016). This observation highlights the robustness of the PrV replication system and motivates for the study of this virus.

Genome analysis

The PrV genome is 6155 nt in length and does not have a 5’ cap or a 3’ poly-A tail (Walter *et al*., 2010). The genome is reported to encode three open reading frames (ORFs; Figure 3.1). The first ORF (45 – 3707 nt) produces a protein of 130 kDa (p130); there are no protein homologs reported to date and the function of this protein is still unknown. Nakayinga (2019)

recently published a bioinformatic analysis that suggests that p130 may function as an RNA chaperone protein due to the presence of an arginine-rich amino acid sequence, however, there is no biological data to support this hypothesis, at this time. A 2A-like processing site is present within the p130 sequence and is predicted to be functional (Luke *et al*., 2008). This would result in translation products of 17 and 113 kDa, respectively (Walter *et al*., 2010). The second ORF (1027 – 3775 nt) produces the RNA dependent RNA polymerase (RdRp), p104, as well as replication accessory protein, p40 in the same frame. ORF2 contains a type 1 read through stop (RTS) codon (UAG-CAA-CUA) which permits the production of p40 and p104 at a ratio of approximately 10 to 1 (Walter *et al*., 2010). Type 1 RTS have no secondary RNA structures that result in termination. Instead there is a 6 nucleotide sequence (UAG-CAR-YYA) following the UAG codon where R represents a purine and Y represents a pyrimidine (Skuzeski *et al*., 1991). Type 1 RTS codons are most commonly associated with plant viruses, for example the tobacco mosaic virus (Firth and Brierley, 2012; Harrell *et al*., 2002), and it has been identified here in the sequence of PrV. The third ORF (3779 – 6044 nt) encodes the capsid precursor protein, p81. The capsid protein is produced via the production of a sgRNA (Pringle *et al*., 2003; Walter *et al*., 2010). ORF3 contains two 2A-like processing sites which result in translated products of 7, 8 and 68 kDa. The immature capsid protein, p68, is autoproteolytically cleaved to form the mature capsid proteins of 60 kDa and 7.4 kDa, described as the β and γ peptides, respectively (Pringle *et al*., 2003; Walter *et al*., 2010). Walter *et al*. (2010) confirmed the expression of predicted p40, p104 and the capsid proteins (68 and 60 kDa) using *in vitro* TnT reactions. No protein that could represent p130 was detected, however this was attributed to the lack of suitable antibodies.

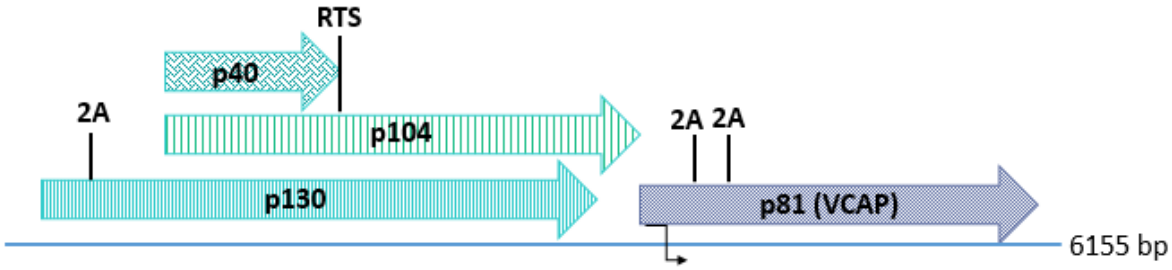


Figure 3.1: Schematic diagram of the PrV genome. There are three recognised ORFs in the PrV genomes: ORF1 – encodes p130; ORF2 – encodes PrV replicase p104 as well as the replication accessory protein p40 via a read through stop (RTS) codon; ORF3 – encodes the capsid protein (VCAP) which is translated from the sgRNA (indicated by the arrow). p130 and VCAP both contain 2A-like processing sites, indicated by the notation 2A.

Phylogenetic analysis of the RdRp resulted in the PrV p104 aligning with the carmo-like replicase supergroup rather than those of the insect Tetraviruses. The PrV p104 was more closely related to the polymerases belonging the *Tombusviridae* and the Umbraviruses

(Walter *et al*., 2010). For this reason, with the *Tetraviridae* were reclassified, PrV was assigned to its own virus family, *Carmotetraviridae*, of which it is the sole member (Short *et al*., 2013; Walter *et al*., 2010).

Replication biology

Nakayinga (2013) studied the PrV p104 and p40 proteins; IP reactions were used to demonstrate an interaction between the p104 replicase and p40 replication accessory proteins. Immunofluoresence analysis using confocal microscopy showed that fluorescence attributed to p104 and p40 co-localized with that of double-stranded RNA (dsRNA) in MG8 cells (Nakayinga, 2013; Short *et al*., 2013). In addition, the fluorescence formed distinct punctate structures within the cytoplasm of the insect cells (Nakayinga, 2013; Short *et al*., 2013). These observations led the authors to propose that p104 and p40 were part of the replication complex in the MG8 cells. Short *et al*. (2013) used confocal immunofluorescence microscopy to show that the fluorescence attributed to the PrV replication complexes co-localised with the Golgi apparatus and secretory vesicles in insect cells. Short *et al*. (2013) treated PrV-infected cells with detergents and used western blot analysis to determine the location of the replication complexes. The authors, based on their data, concluded that the PrV replication complexes were associated with detergent resistant membranes in insect cells.

Classification of PrV

PrV is an interesting virus as it shares features with viruses belonging to two very different families, the *Tetraviridae* and the *Tombusviridae*. PrV shares *T = 4* icosahedral capsid symmetry and genome organisation (monopartite and sgRNA) with members of the *Tetraviridae* family. The PrV replicase however, was found to be most similar to viruses belonging to the *Tombusviridae* family.

Tetraviridae

The *Tetraviridae* are a family of insect viruses, with *T* = 4 icosahedral capsids that encapsidate single-stranded, positive-sense RNA (Hanzlik and Gordon, 1997). The particles are non-enveloped and range in size from 35 to 41 nm. They have a narrow host range restricted to insects of the Order Lepidoptera (Hanzlik and Gordon, 1997; Gordon *et al*., 1999). The family is divided into two genera based on the organisation of their genomes. Betatetraviruses have monopartite genomes and produce a sgRNA for the translation of their capsid protein. In most cases, the genomic and sgRNAs are packaged in the viral capsid (Hanzlik and Gordon, 1997; Gordon *et al*., 1999). In contrast, omegatetraviruses have bipartite genomes, both of which are packaged in the virion (Hanzlik and Gordon, 1997; Short *et al*., 2010). There is little

information about the replication biology of these viruses because of a lack of available *in vitro* systems and possibly, because of a perceived lack of economic importance.

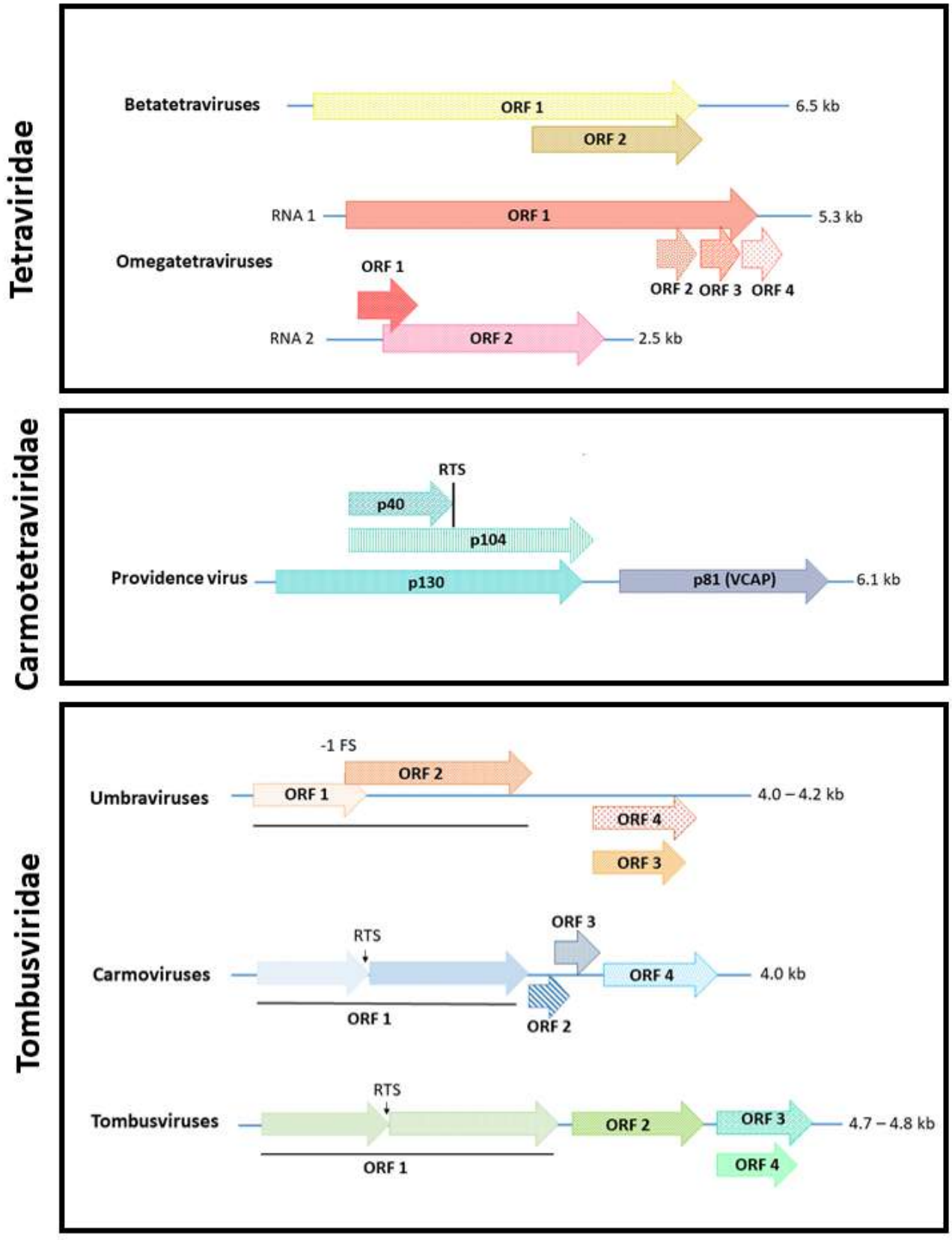


Figure 3.2: Schematic representation of the ORFs present in viruses belonging to the *Tetraviridae, Carmotetraviridae* and *Tombusviridae* families. Omegatetraviruses have bipartite genomes with RNA 1 encoding the replicase (ORF1) and three smaller proteins of unknown functions (ORFs 2 - 4). RNA 2 encodes the capsid precursor protein (ORF2) and p17 (ORF1), which is predicted to bind RNA (Hanzlik and Gordon, 1997; Gordon *et al*., 1999; Mendes *et al*., 2015). Betatetraviruses are monopartite with two ORFs detected. ORF1 encodes the replicase while ORF2 produces the capsid protein via translation of a sgRNA (Hanzlik and Gordon, 1997; Gordon *et al*., 1999). Tombusviruses and Carmoviruses both have monopartite genomes and ORF1 contains a RTS which permits the expression of both the replicase and a smaller accessory protein. ORFs 2 - 4 are produced via translation from sgRNAs. In the case of the tombusviruses, ORF2 encodes the capsid protein and is translated from sgRNA 1; ORFs 3 and 4 which encodes the movement protein and suppressor protein, respectively, are translated from sgRNA 2 (Russo *et al*., 1994; Li *et al*., 2009). Carmoviruses express ORFs 2 and 3 (movement proteins) from sgRNA 1 and the capsid protein (ORF4) from sgRNA 2 (Marcos *et al*., 1999). Umbraviruses express the full length replicase (ORF1 + ORF2) using a -1 frame shift (FS). ORFs 3 and 4 are produced from the sgRNA and result in the expression of a viral RNA stabilizing protein and the movement protein, respectively (Taliansky and Robinson, 2003). Providence virus genome is described in detail above.

Omegatetraviruses

There are three species that belong to this genus of tetraviruses, *Dendrolimus punctatus* virus (DpTV), HaSV and *Nudarelia capensis* ω virus (NωV), which is the type species. As mentioned above, these viruses have bipartite genomes (Figure 3.2). RNA 1 has four ORFs. ORF1 encodes the replicase (p187 for HaSV) which has three domains, methyltransferase, helicase and RdRp (Hanzlik and Gordon, 1997; Short *et al*., 2010). These domains result in the classification of the omegatetravirus replicase in the alpha-like supergroup (Short *et al*., 2010). At the 3' end of RNA 1 are three small, putative ORFs: ORF2 – p11, ORF3 – p15 and ORF – p8; there is no indication at this time that these proteins are expressed (Short *et al*., 2010). RNA 2 encodes the capsid precursor protein (ORF2), p71 in HaSV, which undergoes autoproteolytic cleavage to form β and γ peptides (Short *et al*., 2010). ORF1 encodes p17 at the 5' terminus of RNA 2 and this protein has been shown to promote the packaging of HaSV RNA 2 (Mendes *et al*., 2015). Note that PrV also has a predicted p17 protein at the 5' terminus but there is no homology between the p17 protein of PrV and that of HaSV.

Little is known about omegatetravirus replication biology. These viruses are unable to replicate in established cell culture lines (Hanzilk and Gordon, 1997; Short *et al*., 2010). Despite this challenge, Short *et al*. (2010) created EGFP-HaSV_replicase expressing constructs to study HaSV replication *in vivo.* Short *et al.* (2010) found that the replicase protein was associated with cytosolic structures that were likely to be membranes, derived from the alternative endocytic pathway. In addition, the membranes associated with the HaSV replicase were detergent resistant. The endocytic pathway is known to be associated with detergent resistant

lipid rafts which led Short and Dorrington (2012) to speculate about the nature of the detergent resistant membranes that associated with the HaSV replicase (Short and Dorrington, 2012). It is interesting to note that the PrV replication accessory protein, p40 is also associated with detergent resistant membranes (Short and Dorrington, 2012; Short *et al*., 2013).

Betatetraviruses

NβV is the type species of the genus, *Betatetravirus*. These viruses have monopartite genomes of 6.5 kb and produce a sgRNA of 2.5 kb (Hanzilk and Gordon, 1997; Gordon *et al*., 1999). Two ORFs are encoded by the genomes (Figure 3.2); the first produces the replicase of approximately 215 kDa (Gordon *et al*., 1999). The second ORF encodes the 70 kDa capsid precursor protein which is produced via the translation of a sgRNA (Gordon *et al*., 1999). Like the omegatetravirus replicases, the NβV replicase also clusters with the alpha-like replicase supergroup as it contains the methyltransferase, helicase and RdRp domains (Gordon *et al*., 1999). With respect to genomes organization, PrV is most similar to the betatetraviruses.

Tombusviridae

Viruses that are part of the *Tombusviridae* are positive-sense, single-stranded RNA viruses with a host range that is restricted to dicotyledonous plants (Russo *et al*., 1994). The virions are approximately 30 nm in diameter with a *T* = 3 icosahedral symmetry, and the viral particles encapsidate the monopartite genomes (Russo *et al.*, 1994). The *Tombusviridae* family is divided into two genera, *Tombusvirus* and *Carmovirus*.

Tombusviruses

The genome length of the tombusviruses range between 4.7 and 4.8 kb and encode 5 proteins (Figure 3.2) (Li *et al*., 2009; Russo *et al*., 1994). Tomato bushy stunt virus (TBSV) is the type species of the genus *Tombusvirus*. The first ORF produces the replicase protein p92 as well as p33 via a type III RTS mechanism (Firth *et al*., 2011). Type III RTS generally involve UAG stop codon with a 3' adjacent G or a purine-rich 8 nucleotide sequence with an RNA structure (Firth and Brierley, 2012). p33 is the replication co-factor that is responsible for recruitment of p92 and the viral RNA to the peroxisome which is the site of replication for TBSV (Nagy and Pogany, 2008). The RTS provides a mechanism to regulate the production of the replicase; p33 is produced at levels approximately 20 fold higher than p92 is produced (Nagy and Pogany, 2008). PrV makes use of a similar control mechanism, the type I RTS regulates the production of p40 to p104 at an approximately 10 to 1 ratio (Firth and Brierley, 2012; Walter *et al*., 2010). ORFs 2, 3 and 4 encode proteins that are translated from sgRNAs. sgRNA 1 is 2.2 kb and the translation of ORF2 results in the production of the 41 kDa capsid protein (Russo *et al*., 1994; White and Nagy, 2004). sgRNA 2 is 0.9 kb and encodes two proteins:

p22 (ORF3) which is the cell-to-cell movement protein and p19 (ORF4) which is the suppressor of virus-induced gene silencing (Hull, 2014; Nagy and Pogany, 2008; White and Nagy, 2004).

Carmoviruses

Carnation mottle virus (CaMV) is the type species of the genus *Carmovirus*. These viruses have a 4.0 kb genome that encodes 4 ORFs that result in the production of at least 5 proteins (Figure 3.2; Marcos *et al*., 1999). The first ORF produces an 88 kDa replicase protein as well as a 28 kDa protein by using the same type III RTS mechanism as the tombusviruses (Firth *et al*., 2011; Marcos *et al*., 1999). The proteins produced from ORFs 2 to 4 are translated from two sgRNA, like the tombusviruses. sgRNA 1 is 1.7 kb and appears to be bicistronic, resulting in two smaller proteins, p7 and p9 that are produced by an unknown mechanism (Garcia-Castillo *et al*., 2003; Marcos *et al*., 1999; Russo *et al*., 1994). p7 and p9 are involved in cell-to-cell movement, p7 has RNA binding potential (Garcia-Castillo *et al*., 2003; Marcos *et al*., 1999) and p9 was found to be an intrinsic membrane protein that inserted into the ER membrane (Vilar *et al*., 2002). sgRNA 2 is 1.5 kb and directs the translation of the 38 kDa capsid protein (Marcos *et al*., 1999).

Umbravirus

Umbraviruses are unconventional plant viruses as they do not encode a gene for a capsid protein. Instead, umbraviruses depend on an assistor virus, usually from the *Luteovirus* family, to package the genomic RNA. The *Umbravirus*, in this hijacked protein coat, is transferred between plants by aphids (Taliansky and Robinson, 2003). Umbraviruses are positive-sense, single-stranded RNA viruses with relatively short genomes, between 4.0 and 4.2 kb. The assistor virus is not required for umbravirus replication; however, some species of umbraviruses do require the assistance of satellite RNAs. For example, the groundnut rosette virus (GRV), which is endemic to sub-Saharan Africa, needs to be associated with its satellite RNA to cause symptomatic groundnut rosette disease (Murant and Kumar, 1990).

The genome organisation of umbraviruses is different to PrV. There are four ORFs (Figure 3.2) with the products of ORFs 3 and 4 produced from separate sgRNAs. ORF1 produces a putative 31 – 37 kDa protein but immediately before the stop codon is a 7 nucleotide sequence which is associated with frame shifting. This would result in a -1 frame shift to produce a protein of 94 - 98 kDa (ORFs 1 + 2). This protein is the RdRp which is closely related to the RdRp of the *Tombusviridae* (Ryabov *et al*., 1998; Taliansky *et al*., 2003; Taliansky and Robinson, 2003). ORF3 (26 – 29 kDa) has been shown to stabilize the viral RNA and to form filamentous ribonucleoprotein complexes in the cytoplasm of plant cells (Taliansky *et al*.,

2003). These complexes could serve to protect the viral RNA from host cell detection and may also be involved in long distance travel through the phloem in plants (Taliansky *et al*., 2003). ORF4 (27 -29 kDa) is the movement protein which aids cell to cell movement via the plasmodesmata, negating the need for a capsid protein (Ryabov *et al*., 1998; Taliansky and Robinson, 2003). This serves as a neat way to spread virus throughout the plant.

Reuter *et al*. (2019) have recently reported a novel tombus-/carmotetravirus-like RNA virus (H14-hedgehog/2015/HUN) that isolated from a wild northern white-breasted hedgehog (*Erinaceus roumanicus*). The authors found five potential ORFs on the viral genome and only two ORFs have been assigned function. ORF1-RT (RdRp) was found to have a 96 % amino acid similarity to the RdRp of PrV, and also contains a type 1 RTS codon. The second ORF produces the capsid protein, potentially through the translation of sgRNA but no 2A-like processing sites were identified (Reuter *et al*., 2019). Due to the similarities between the H14-hedgehog/2015/HUN and the plant viruses, the authors attempted to demonstrate the infection of three plant species with *in vitro* transcripts of the virus genome. They found that the viral RNA was not detected 10 hours post infection, which suggests that no H14-hedgehog/2015/HUN replication occurred within the plant.

One of the most interesting findings was that H14-hedgehog/2015/HUN was identified in the faecal samples, in the blood as well as in the muscle tissue (Reuter *et al*., 2019). This suggests that a virus that shares high levels of sequence similarity with PrV, is able to infect and replicate within a mammalian host. The authors reported no indication of illness in the hedgehogs (Reuter *et al*., 2019); it is therefore interesting to note that cells persistently infected with PrV also show no signs of cytopathic effect (Pringle *et al*., 2003; observations by author).

Summary

The $T = 4$ capsid structure of PrV, the similarity of the structural proteins and the monopartite genome align with the *Tetraviridae*. In contrast, the replicase and the presence of a type 1 RTS, which are utilised by plant viruses, align with the *Tombusviridae* and the Umbraviruses (Firth and Brierley, 2012). For this reason, PrV has been classified as the sole member of the *Carmotetraviridae* family.

Motivation for the use of PrV as a model system

PrV has a demonstrated broad host range and can replicate in insect and mammalian cells as well as in plants. The fact that PrV can replicate *in vivo* provides an opportunity to study the replication biology of PrV, a persistent virus with some very interesting characteristics. More

importantly, the availability of the PrV-based replication system provides an opportunity to develop tools to study viruses that are of interest when the identity of the virus, its host range and its replication mechanisms are unknown. In this study, we propose to use PrV as a model system to develop tools that can be used to evaluate emerging broad-spectrum zoonotic viruses. For this reason, we will consider PrV to be an unknown virus and we will analyse the PrV genome using bioinformatics tools. We will predict ORFs that may be present and then use PrV-specific antibodies to determine whether the identified proteins are expressed and detected. We will use this information to study PrV replication in mammalian cells, namely HeLa cells.

3.2 Results

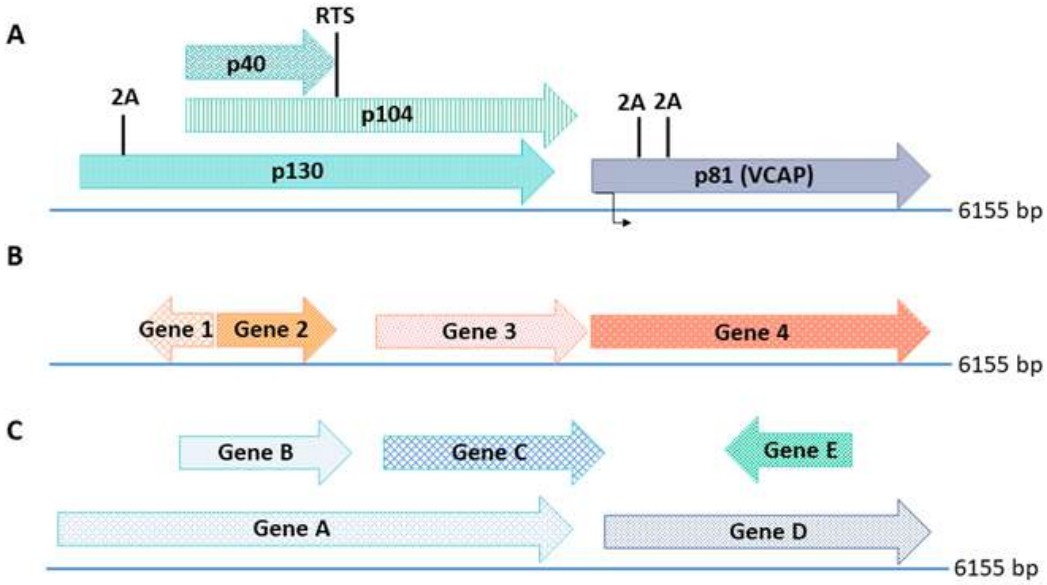
3.2.1. Bioinformatic analysis of the PrV genome

Walter *et al*. (2010) performed the first detailed analysis on the genome organization of PrV and validated this analysis by studying protein production *in vitro* in insect systems as well as *in vivo* in insect cell lines. The authors identified three ORFs (Figure 3.3A); ORF1 was thought to produce p130 and to a have a 2A-like processing site that would result in the production of two proteins, p17 and p113. ORF2 produced the RdRp p104 as well as the replication accessory protein p40 via a mechanism utilizing a RTS. The capsid protein was generated from a sgRNA. As these analyses were performed for and in insect-based systems, we hypothesized that PrV may express different proteins in mammalian systems. With this in mind, the PrV genome sequence was analysed using two ORF prediction sites, GeneMarkS and SoftBerry (Figure 3.3).

GeneMarkS predicts the ORFs of prokaryote sequences (Besemer *et al*., 2001) and SoftBerry specifically predicts virus ORFs. Four ORFs were predicted by GeneMarkS (Figure 3.3B and 3.3D). The first ORF (Gene 1) produced a protein of 8.4 kDa on the negative sense strand. Gene 2 is in the same frame as p40/p104 and utilizes the same stop codon but starts after the start site predicted by Walter *et al*. (2010) and produces a protein of 34.2 kDa. Gene 3 is also in the same frame as p40/p104 and shares a stop codon with p104 but begins after the predicted RTS sequence on the PrV genome, resulting in the production of a 59.9 kDa protein. The final ORF (Gene 4) matches the sequence reported for p81 in its entirety (Walter *et al*., 2010).

SoftBerry predicted genes in positions almost identical to the PrV genome published by Walter *et al*. (2010). Genes A, B and D match the sequence of p130, p40 and p81. Gene B is the exception; it aligns with Gene 3 predicted by GeneMarkS and predicts a protein that is 59.9 kDa in size (Figure 3.3C and 3.3D). This suggests that both GeneMarkS and Softberry do not

recognise RTS sequences, which is particularly interesting when referring to SoftBerry which is designed to detect viral sequences. The final ORF predicted by Softberry, Gene E, produces a 20.3 kDa protein from the negative-sense RNA strand. When the Gene E putative protein sequence was compared to the proteins presents in the BLAST database, no proteins with sequence similarity were identified. The same result was achieved when the Gene 1 protein sequence was submitted to the BLAST database.



D

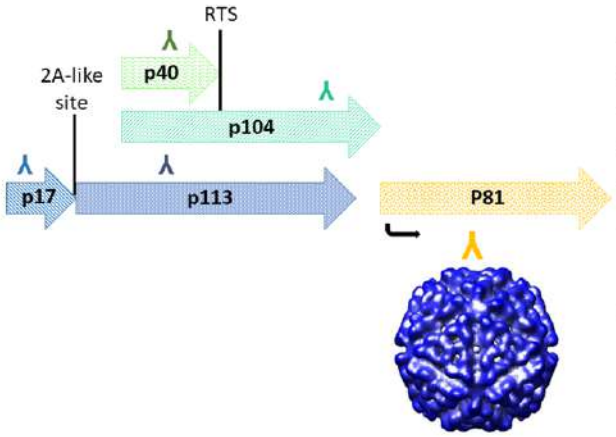
PrV genome (published by Walter *et al.*, 2010)				GeneMarkS predicted				SoftBerry predicted			
Gene	ORF	Length (nt)	Protein (kDa)	Gene	ORF	Length (nt)	Protein (kDa)	Gene	ORF	Length (nt)	Protein (kDa)
p130	+ 3rd	45-3707	130.7	1	- 1st	974-1205	8.4	A	+ 3rd	45-3704	130.7
p40	+ 1st	1027-2106	40.5	2	+ 1st	1207-2106	34.2	B	+ 1st	1027-2103	40.5
p104	+ 1st	1027-3774	104	3	+ 1st	2194-3774	59.9	C	+ 1st	2194-3771	59.9
p81 (VCAP)	+ 2nd	3779-6043	81.4	4	+2nd	3779-6043	81.4	D	+ 2nd	3779-6040	81.4
								E	- 3rd	4885-5466	20.3

Figure 3.3: Comparison of the published PrV genes to gene identified by GeneMarkS and Softberry prediction sites. (A) The PrV genome (NC_014126.1) was published by Walter *et al*. (2010). The sequence for the PrV genome was analysed using **(B)** GeneMarkS

and **(C)** SoftBerry. GeneMarkS predicted four ORFs, Gene 1-4, with Gene 1 being translated from the negative-sense RNA. SoftBerry predicted five ORFs, Gene A-E, with Gene E being translated from the negative-sense RNA strand. **(D)** A summary of the location of the ORFs published by Walter *et al*. (2010) or identified by GeneMarkS and SoftBerry. Here, the ORFs identified are specified as being in the 1st, 2nd or 3rd frames, and either on the forward (+) or reverse strand (-) of the RNA. The images presented above were created in SnapGene**.**

3.2.2 The available PrV-specific antibodies

The Dorrington research group (Appendix A5) have generated antibodies to the proteins identified by Walter *et al.* (2010). These include anti-VCAP and IgG-biotin anti-p40 (Walter, 2008), anti-p104 (Nakayinga, 2013), anti-p17 and anti-p113 (Mpho Peter, research in progress). The antibody recognition sites are depicted on the schematic of the PrV genome in Figure 3.4. Anti-p17 and anti-p113 were generated to regions of p130 and would interact with the full protein. In addition, the two antibodies would also evaluate whether the predicted 2A-like site present in the p130 ORF was functional. The anti-p40 antibody was raised in rabbits; when tested this antibody showed high levels of non-specific interactions with cellular proteins. To improve specificity, the rabbit-specific epitopes were masked and the antibody was biotinylated. This IgG-biotin anti-p40 antibody demonstrated higher specificity than the original anti-p40 (rabbit) antibody. This antibody will detect both p40 and p104 proteins in the cell because the region to which it is targeted is present in both proteins. In contrast, the anti-p104 antibody was generated to recognise the C-terminal of p104 permitting differentiation between p40 and p104. The anti-VCAP antibody was generated against purified PrV particles and therefore recognises the full virus particle.



Antibody	Binding site (amino acid)
P17	1-130 (F 3rd ORF)
P113	595-611 (F 3rd ORF)
P40	248-331 (F 1st ORF)
P104	889-902 (F 1st ORF)
VCAP	Epitope recognition based on viral particle conformation

Figure 3.4: The recognition sites of PrV-specific antibodies used in this project. A schematic diagram shows the recognition sites of the PrV antibodies. The table on the right lists the position of the antibody recognition sites in amino acids residues, the strand on which the recognition site is based (F) and the frame on which the ORF is present. The image of the PrV particle was acquired from The Scripps Institute Cyro-EM Database

(http://viperdb.scripps.edu/emdb/html/em_info_page.php?vipPDB=em_1prv; accessed 3/12/2018).

These PrV specific antibodies were used to detect PrV proteins in PrV-infected HeLa cell lysates that were separated using SDS-PAGE and subjected to western blot analysis (Figure 3.5).

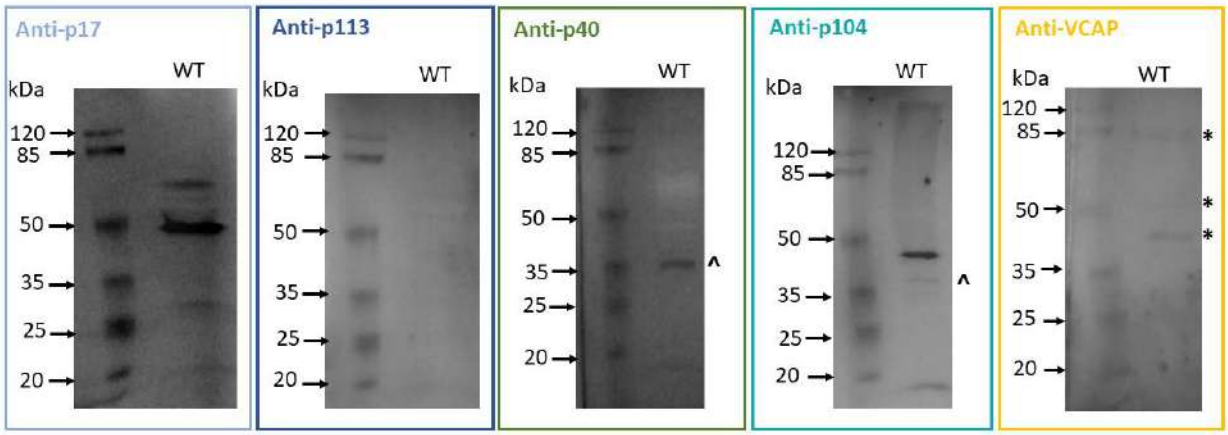


Figure 3.5: Western analysis of PrV proteins using PrV-specific antibodies. PrV-infected HeLa cells were separated using 10 % SDS-PAGE, transferred onto PVDF membranes and probed with the following antibodies: anti-p17 (rabbit), anti-p113 (mouse), IgG-biotin anti-p40, anti-p104 (rabbit) and anti-VCAP (rabbit). The black ^ indicates p40. The black * indicate bands present on the anti-VCAP membrane that were more visible on the original image. WT – wild-type, PrV-infected HeLa cells.

When probed with the anti-p17 antibody, three bands of approximately 60, 50 and 30 kDa were detected but no definitive bands that could represent p130 or p17 were detected at 130 kDa or below 20 kDa (Figure 3.5). When probed with anti-p113, no bands were detected in the HeLa cell lysates. When probed with IgG-biotin anti-p40 antibodies, a clear band was present at approximately 40 kDa. There was a faint band present at 40 kDa and a prominent band present at approximately 45 kDa when the membrane was probed with anti-p104, but no p104 was detected in the HeLa cell lysate. This is not surprising, Walter *et al*. (2010) reported that the p104:p40 are produced at a 1:10 ratio; the levels of p40 indicate that it would be unlikely to detect p104. It is interesting to note that the p104 antibody, which specifically targets the C-terminal of the p104 protein, is clearly recognising a protein of approximately 45 kDa and showing some interaction towards a protein that is 40 kDa in size. The p104 protein used to generate the anti-p104 antibody was generated towards a peptide fragment so these 45 and 40 kDa proteins are unlikely to be p40. It is more likely that these proteins are cellular proteins that are detected through non-specific interactions. The anti-VCAP antibody detected three proteins of approximately 80, 50 and 45 kDa in the HeLa cell lysates (Figure 3.5, highlighted by the black *). None of these bands are at the sizes expected for PrV VCAP,

which would be 68 kDa for the immature capsid proteins and p60 for the mature capsid proteins. While 81 kDa would represent the full length product of the sgRNA, this protein has not been detected before this experiment. These bands in the anti-VCAP western analysis may also be non-specific detection of cellular proteins.

We hypothesized that the levels of the PrV proteins in the HeLa cells were much lower than those present in the insect cells. If this was the case, we may not have detected the proteins of interest because of the low levels within the cells. For this reason, we performed IP reactions using the PrV-specific antibodies on PrV-infected HeLa cell lysates. These immunoprecipitated proteins were separated using SDS-PAGE and subjected to western blot analysis (Figures 3.6).

After IP, not surprisingly, more bands were detected on the western blots. First, there were one or two bands, present at approximately 60 and 30 kDa present on the membranes probed with anti-PrV antibodies. These bands were attributed to the heavy and light chains of the antibody that was used in the IP (Janeway *et al*., 2001). In addition, and in general, more bands were detected during western analysis for immunoprecipitated proteins than in the analysis of the HeLa cell lysates. This is not surprising as the IP process results in the concentration of the proteins in the lysates of interest. This indicates that in HeLa cells, the levels of the PrV proteins are at levels that are much lower than those in insect cells (Nakayinga, 2013; Walter, 2008).

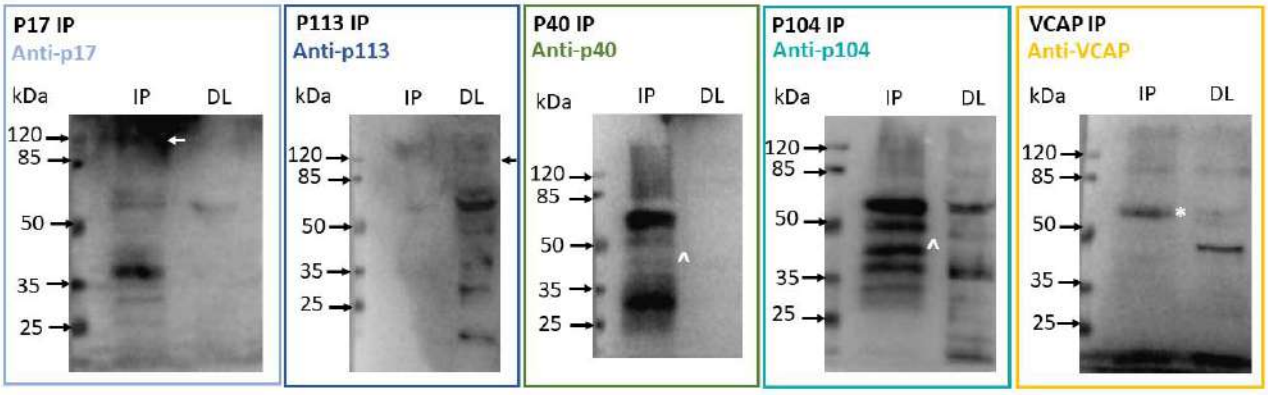


Figure 3.6: Detection of proteins immunoprecipitated from PrV-infected HeLa cells, using PrV-specific antibodies. PrV-infected HeLa lysates were incubated with anti-p17 (rabbit), anti-p113 (mouse), IgG-biotin anti-p40, anti-p104 (rabbit) or anti-VCAP (rabbit) antibodies overnight, at 4 °C with rotation. The antibody-lysate mixture was then incubated with A/G coated magnetic beads overnight at 4 °C, with rotation. The beads were collected using a magnet and the depleted lysate (DL) was removed. The membranes were probed with the same antibody that was used for the immunoprecipitation (IP). The white arrow in p17 IP highlights a band that may be p130; the black arrow in the p113 IP highlights a band that may

be p113 in the DL; the white ^ in the p40 and p104 IPs identify p40 and the white * in the VCAP IP highlights the 60 kDa capsid protein.

When probed with anti-p17, a protein of approximately 130 kDa was identified in the HeLa cell lysate (marked by a white arrow in Figure 3.6) derived from PrV-infected cells. There were also bands visible at 65, 40 and 25 kDa. When proteins from PrV-infected HeLa cell lysates were immunoprecipitated with anti-p113 antibodies, we detected one band at 130 kDa and a second band at 60 kDa. It is tempting to speculate that the 130 kDa protein represents PrV p130. Luke *et al*. (2008) predicted that the 2A-like site at the N terminus of p130 protein was active which would result in the generation of p17 and p113. We did not detect any proteins that corresponded to p113 or p17. This suggests that either the levels of p113 and p17 were too low to detect or alternatively, the 2A-like site was not functional in mammalian cells. When the PrV-infected HeLa cell lysate was subjected to IP with IgG-biotin anti-p40 antibodies, a number of bands were detected which could represent p104 and p40 (marked by a ^ in Figure 3.6). Similar bands were detected when the HeLa cell lysate was immunoprecipitated with anti-p104 antibodies. In addition to p104 and p40, the anti-p104 probed western blot showed bands at 50 and 45 kDa. It is also interesting to note that there is a band detected at 60 kDa; both GeneMarkS and SoftBerry (Figure 3.3) predicted a 60 kDa protein. The IP reaction with anti-VCAP antibodies resulted in the detection of a 60 kDa band, which is the expected size of the mature capsid protein (marked by a * in Figure 3.6).

These western blots with PrV-specific antibodies were surprising. We naively expected that antibodies raised to specific proteins would detect the target proteins specifically and at high efficiency. In reality, we detected bands at sizes that could represent the PrV proteins of interest. In addition, we detected a number of unexpected bands and we are not sure of their identity. For example, in the IP with anti-VCAP antibodies, we detected a band at approximately 85 kDa. It is possible that this is the full length product from the sgRNA, however it is very unlikely. In an effort to gain clarity, we analysed the proteins immunoprecipitated with PrV-specific antibodies with LC-MS/MS (Table 3.1).

Table 3.1: Mass spectrometry analysis of PrV proteins.

Identified protein	Possible Coverage (%)	Coverage (%)	Confidence (%)
p17			
Gene 1	58.44	0	0
P130 and Gene A	81.07	0	50
VCAP, Gene 4, Gene D	68.97	0	0
VCAP (p8), Gene 4, Gene D	68.97	0	50
Gene E	64.43	0	0
P104	87.96	0	0
p113			
Gene 1	58.44	0	0
P130 and Gene A	81.07	0	0
VCAP, Gene 4, Gene D	68.97	0	0
Gene E	64.43	0	0
P104	87.96	0	0
p40			
P130 and Gene A	81.07	0	0
VCAP, Gene 4, Gene D	68.97	0	0
Gene E	64.43	0	0
Gene 1	58.44	0	0
P104	87.96	0	0
p104			
P130 and Gene A	81.07	0	0
VCAP, Gene 4, Gene D	68.97	0	0
Gene E	64.43	0	0
Gene 1	58.44	0	0
P104	87.96	0	0
VCAP			
Gene 1	58.44	0	0
P130 and Gene A	81.07	2.46	100
P104, Gene 3, Gene C	87.96	8.64	100
VCAP, Gene 4, Gene D	68.97	2.92	100
Gene E	64.43	0	0
P104	87.96	8.64	0

The peptide sequences identified by LC-MS/MS were compared to theoretical libraries of PrV-derived peptides including those predicted by GeneMarkS and Softberry. As decoys, we used reversed PrV peptide sequences. The possible coverage and coverage represent the

coverage of the target sequence based on the specific search criteria. These were informed by the conversations with Dr Mare Vlok, a researcher with extensive experience in protein preparation, protein analysis and data interpretation. The confidence of the hit is associated with a posterior error possibility (PEP) score that is commonly used in proteomics to determine the validity of the match. Here, a higher confidence is associated with a hit that occurs in the target library before it occurs in the decoy library. The confidence is therefore affected by the size of the target and decoy libraries, and a smaller library has a negative impact on the PEP score. With respect to the LC-MS/MS data, we detected possible coverage of PrV proteins immunoprecipitated by p113, p40 and p104 but not with any confidence. These data indicate that all the PrV-specific antibodies are interacting not only with their specific protein target but also with other PrV proteins. This is difficult to explain as it suggests a lack of antibody specificity. The validation of the PrV-specific antibodies continues to be an important part of ongoing research. It is important to note that only the analysis of the anti-p17 and anti-VCAP antibody associated proteins yielded data with any degree of confidence. For stringency, we will consider only the data that was highlighted as having a confidence level of over 50%.

The IP reaction with the anti-p17 antibody contained proteins that were similar to p130 and the p8 region of the VCAP sequence (Table 3.1). The anti-p17 antibody recognises the N-terminal region of p130 and the fact that this is detected by LC-MS/MS suggests that the PrV-infected HeLa cells are producing p130. There is still no clarity on whether the 2A site in the p130 protein is functional and whether it generates p17 and p113. It is interesting that the anti-p17 antibody recognises the p8 protein that would be generated from the action of a functional 2A site in the VCAP sequence. This indicates the importance of the ‘small proteins’ that have been overlooked thus far when considering PrV replication. The IP reaction with the anti-VCAP antibody immunoprecipitated p130, p104 and VCAP. This is quite interesting; it suggests that the antibody for VCAP is interacting with p130, p104 and VCAP. Alternatively, it means that VCAP protein interacts with p130 and p104 and so when VCAP is immunoprecipitated, it also precipitates the other PrV proteins. Regardless of the scenario, this substantiates that the p130, p104 and VCAP proteins are being produced in the PrV-infected HeLa cells, albeit at low levels.

3.2.3 Providence virus replication

The detection of PrV proteins in HeLa cell lysates by western analysis as well as using LC-MS/MS indicated that PrV had established a productive infection in mammalian cells. We do however have no information to date on the location of the replication complexes nor on the mechanisms of replication. For this reason, we pursued research to characterise PrV replication in mammalian cells and to compare it to what occurs in insect cells.

In insect cells, the PrV replication complexes involve p40/p104 and dsRNA, and the replication complexes present as discrete punctate structures in the cytoplasm of the PrV-infected cells (Nakayinga, 2013; Short *et al*., 2013). These structures co-localised with markers of the Golgi apparatus and secretory vesicles leading Short *et al*. (2013) to propose that the PrV replication complexes in insect cells were localized to the membranes of these organelles. To evaluate whether the PrV replication complexes in mammalian cells were located in the same place, PrV-infected HeLa cells were probed with antibodies targeted to the replication accessory protein/replicase (p40/p104) and PrV VCAP. In addition, antibodies that recognise dsRNA, the replicating form of RNA for a single-stranded positive-sense RNA virus. These cells were examined by confocal microscopy (Figure 3.7, uninfected HeLa cells included in Appendix A6).

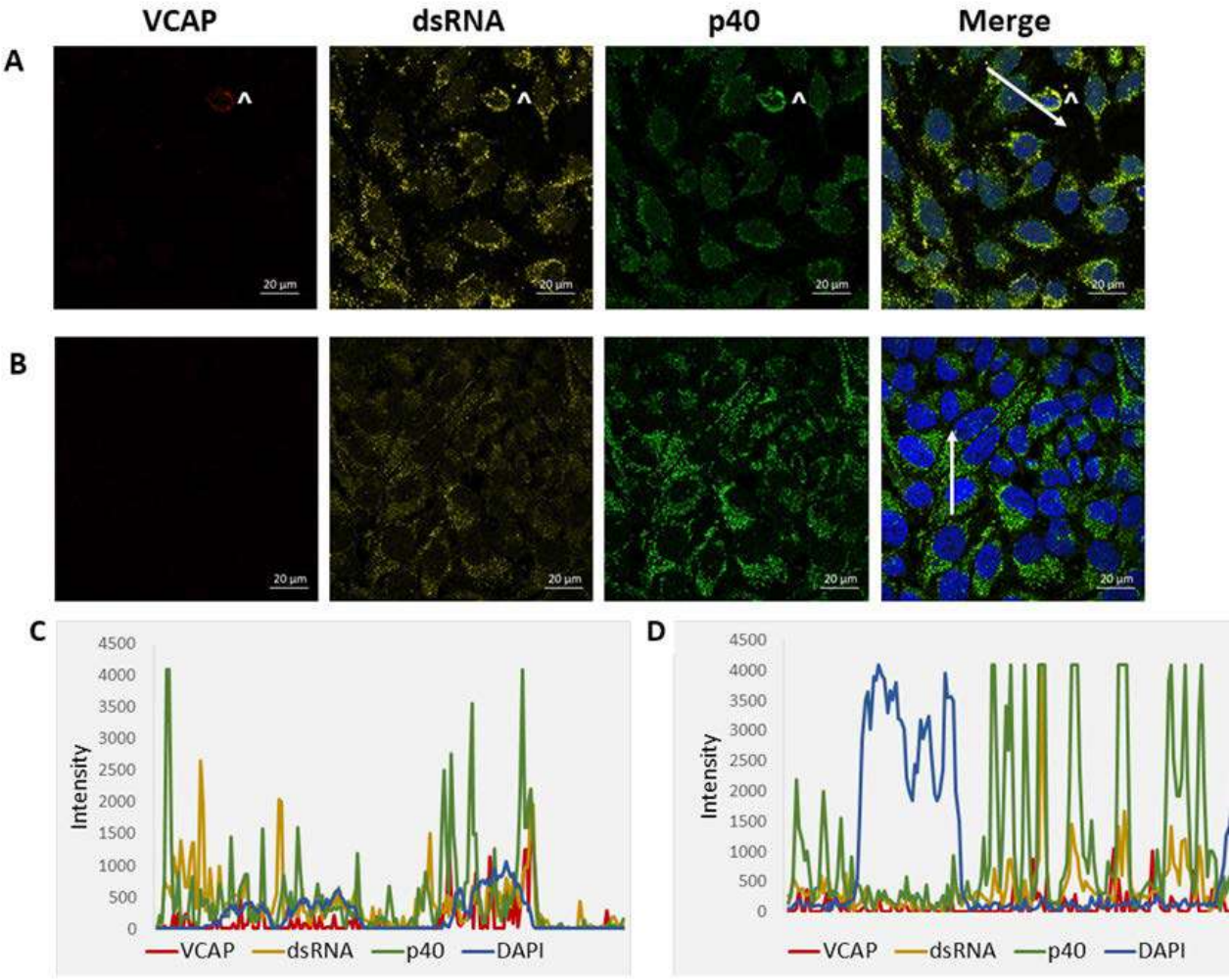


Figure 3.7: Detection of PrV-specific replication proteins in HeLa cells. (A and B) PrV-infected HeLa cells were fixed with 4 % paraformaldehyde, permeabilised and incubated with the following antibodies: IgG-biotin anti-p40 with streptavidin AF 488; anti-dsRNA (mouse) and

goat anti-mouse AF 546 and anti-VCAP (rabbit); and goat anti-rabbit AF 633. The nuclei were stained with DAPI. **(C and D)** The white arrows indicate a section that was used to generate a profile view of the immunofluorescence in each channel for the merged images in A and B respectively. A cell producing detectable signal for PrV VCAP is highlighted by the white ^ marker in panel A. The scale bar represents 20 µm.

The replication complexes in the HeLa cells formed discrete, punctate structures in the cytoplasm of the cells (Figures 3.7A and 3.7B). The signal for p40 (and p104) and dsRNA fluorescence co-localised (Figures 3.7C and 3.7D) in the cells. The overlap in p40 and dsRNA was also observed by Short *et al.* (2013) in PrV-infected insect cells. Very few cells were observed that produced detectable PrV VCAP signal and one such cell is present in Figure 3A. In these cells, there was an overlap of p40, dsRNA and VCAP (Figure 3.7A). These cells appeared to be collapsed and we hypothesized that these cells were undergoing cell death-related processes. It is possible that PrV VCAP is produced late in the infection and only when the PrV-infected cell was dying. It is also interesting to note that all the HeLa cells show the presence of p40 and dsRNA. This suggests that PrV has established a persistent viral infection in the HeLa cells and that this infection is not cytopathic to HeLa cells. The lack of PrV VCAP also suggests that PrV may have an alternative way in which to transmit itself from cell-to-cell. It is interesting to speculate that PrV employs a mechanism similar to that employed by the umbraviruses, which permits the movement of RNA as an infective agent, between cells.

It is important to highlight the differences in antibody sensitivity between the western analyses and the confocal microscopy, for example the activity of IgG-biotin anti-p40 (Figures 3.5 and 3.6 vs. Figure 3.7). The sensitivity and specificity of the IgG-biotin anti-p40 and anti-VCAP antibodies appears to be higher in confocal microscopy than in western analyses. In insect cell, PrV replication complexes are reported to associate with detergent resistant membranes (Short and Dorrington, 2013; Short *et al*., 2013); it is therefore possible that p40 and VCAP were not being detected because they were associated with the cellular membranes. When the PrV-infected HeLa cell lysates were clarified, by centrifugation, the PrV proteins that were membrane-associated were discarded. We would then only detect the soluble and free proteins, and this may account for the low levels of PrV proteins detected.

The observation that the PrV replication complexes were localized to the cytoplasm of PrV-infected HeLa cells and the differential detection of the PrV proteins in biochemical analyses led us to query whether PrV replication complexes in mammalian cells were also associated with membranes. The localisation pattern of the replication complexes was compared to published localisation patterns of cellular organalles and it was found that PrV replication complex localisation was most similar to that of the mitochondria. This prompted the

bioinformatic investigation. The critical proteins in the replication complexes are p40 and p104; these amino acid sequences for these proteins were subjected to bioinformatics analyses to predict their localization within the HeLa cells (Table 3.2).

All three bioinformatics analyses predicted that p40 and p104 would localise to membranes, and specifically the membranes of the mitochondria (Table 3.2). The most confident predication was returned by UCL Bioinformatics. UCL bioinformatics used FFPred3 to assign gene ontology (GO) terms to human and other eukaryotic sequences (Cozzetto *et al*., 2016). UCL Bioinformatics reported the mitochondrial and membrane localization for PrV p40 and p104 with 97% confidence, respectively. While the other bioinformatic tools did not report high confidence levels, their predicted localization aligned with that reported by UCL Bioinformatics. PredictProtein made use of LocTree3 (Goldberg *et al*., 2014) as a subcellular prediction tool which is specific for eukaryotes and bacteria. This site assigned the location of both p40 and p104 as the mitochondria. Of the three selected bioinformatics sites selected, only CELLO2GO (Yu *et al*., 2014) permitted a 'virus' specific input. While the results did not show a high confidence in the predicted localization, the site of localization correlated with the other bioinformatics data. The predicted localization of p40 and p104 to the mitochondria is highly unexpected. In insect cells, PrV replication complexes are associated with the Golgi apparatus and the secretory vesicles (Short *et al.,* 2013).

Table 3.2: Subcellular predictions of PrV replication proteins p40 and p104.

Prediction site	P40		p104	
	Organelle	**Confidence (%)**	**Organelle**	**Confidence (%)**
UCL Bioinformatics	Mitochondrial envelope	97	Integral component of membrane	96.6
Predict Protein	Mitochondria	33	Mitochondria	36
CELLO2GO	Plasma membrane	52.8	Plasma membrane	33.5
	Mitochondria	12.2	Mitochondria	29.1

To validate the bioinformatics data *in vivo*, PrV-infected HeLa cells were probed for PrV-specific signal in conjunction with probing with mitochondria-specific antibodies and stain. First, PrV-infected HeLa cells were stained with MitoTracker and probed with antibodies for

dsRNA and p40 (Figures 3.8A and 3.8B). The PrV-infected HeLa cells were also probed with mitochondria-specific antibodies targeted to ATP5B, an inner mitochondrial membrane protein, and dsRNA (Figures 3.8C and 3.8D). The prepared cells were analysed using confocal microscopy (Figure 3.8).

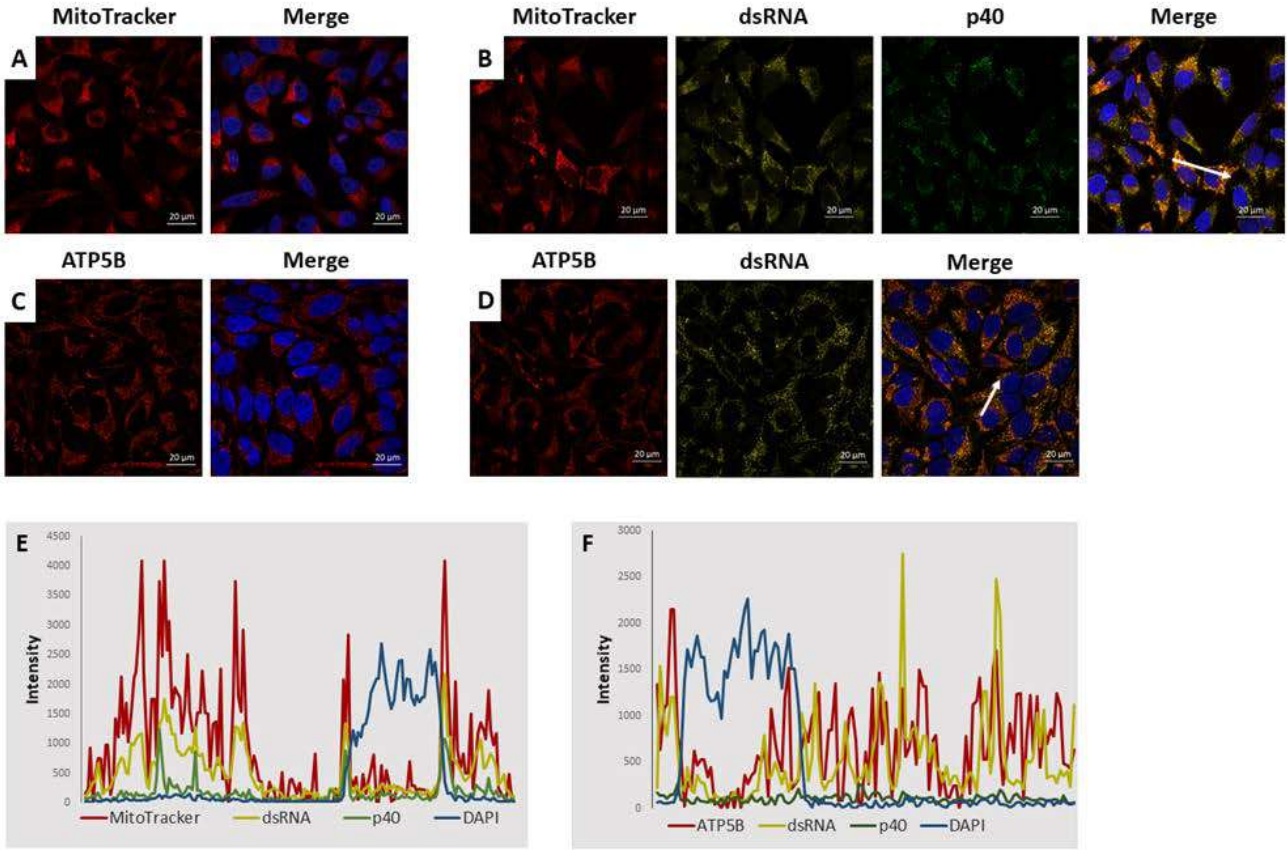


Figure 3.8: PrV replication is localised to the mitochondria in HeLa cells. (A and B) PrV-infected HeLa cells were plated on glass cover slips and settled overnight. Live cells were stained with MitoTracker for 4 hrs and then prepared for microscopy. **(B)** The PrV-infected HeLa cells were probed with antibodies including IgG-biotin anti-p40 with streptavidin AF 488, anti-dsRNA (mouse) and goat anti-mouse AF 546. **(C and D**) PrV-infected HeLa cells were probed with anti-ATP5B (goat) and donkey anti-goat AF 633 and anti-dsRNA (mouse) with donkey anti-mouse AF 546. Control cells were probed with MitoTracker **(A)** and ATP5B **(C)** to demonstrate signal in the absence of PrV-specific signal. The nuclei were stained with DAPI. The scale bar represents 20 µm. **(E and F)** The white arrows indicate a section that was used to generate a profile view of the immunofluorescence in each channel for the merged images, E – MitoTracker, F – ATP5B.

HeLa cells probed with MitoTracker stain (Figure 3.8A) and ATP5B (Figure 3.8C), in the absence of PrV-specific antibodies, showed similar staining in the two sets of samples. When PrV-specific antibodies were added to the mitochondrial stain/antibody, we observed that the PrV-specific signals for dsRNA and p40 (Figure 3.8B) and dsRNA alone (Figure 3.8D) overlapped with the signal for the mitochondria (Figure 3.8B and 3.8D). This is also apparent

in the profiles taken of HeLa cells that were probed for the mitochondria as well as the PrV-specific proteins (Figure 3.8E and 3.8F). The profile of the fluorescent channels show that where there is a peak in red (MitoTracker) there is also a peak in yellow (dsRNA) and green (p40) fluorescence, indicating that dsRNA and p40 co-localise with the mitochondria in mammalian cells. The same trend can be seen in Figure 3.8D. Here, the antibody for p40 was not used as the secondary goat anti-streptavidin antibody would have cross-reacted with the primary ATP5B antibody. From these data, we concluded that PrV p40 and dsRNA form replication complexes that are associated with the mitochondrial membranes in HeLa cells.

3.2.4 PrV translation

The localization of PrV to the mitochondria raises the question, where is the PrV RNA being translated? The mammalian mitochondria encode their own translation machineries; if the PrV complexes are localised to the membranes of the mitochondria, then it is possible that PrV may be translated by the mitochondria or by the translation mechanisms present in the cytosol of the HeLa cell. To determine the site of translation, PrV-infected HeLa cells were treated with chloramphenicol (20 µg/ml chloramphenicol for 48hrs) or cycloheximide (10 µg/ml cycloheximide for 24hrs). Chloramphenicol is an antibiotic that selectively inhibits the mitochondrial translation machinery in eukaryotic cells without affecting the cytoplasmic machinery (Lipton and McMurray, 1977; McKee *et al*., 2006). Cycloheximide is a eukaryotic antibiotic that inhibits protein synthesis by binding to the 60S ribosomal subunit and preventing elongation during translation (Schneider-Poetsch *et al*., 2010). After treatment, the HeLa cells were lysed and the proteins immunoprecipitated using anti-p40 antibodies. The proteins were separated on SDS-PAGE gels and were then analysed by western blot analysis using either anti-p40 or anti-p104 antibodies (Figure 3.9).

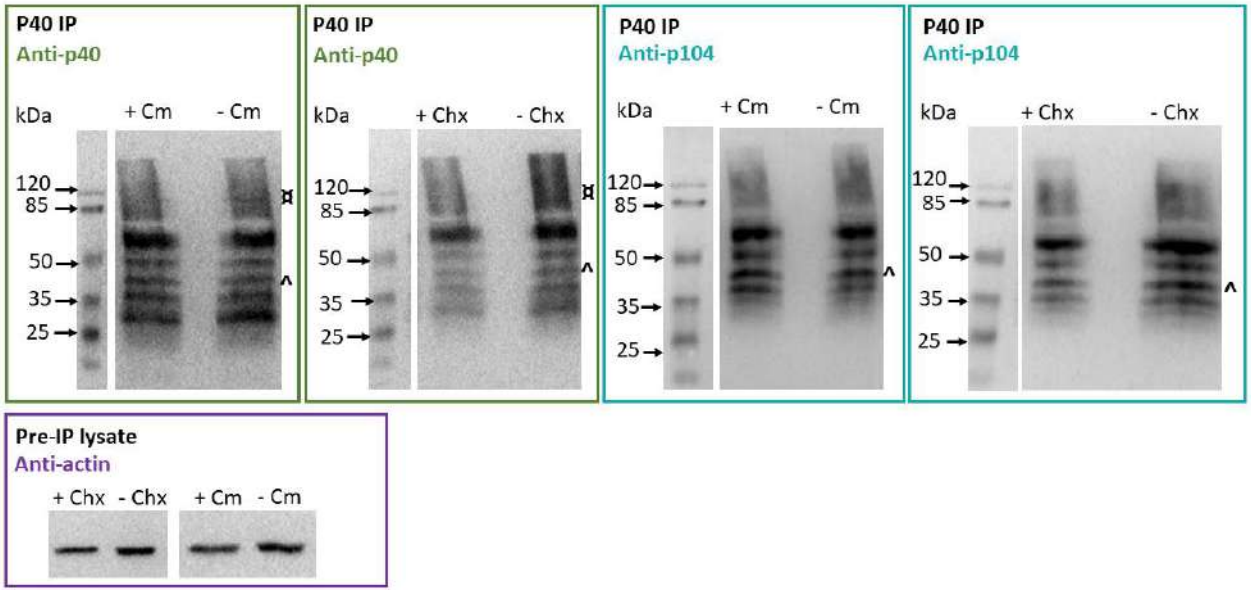


Figure 3.9: PrV makes use of host cytosolic translation machinery. PrV-infected HeLa cells were treated with either 20 µg/ml chloramphenicol (Cm) for 48hrs or 10 µg/ml cycloheximide (Chx) for 24hrs.The cells were then collected, lysed and incubated with IgG-biotin anti-p40 antibodies overnight at 4 °C, rotating. The antibody-lysate mixture was then mixed with A/G coated magnetic beads and incubated overnight at 4°C, with rotation. The beads-protein complexes were collected and separated using SDS-PAGE and analysed using western blots. The membranes were probed with the following antibodies; IgG-biotin anti-p40 with streptavidin HRP, anti-p104 (rabbit) with goat anti-rabbit HRP and anti-actin (goat) with donkey anti-goat HRP. The black ^ indicates p40 protein and the black ¤ indicates p104.

PrV-infected HeLa cells that were treated with chloramphenicol showed no change in the levels of p40 and p104 detected when probed with anti-p40 and anti-p104 antibodies (Figure 3.9). The levels of protein detected was comparable between the ± chloramphenicol-treated samples. The cells treated with cycloheximide showed lower levels of protein when probed with anti-p40 and anti-p104 antibodies. These data suggest that PrV utilises the cytoplasmic translation machinery as opposed to the mitochondrial translation machinery within the cell. Actin was used as a loading control in this experiment. The levels of actin are not affected by treatment with either chloramphenicol or cycloheximide, which indicates that actin is a stable protein that is not rapidly turned over in the cell.

3.3 Discussion

In this chapter, we introduced PrV as a model system to study RNA virus replication in mammalian cells. It is important to highlight here that we propose that PrV be used as a model for the development of analytical systems that can be applied to the study of other viruses, rather than PrV as a model system for other viruses. We performed bioinformatics analyses and then validated the data experimentally. Our research, when taken with the knowledge

about the replication of PrV in insect cells, permits us to better understand PrV and its replication strategies in mammalian cells.

Here, we report that PrV was localized to the mitochondria in the mammalian cells. All single-stranded, positive-sense RNA viruses require an association with organelle membranes to form replication complexes. The membranes play a vital role in the structure and organisation of the replication factory. In addition, the organelle associated replication complexes protect the dsRNA replication intermediates from the host cell antiviral responses (Martinez-Turino and Hernandez, 2012). There are only a small number of viruses that are known to replicate in or on the mitochondrial membrane. These include members of the *Tombusviridae* (e.g. pelargonium flower break virus and melon necrotic spot virus; Gómez-Aix *et al*., 2015; Martínez-Turiño and Hernández, 2012) and the *Nodaviridae* (Flock House virus; Miller *et al*., 2001).

Flock House virus (FHV) is a nodavirus that replicates on the outer mitochondrial membrane in structures known as spherules. Protein A, the RdRp of FHV, localises to the outer mitochondrial membrane and forces the membrane to invaginate and these invaginations permit the formation of the spherical structures that protect the dsRNA intermediate from being detected by the host cell antiviral responses (Miller *et al*., 2001; Short *et al*., 2016). It is tempting to speculate that PrV forms similar protective structures, using p40 and p104 as it would explain the protection of the dsRNA and the lack of detectable antiviral responses in the host cell. Our attempts to visualize the membrane of the mitochondria using transmission electron microscopy (TEM) were not successful, this remains an aspect we need to address.

Pelargonium flower break virus (PFBV; genus- *Carmovirus*) and melon necrotic spot virus (MNSV; genus- *Carmovirus*) are members of the *Tombusviridae* and share RdRp homology with PrV. These two viruses are also known to replicate in modified mitochondrial membranes (Gómez-Aix *et al*., 2015; Martinez-Turino and Hernandez, 2012). PFBV produces p86 (RdRp) and p27 via an RTS. P27 has ssRNA binding potential and could therefore play a role in the recruitment of viral RNA to the replication complex. PrV p40, which is also produced as a result of a RTS, has predicted RNA binding sites. Martinez-Turino and Hernandez (2012) demonstrated that PFBV p27 was targeted to the mitochondria and was tightly associated with the mitochondrial membranes. They also performed bioinformatic analysis on p27 and detected no definitive mitochondrial localisation sequences (MLS) on the protein. This was also the case for p40. The lack of a MLS does not rule out the localisation of the protein to the mitochondrial membranes. Most outer membrane proteins lack an MLS and instead they

contain internal targeting information, which is difficult to predict (Chacinska *et al*., 2009). This makes it even more difficult to identify proteins that are translocated to the mitochondria.

Eukaryotic viruses require host translation machineries to produce their proteins but unlike eukaryotic mRNAs, viruses express more than one protein from a single transcript. Most viruses employ non-canonical translation mechanisms to access multiple ORFs within the viral genome. Viral mechanisms include internal ribosomal entry sites (IRES), leaky scanning, non-AUG initiation, ribosomal shifting and read through stop codons (Firth and Brierley, 2012). PrV uses RTS codons to produce p40 from the p104 sequence, which is a mechanism employed by plant viruses (Firth and Brierly, 2012) and the recently reported H14-hedgehog/2015/HUN virus (Reuter *et al.*, 2019). The PrV capsid precursor protein is translated from a sgRNA as a single protein and is then autoproteolytically cleaved to produced two proteins (60 and 8 kDa). It is possible that non-AUG start codons or other alternative mechanisms may also be utilized by PrV to produce the proteins detected in the IPs.

The western analyses and LC-MS/MS analyses indicate the importance of antibodies to studies like these. Despite extensive attempts to optimize the protocols that required the use of antibodies, we are concerned about the sensitivity and the specificity of the PrV-specific antibodies. This concern is higher in this study because the levels of the PrV proteins in the infected cells are much lower than those detected in insect cells. Our data suggest that PrV may express proteins that we are unaware of to date. Viruses employ multiple mechanisms to express the proteins required to establish a successful infection. The *Tombusviridae* produce multiple small proteins that are required for cell-to-cell movement. PrV is closely related to the *Tombusviridae* and so it is not unreasonable to hypothesize that PrV also produces small movement proteins, for example p6 and p8. These would permit cell to cell spread without the need for a viral capsid. Poliovirus, a small non-enveloped positive-sense RNA virus uses non-lytic transfer to move viral material between cells (Bird *et al*., 2014; Bird and Kirkegaard, 2015). The virus forms replication organelles with double membranes in the cytoplasm during infection and these structures closely resemble autophagosomes (Bird *et al*., 2014). The authors demonstrated that poliovirus relies on the autophagy pathway to exit the cell without causing in cell lysis. This challenges the current understanding of the replication cycle of non-enveloped viruses. With this information, it is reasonable to suggest that PrV may use similar mechanisms to transfer material between cells without causing cell death.

It is interesting to note that FHV is the only virus, other than PrV, that has been shown to replicate in plant, mammalian and insect cells (Miller *et al*., 2001). This leads us to question whether the site of replication influences the host range of the virus. With the discovery of new

viruses, and the expansion of the phylogenetic trees, we expect the discovery of more viruses with broad host ranges. The ability of PrV to cross Kingdom boundaries (Jiwaji *et al*., 2016; Jiwaji *et al*., 2019; Pringle *et al*., 2003), highlights the need to development tools to study viruses that pose a potential threat. The model system based on PrV provides an opportunity to develop the tools required to study unknown virus replication, viral persistence and the broad spectrum host range of certain viruses.

Chapter 4: Developing an *in vitro* tool to study the function of viral proteins

4.1 Introduction

Proximity dependent labelling of proteins

In 2012, Roux *et al*. used a promiscuous biotin ligase (BirA) to label proteins that interact with lamin A. Unlike traditional protein-protein interaction assays, which are reliant on the strength of the protein interactions and are fraught with the challenges of optimizing the binding efficiency of bait and prey proteins *in vitro,* this proximity based system permitted the identification of proteins that were proximal to the protein of interest, *in vivo.* The BirA was isolated from *Escherichia coli* (*E. coli*); and in bacterial cells, BirA activity is both highly specific and stringently controlled (Choi-Rhee *et al*., 2004). The enzyme uses a two-step biotinylation process; the first uses biotin and ATP to generate the highly reactive bioAMP (biotinyl-adenylate) intermediate. BirA holds the reactive intermediate in the active site until a protein with an exposed ε-amino group on the target lysine is available (Choi-Rhee *et al*., 2004; Kwon and Beckett, 2000). In 2000, Kwon and Beckett modified the wild-type *E. coli* BirA enzyme to produce an R118G mutant; this BirA* enzyme was less stringent than the wild-type BirA. BirA* formed bioAMP but instead of holding the reactive intermediate in the active site, the bioAMP was released into the surrounds of BirA. The free and highly reactive bioAMP targeted any proximal proteins with exposed primary amines, primarily lysine, and biotinylated the residues (Choi-Rhee *et al*., 2004; Kwon and Beckett, 2000; Roux *et al*., 2018). Choi-Rhee *et al*. (2004) showed that BirA* promiscuously biotinylated both BirA* as well as proximal proteins.

In the Roux *et al.* (2012) study, BirA* fused to human lamin A was introduced into mammalian cells in the presence of excess biotin. The BirA*-human lamin A labelled BirA*, lamin A and proteins that were proximal to lamin A with biotin. The biotinylated proteins were then isolated by IP and identified using mass spectrometry. The authors named this system BioID (Roux *et al*., 2012). The BioID system is a complementary alternative to methods like the yeast-2-hybrid system and affinity complex purification as it is more sensitive to transient interactions as well low frequency associations (Roux *et al*., 2018). It is interesting that the developers of BioID recommended that this system should be used as a screening platform for potential interactors rather than a mechanism of validation (Roux *et al*., 2012; Roux *et al.,* 2018). This may be because of the abundance of endogenously biotinylated proteins in the cell. In addition, the developers of BioID highlighted that the extent of biotinylation should not be used as an indicator of the strength of the interaction (Roux *et al*., 2012; Roux *et al.,* 2018). The extent of biotinylation is dependent on the number of exposed primary amines. This in turn means that the lack of biotinylation also does not rule out interaction with the protein of interest (Roux *et al*., 2012; Roux *et al.,* 2018).

In 2015, Le Sage *et al*. and Ritchie *et al*. both used BioID system to study the HIV-1 Gag polyprotein. Le Sage *et al*. (2015) fused BirA to the N-terminal of HIV-1 Gag polyprotein to investigate the host proteins that interact with the HIV structural proteins. Two Gag-interacting proteins identified using mass spectrometry analysis, DDX17 and RPS6, were confirmed by using IP reactions. Ritchie *et al*. (2015) made use of a different strategy, they inserted the promiscuous BirA in between the HIV matrix and capsid proteins. The authors also generated a mutant fusion protein where the matrix protein had been removed. In the study, these constructs were used to study viral protein interactions as well as virus-host protein interactions. Lajko *et al*. (2015) also used the foundation of the BioID system in a proximity dependent biotinylation assay format to study herpesviruses, and specifically the glycoproteins of HSV-1 and Epstein-Barr virus (EBV). Here, the candidate protein was tagged with BirA and a second protein that had been identified as a potential interactor was tagged with an acceptor peptide. The acceptor peptide would be biotinylated if the interactor and target proteins were in proximity. The authors found this technique lacked specificity and hypothesized that because the candidate protein was bound to a membrane, the frequency of random collisions was increased and so the high number of false positives.

The BioID system has also been used to study enveloped viruses as well as insoluble and membrane-associated proteins (Le Sage *et al*., 2015; Lajko *et al*., 2015; Ritchie *et al*., 2015; Roux *et al*., 2018). With these successful applications of the BirA-based BioID system, we decided to adapt it to study virus replication in mammalian cells. We decided to use PrV as the replication model, and selected the best studied PrV protein, p40, as the target for study.

4.2 Results

The PrV replication accessory protein, p40 was chosen as the target of this study for three main reasons. Firstly, it remains the best studied PrV protein and there is some understanding about the protein and its interactors within the cells. Secondly, there are biological resources available for p40 including validated plasmids and antibodies. These would be critical for the progress of this project. Thirdly, we have both *in vivo* and *in vitro* methods to work with p40 and PrV, which offers a great deal of flexibility when developing a biological tool to study virus biology.

4.2.1 Validation of BirA-p40 system

Two plasmids were generated using information published by Roux *et al*. (2012); pBirA, which encoded the 37.4 kDa BirA* protein (hereafter referred to as BirA), and pBirA-p40, which expressed the 77.5 kDa BirA-p40 fusion protein (Figure 4.1). The myc-epitope tagged BirA

and BirA-p40 proteins were expressed under the control of the CMV promoter in eukaryotes. The presence of a T7 promoter permitted the same plasmid to be used for expression of the proteins of interest in bacterial systems or *in vitro* in a T7-driven TnT reaction. The EGFP sequence was inserted downstream of the BirA and the BirA-p40 sequences and an IRES, resulting in the production of the fluorescent marker protein in cells containing the plasmid construct DNA. Both plasmids also encoded the NeoR gene, which would confer resistance to Geneticin, permitting the generation of stable protein-expressing mammalian cell lines, if required.

Competent *E. coli* DH5α cells were transformed with pBirA and pBirA-p40; plasmid DNA was purified and validated using restriction enzyme digests with *Bam* HI and *Xba* I (Appendix A, Figure A2). The expected bands of 1400 and 6900 bp for pBirA-p40 (*Bam* HI and *Xba* I double restriction enzyme digest) and 890 and 6300 for pBirA (*Bam* HI and *Xba* I double restriction enzyme digest) were detected when the restricted DNA was separated on a 1 % agarose gel.

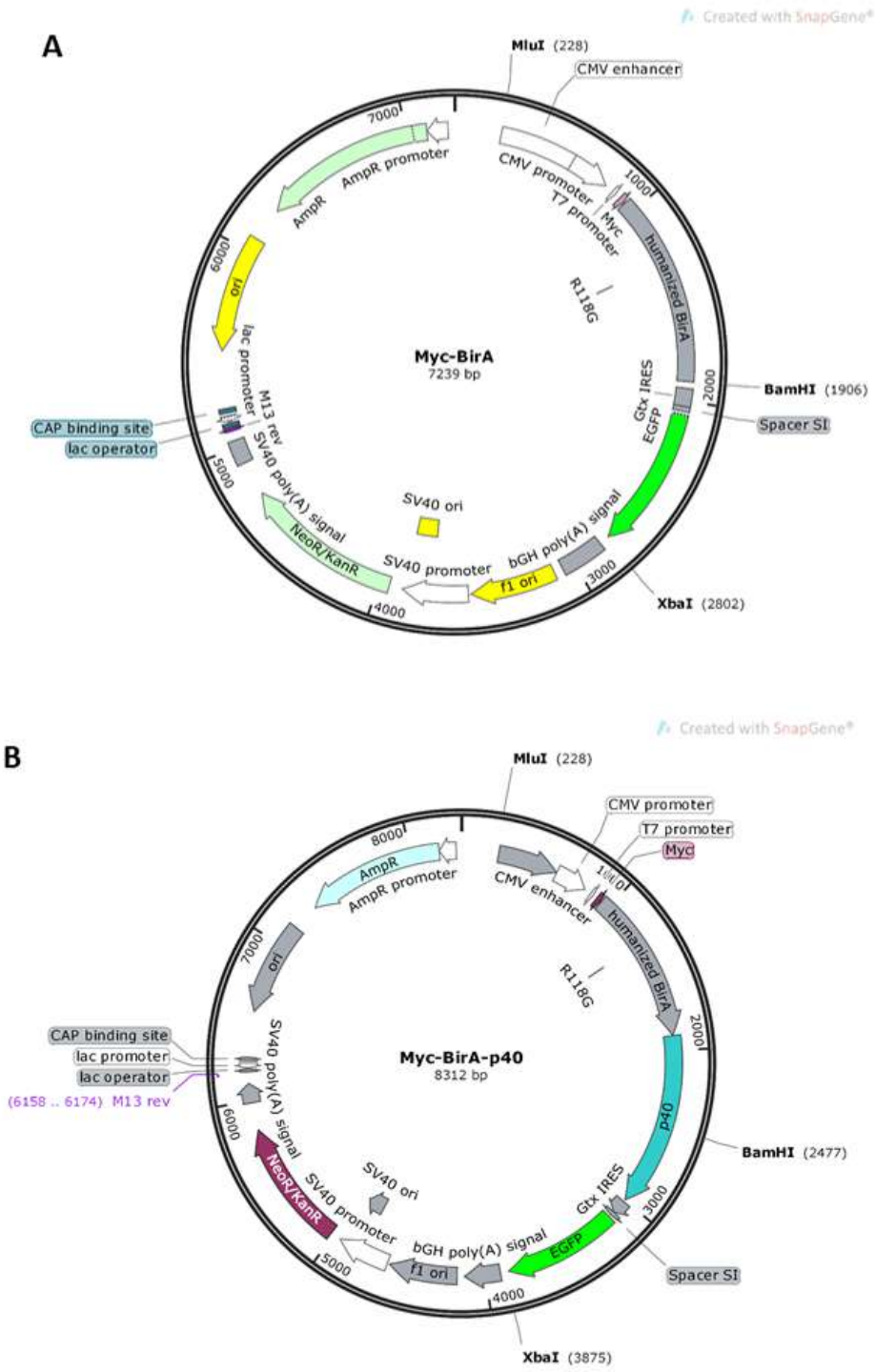


Figure 4.1: Plasmid maps of (A) pBirA and (B) pBirA-p40. The plasmids encode **(A**) Myc-tagged BirA **(B)** Myc-tagged BirA-p40 under the control of a CMV promoter for expression in eukaryotes and T7 promoter for use in bacteria. Downstream of these genes is an IRES and the sequence for EGFP. The plasmids encode resistance to the antibiotic Neomycin. Restriction enzyme sites for *Bam* HI and *Xba* I are shown; these sites were used for plasmid confirmation using restriction enzyme digests.

HeLa cells were transiently transfected with pBirA or pBirA-p40, and the cells were grown in the presence of 50 µM biotin. Protein expression was permitted for 24 hrs before the transiently transfected cells were collected for analysis. The cells were lysed and the cell debris removed by centrifugation. Proteins in the cell lysate were separated using SDS-PAGE and analysed using western blot analysis. Antibodies targeted towards the myc epitope and biotin moieties were used to probe the western blots (Figure 4.2).

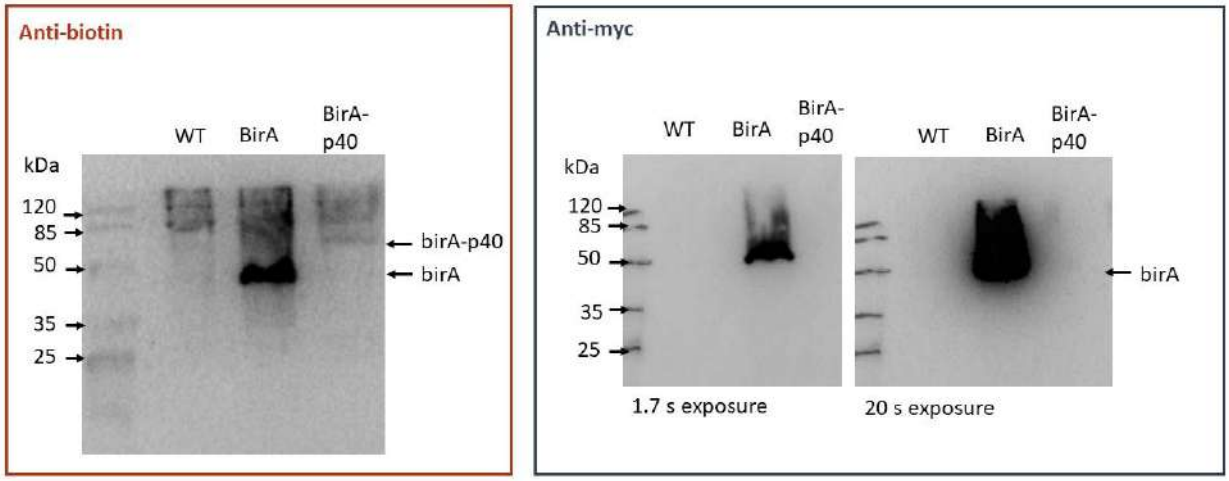


Figure 4.2: Analysis of BirA and BirA-p40 expression in transiently transfected HeLa cells. HeLa cells were transfected with pBirA or pBirA-p40 and protein was expressed for 24 hrs. The proteins were separated by SDS-PAGE (on 10 % gels) and the proteins were transferred to PVDF membranes. The western blots were probed with antibodies targeted towards anti-biotin (rabbit) with goat anti-rabbit HRP or anti-myc (mouse) with goat anti-mouse HRP.

HeLa cells transiently transfected with pBirA robustly expressed BirA and this protein was biotinylated (Figure 4.2). In contrast, very low levels of BirA-p40 were detected using the anti-biotin antibody and no BirA-p40 was detected using the anti-myc antibody (Figure 4.2). The BirA-p40 protein that was detected was biotinylated which gave confidence that the construct pBirA-p40 was functional, and that the system had merit. In addition to the bands of interest, we detected bands at unexpected sizes in the wild-type HeLa lysate as well as those from the transiently transfected cells. These we attributed to endogenously biotinylated proteins, which are present in the HeLa cells. The levels of BirA-p40 detected in transiently transfected HeLa cells were low. In an attempt to enhance the detection of BirA-p40, the proteins were immunoprecipitated with anti-myc antibodies and then used for western analysis (Figure 4.3).

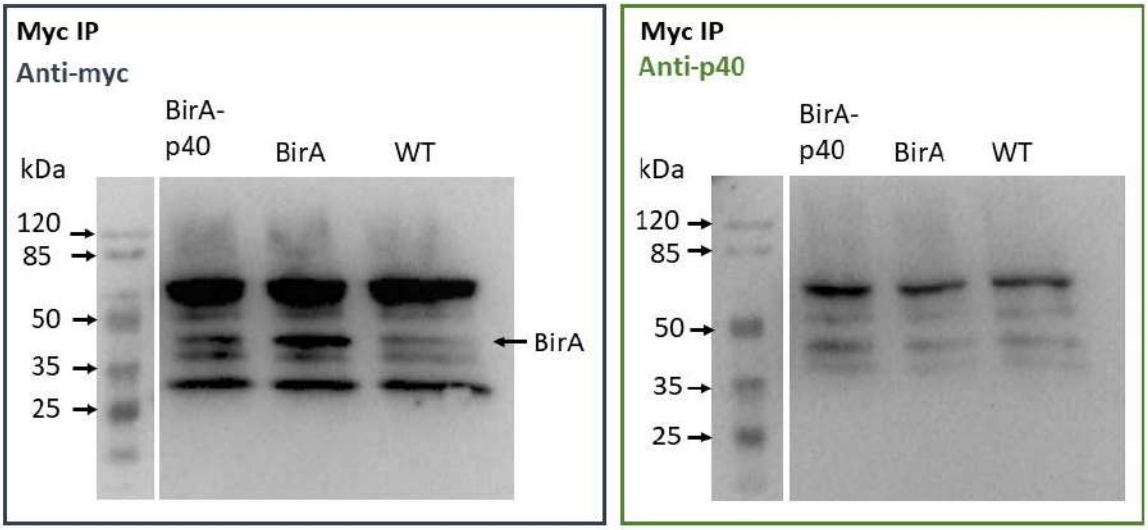


Figure 4.3: Detection of immunoprecipitated myc-tagged BirA and BirA-p40 from transiently transfected HeLa cells. HeLa cells were transfected with pBirA and pBirA-p40 and proteins were expressed for 24 hrs before the cells were collected. The cells were lysed and then incubated with anti-myc (mouse) antibodies overnight at 4 °C, with rotation. The antibody-lysate mixture was then mixed with A/G coated magnetic beads and incubated overnight at 4 °C, with rotation. The beads were collected, and the proteins separated using 10 % SDS-PAGE and then transferred to a PVDF membrane. The western blots were probed with anti-myc (mouse) and IgG-biotin anti-p40 antibodies.

The IP reactions concentrated the proteins and so more bands were clearly visible on the western analysis (Figure 4.3). The bands at 60 and 30 kDa represent those of the heavy and light chains of the anti-myc antibody used for the IP reaction. BirA was detected by the anti-myc antibody but no BirA-p40 was detected with either the anti-myc or the anti-p40 antibodies. We noted that the anti-myc antibody did not detect any BirA-p40 protein (Figures 4.2 and 4.3) but we initially attributed this to low transfection efficiencies. To evaluate the expression of BirA and BirA-p40 in HeLa cells, we prepared transiently transfected cells for confocal microscopy and probed with anti-myc antibodies for myc-tagged BirA and BirA-p40 (Figure 4.4).

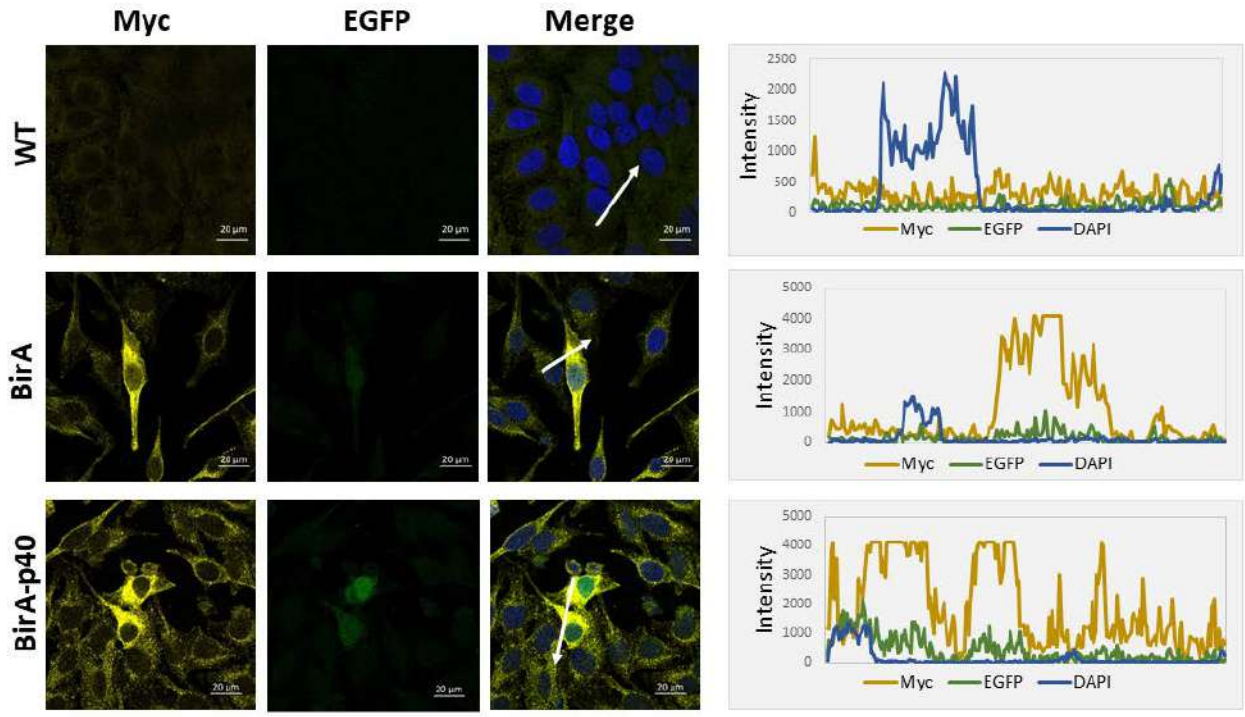


Figure 4.4: Analysis of BirA and BirA-p40 protein expression in transiently transfected HeLa cells. HeLa cells were transfected with pBirA or pBirA-p40 and permitted to express protein for 24 hrs. Cells were prepared for confocal microscopy. The cells were fixed with 4 % paraformaldehyde, permeabilised and then incubated with anti-myc (mouse) and goat anti-mouse AF 546. The WT (wild-type) cells represent untransfected control cells. The nuclei were stained with DAPI. The scale bar represents 20 µm. The white arrows indicate sections that were used to generate a profile view of the immunofluorescence in each channel for the merged images.

Transiently transfected HeLa cells expressed low levels of EGFP (Figure 4.4). When proteins are produced using an IRES, the levels of the protein are often lower (Komar and Hatzoglou, 2011; Mizuguchi *et al*., 2000). The number of transfected cells however supported the observation that the transfection efficiency was low; this could explain the low levels of protein detected in the western analyses. In cells that were transfected with pBirA and pBirA-p40, high levels of fluorescence were detected. In fact, the levels for BirA were comparable to BirA-p40 (Figure 4.4). This is surprising because we detected BirA protein in the western analyses (Figure 4.2 and 4.3) but detected no, or low levels, of BirA-p40 protein. We performed bioinformatic analyses of p40 using UCL Bioinformatics and found that the protein contained two putative transmembrane domains. This suggested that p40 may be firmly integrated in cellular membranes. If this was the case, then it is likely that the clarifying spin when preparing the cell lysates would cause the precipitation of both p40 and BirA-p40; this could explain the lack of detection in the western blots. We investigated the options to increase the expression of BirA-p40 in the HeLa cells and decided to create a HeLa cell line that stably expressed the BirA-p40 protein from chromosomally integrated DNA.

4.2.2 Generation of a stable HeLa cell line expressing BirA-p40

The integration of BirA-p40 DNA into the chromosome would result, in theory, in a cell line that produced the BirA-p40 protein at consistent levels in all the cells. This would circumvent the problems with the low transfection efficiencies and result in higher levels of proteins in the HeLa cells that could be used to perform biological analyses of interest. Plasmids pBirA and pBirA-p40 were transfected in HeLa cells and Geneticin was used to select for chromosomal integrants. The polyclonal cell line would contain a population of cells where the desired DNA, be it BirA or BirA-p40, was integrated into different positions of the chromosomal genome. The BirA and BirA-p40 expressing stable cell lines were analysed using western analysis (Figure 4.5) and confocal microscopy (Figure 4.6).

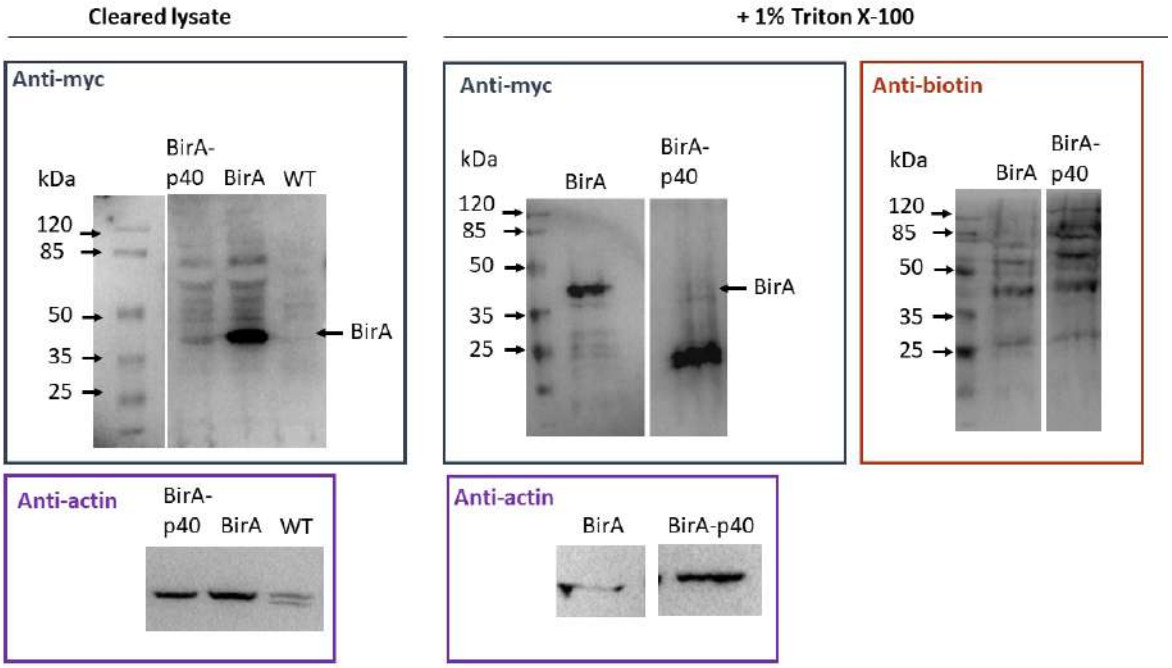


Figure 4.5: Western blot analysis of detergent-treated BirA-p40 expressed in stable HeLa cell lines. HeLa cells expressing BirA or BirA-p40 were collected and lysed in Cell Lytic M, the cell lysate was cleared and the cell debris discarded. The supernatant was loaded for western analysis. Alternatively, HeLa cells expressing BirA or BirA-p40 were incubated in Cell Lytic M containing 1 % Triton X-100. The samples were collected and the whole cell lysate was analysed using SDS-PAGE and western blot analysis. The western blots were probed with anti-myc, anti-biotin and anti-actin antibodies.

Stable HeLa cell lysates were prepared in the absence and presence of 1 % Triton X-100 to enhance the solubility of membrane-associated proteins. In the absence of the detergent, BirA was detected using anti-myc antibodies but no BirA-p40 was detected (Figure 4.5). When detergent was added, BirA was robustly detected but no BirA-p40 was detected with anti-myc antibodies. In addition, while biotinylated proteins were detected using the anti-biotin

antibodies, no proteins of the expected sizes for BirA and BirA-p40 were detected in the presence of detergent (Figure 4.5). The addition of detergent significantly complicated the SDS-PAGE and western blot process as it affected how the protein samples handled before loading on the SDS-PAGE and how the proteins separated in SDS-PAGE.

The BirA and BirA-p40 expressing stable cell lines were plated on glass coverslips and prepared for confocal microscopy (Figure 4.6). The BirA-p40 stable cell line showed the presence of myc-tagged BirA-p40 as well as EGFP in all the cells. The proteins were expressed at levels that were lower than those present in transiently transfected HeLa cells (Figure 4.6 vs. 4.4, activity profile graphs). The BirA-p40 was present in discrete punctate structures in the cytosol of the HeLa cells (Figure 4.6). The punctate structures are similar to those observed for p40 in PrV-infected HeLas (Figure 3.7). Unfortunately, the polyclonal BirA-expressing stable HeLa cell line expressed negligible levels of BirA and no EGFP was detected (Figure 4.6). When the stable cell line for BirA was generated, very few Geneticin-resistant cells were observed. While these were expanded to form a population, this line grew slowly. We hypothesized that the polyclonal population was made of HeLa cells with only a few integration events, and that these integration events may have occurred in a critical gene or region. This BirA-expressing stable cell line did not survive long after this experiment.

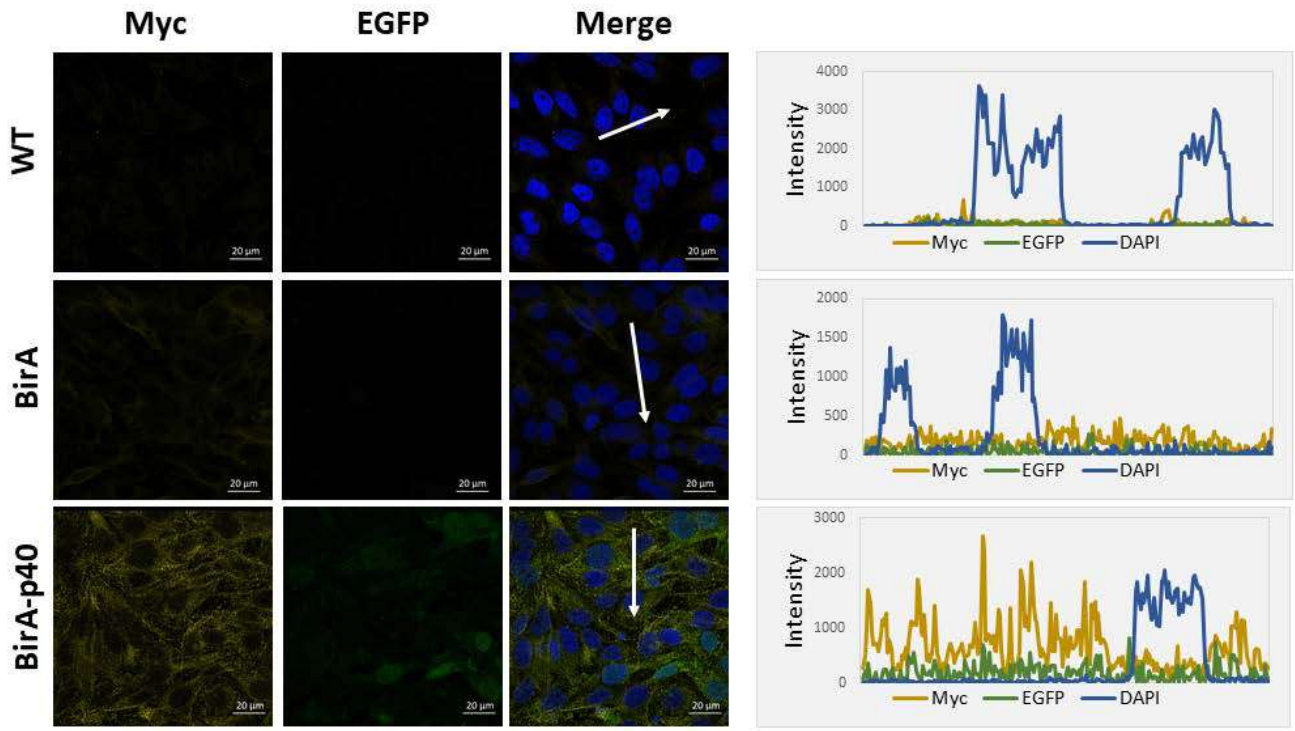


Figure 4.6: Detection of BirA-40 protein expression from a polyclonal stable HeLa cell line. The HeLa cells were plated on glass coverslips and prepared for confocal microscopy as

described previously. The cells were probed with anti-myc (mouse) and goat ant-mouse AF 546 and the nuclei were stained with DAPI. The scale bar represents 20 µm.

What is evident in the data from the BirA-p40 expressing stable HeLa cell line is that there is a clear difference between the detection of the BirA-p40 protein in western blot analysis and in confocal microscopy. The levels of the proteins detected in confocal microscopy were much higher than those detected in western analyses. The difference between western analysis and confocal microscopy is that the western analysis requires the isolation of the proteins of interest from the cells whereas in confocal microscopy, the proteins are detected *in situ.* To determine whether we had the desired protein in the cells, as visualized by confocal microscopy, but just could not extract them for western analysis, we decided to use *in vitro* TnT reactions.

4.2.3 *In vitro* expression of BirA-p40

In vitro TnT reactions contain the nuclear and cytosolic components required for transcription and translation in mammalian cells. The plasmid pBirA-p40 was added to the *in vitro* TnT reaction and the proteins were expressed. Proteins were separated using SDS-PAGE and analysed by western analysis. Membranes were probed with anti-myc and anti-biotin antibodies (Figure 4.7).

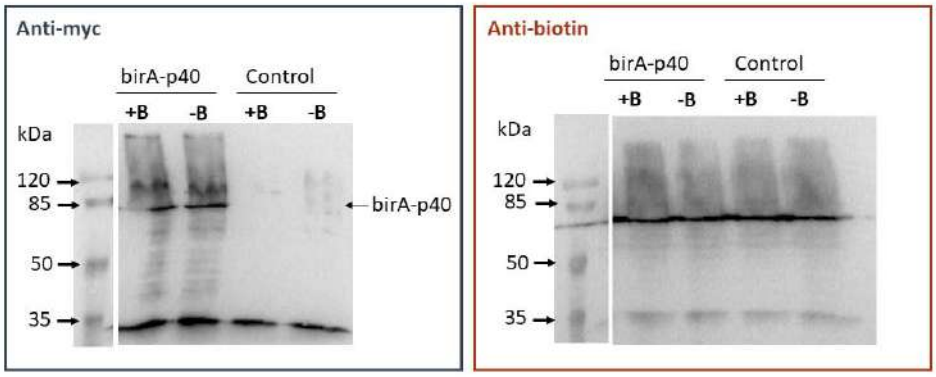


Figure 4.7: Expression of BirA-p40 using *in vitro* coupled transcription and translation reactions, in the presence and absence of biotin. The plasmid pBirA-p40 was used in the coupled TnT reactions in the presence or absence of 5 µM biotin. The control reaction contained no DNA. The reactions were incubated at 30 °C for 90 mins. The samples were separated using SDS-PAGE (7.5 % gels) and the proteins were transferred to a PVDF membrane. The membranes were probed with anti-myc (mouse) and anti-biotin (rabbit) antibodies.

Myc-tagged BirA-p40, a 78 kDa protein, was robustly expressed in the *in vitro* TnT reactions (Figure 4.7). When the same protein was probed using anti-biotin antibodies, no biotinylated BirA-p40 detected (Figure 4.7). The TnT reaction mix was high in salts and proteins, and we

thought that the quality of the western blot could be improved if the desired proteins were immunoprecipitated from the *in vitro* TnT reactions (Figure 4.8).

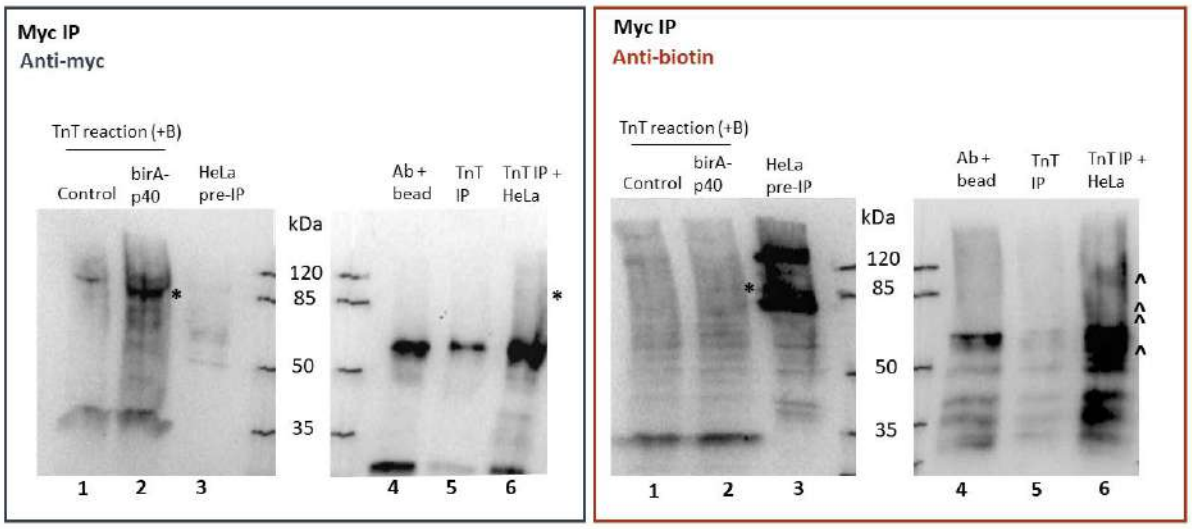


Figure 4.8: Biotinylation of proteins that interact with BirA-p40 produced *in vitro*. BirA-p40 was generated in an *in vitro* TnT reaction. The BirA-p40 was incubated with anti-myc antibodies (Ab) and then bound to A/G coated magnetic beads (TnT IP). The TnT IP was then incubated with PrV-infected HeLa cell lysate. The magnetic beads were collected and the samples analysed using western analysis. The membranes were probed with anti-myc (mouse) and anti-biotin (rabbit) antibodies.

We generated BirA-p40 protein *in vitro,* in the presence of biotin, attached the BirA-p40 protein to anti-myc antibodies and then to magnetic beads. These BirA-p40 coated magnetic beads were incubated with cell lysate from PrV-infected HeLa cells to attract any BirA-p40-interacting proteins. We detected myc-tagged BirA-p40 in the *in vitro* TnT reaction (Figure 4.8, marked with a *, anti-myc western, lane 2). Very little of the BirA-p40 protein that was used in the IP with the HeLa cell lysate was visible in the western analysis (Figure 4.8, marked with a *, anti-myc western, lane 6). In addition, a very faint band is visible for biotinylated BirA-p40 in the *in vitro* TnT reaction (Figure 4.8, marked with a *, anti-biotin western, lane 2). When bound to the HeLa cell lysate, there are bands visible in the western blot of the IP using BirA-p40 from PrV-infected HeLa cell lysate probed with anti-biotin antibodies that are not present in the western probed with anti-myc antibodies (Figure 4.8, anti-biotin lane 6 vs. anti-myc lane 6, marked with ^). These bands are 100, 84, 70, and 52 kDa in size. It is possible the 100 kDa band may be p104 and the 70 kDa band may be the PrV VCAP (68 kDa). It is not possible to speculate about the identity of the 84 and 52 kDa protein bands. The efficiency of biotinylation is poor and these samples are unfortunately not of sufficient quality to analyse by LC-MS/MS.

As the biotinylation of BirA-p40 was not efficient *in vitro*, we wanted to evaluate whether the challenges being experienced were linked to the protocol or to the subject of study, BirA-p40. We have previously used pBirA-CNK in our research group. In this construct, CNK is a scaffold protein that is involved in cellular signalling. A colleague, Jessica Swan transiently transfected HeLa cells with BirA-CNK and expressed the BirA-CNK protein. She prepared cell lysates and performed an IP with anti-biotin antibodies. Swan (2018) separated the proteins by SDS-PAGE and transferred them to PVDF membranes. The western blots were probed with anti-CNK (mouse) antibodies (Figure 4.9).

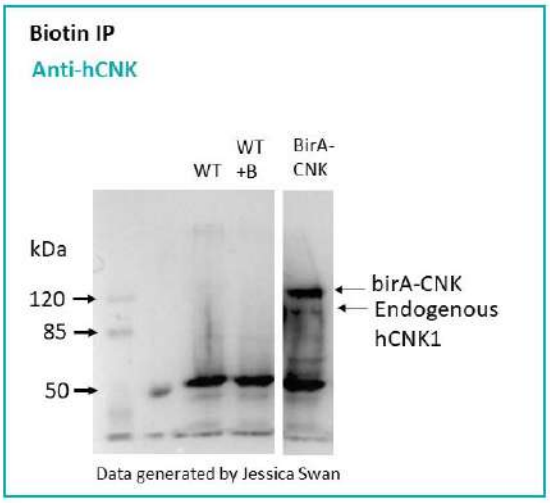


Figure 4.9: Biotinylation of BirA-CNK *in vivo*. HeLa cells were transiently transfected with pBirA-CNK and protein was expressed for 48 hrs in the presence of 50 µM biotin. The lysate was incubated with anti-biotin (rabbit) antibodies and then bound to A/G coated magnetic beads. The proteins in the IP reactions were separated by SDS-PAGE and the proteins transferred to a PVDF membrane. The western blots were probed with anti-hCNK1 (mouse) antibodies. WT – untransfected wild-type HeLa cells, in the absence or presence of biotin. BirA-CNK cells, in the presence of biotin. Data was generated by Jessica Swan.

When HeLa cells were transfected with pBirA-CNK, the cells expressed BirA-CNK. The IP reaction was performed with anti-biotin antibodies, therefore only biotinylated proteins would be immunoprecipitated. The western blots were probed with anti-hCNK1 antibodies. In the proteins isolated from HeLa cells transfected with pBirA-CNK, we observe bands for BirA-CNK as well as endogenous hCNK1. This indicates that the BirA-CNK is being produced and biotinylated *in vivo.* In addition, these data indicate that BirA-CNK is interacting with endogenous CNK1, and biotinylating this protein as well (Figure 4.9). These data indicate that the BirA is functional, and that the system works in HeLa cells *in vivo*, in our research group. To evaluate whether the BirA-CNK system would function *in vitro,* we performed an experiment with pBirA, pBirA-p40 and pBirA-CNK in the *in vitro* TnT reactions (Figure 4.10).

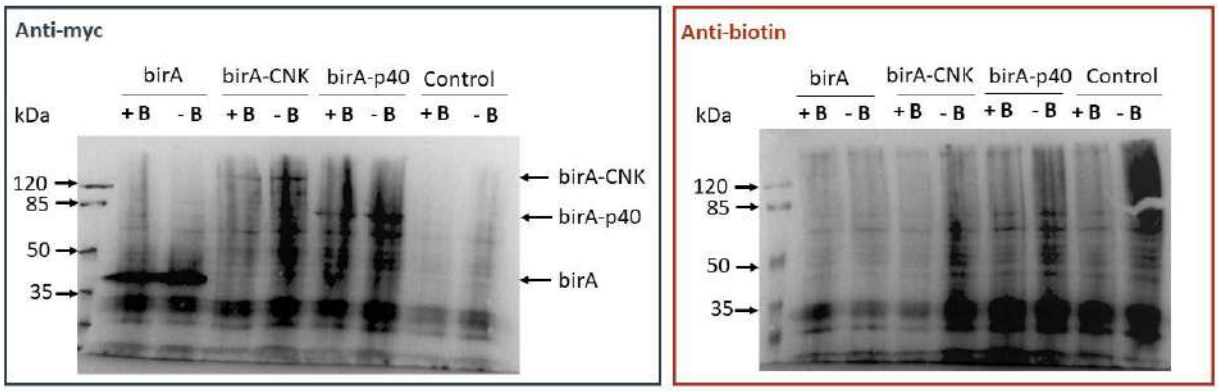


Figure 4.10: Testing the ability of BirA to biotinylated proximal proteins *in vitro*. The *in vitro* TnT reactions were repeated as described previously using the plasmids pBirA, pBirA-p40 and pBirA-CNK. The reactions were performed in the absence of or the presence of 10 µM biotin. Proteins in the *in vitro* TnT reactions were separated by SDS-PAGE and transferred to PVDF membranes. The western blots were probed with anti-myc (mouse) and anti-biotin (rabbit) antibodies.

Western blots of *in vitro* TnT reactions analysed with anti-myc antibodies showed the presence of BirA, BirA-p40 and BirA-CNK (Figure 4.10). All three constructs were functional and produced detectable and myc-tagged proteins from the constructs in the *in vitro* TnT reactions. When the same lysates were probed with anti-biotin antibodies, no significant bands were detected for BirA, BirA-p40 or BirA-CNK (Figure 4.10). There may be faint bands representing the proteins, but the levels of the protein are not sufficient. These data indicate that the BirA protein is unable to biotinylate proximal proteins *in vitro.*

4.3 Discussion

PrV has been the focus of a number of studies (Jiwaji *et al*., 2016; Jiwaji *et al*., 2019; Nakayinga, 2013; Pringle *et al*., 2003; Short *et al*., 2013; Walter *et al*., 2010) however little is known about the replication mechanisms employed by PrV, both in insect and in mammalian cells. The availability of the BioID system (Roux *et al*., 2012) provided the opportunity to address this limitation. The replication accessory protein p40 was selected as focus of this study because it is the best studied of the PrV proteins. A construct expressing BirA-fused p40 was generated and validated such that it could be used to identify p40-interacting proteins *in vivo.* While we were aware of previous research reporting that PrV was associated with detergent resistant membranes (Short *et al*., 2012) and our own bioinformatics analyses indicating that p40 was associated with membranes, we did not consider these to be limitations of sufficient magnitude to deter us from selecting p40 for these studies. We actually considered the fact that p40 was associated with the membrane a positive factor when using the BioID system; Roux *et al*. (2012) had used BioID to study insoluble and membrane associated proteins.

Our studies were hampered by the low levels of transfection efficiency; these were a result of the increasing sizes of the plasmids to express BirA, BirA-p40 and BirA-CNK. Despite the low transfection efficiency, BirA and BirA-p40 were shown to biotinylate themselves *in vivo*. This finding encouraged us to generate a stable cell line expressing BirA-p40 and we routinely detected myc-tagged BirA-p40 in the cells using confocal microscopy. In contrast, we did not detect the BirA-p40 protein with the same efficiency when performing western blot analyses. It is tempting to dismiss this as a non-specific or low sensitivity antibody. There is much to be said for more stringent validation of antibodies and also that antibodies be validated in a context and application specific manner (Älgenäs *et al*., 2014; Uhlen *et al*., 2016). There are differences in protein conformation and epitope accessibility depending on the technique used. For example, proteins are denatured before SDS-PAGE and western blot analyses; and they return to some variant of their initial confirmation after the removal of the detergents. In contrast, in ELISAs, proteins remain in their native conformation (Älgenäs *et al*., 2014; Uhlen *et al*., 2016). In our studies, the proteins analysed by western blot are irreversibly denatured and the proteins analysed by immunofluorescence are fixed by paraformaldehyde in their native conformation. Immunofluorescence also permits analysis on a single cell level so proteins can be visualized in individual cells, even at low concentrations. This highlights the importance of specificity and sensitivity required from the antibodies that are used (Älgenäs *et al*., 2014).

However, having considered the data collected during this project, we would suggest that the differential detection of p40 in confocal and western analysis can also be attributed to the nature of the protein that was the object of study. In this study, our bioinformatics analyses have indicated that p40 is likely to be membrane associated and it has two transmembrane domains. Previous reports have indicated that the replication complexes, which include p40, are associated with detergent resistant membranes (Short *et al.,* 2013). With these pieces of information, it is reasonable to detect higher levels in confocal microscopy because the proteins will be present in the cell, in their entirety. During SDS-PAGE and western analyses, proteins need to be extracted from their membranes and then they need to reform to reveal the epitopes recognised by the antibodies. This could explain why the western analyses detect lower levels of p40.

Our inability to perform expression and analyses studies with BirA-p40 *in vivo* led us to consider the *in vitro* expression of BirA-p40. We were able to express BirA-p40 at significantly higher levels *in vitro* compared to *in vivo.* It is interesting to note that the BirA-fused protein was unable to biotinylate itself *in vitro*. One possible reason could be that the BirA-fusion

protein was not correctly folded when expressed *in vitro.* This is contradicted by the report of Choi-Rhee *et al*. (2004) who studied the BirA (R11G mutant) *in vitro* and observed successful biotinylation. The natural binding partner of BirA, BCCP (biotin carboxylase-biotin carboxyl carrier protein) was biotinylated to capacity after 30 minutes however biotinylation of the acceptor protein RNase A was detected only after 2 hours and reached maximum biotinylation after 24 hours (Choi-Rhee *et al*., 2004). It is possible that the low levels of biotinlyation detected in the *in vitro* TnT reactions was due to the short incubation time (90 min). When the BirA-p40 was used in IP reactions however, the samples were incubated overnight. If time of incubation was a factor, we would have observed higher levels of biotinylation in IP experiments, and we did so. We detected biotinylated proteins that were pulled down by BirA-p40 in an IP reaction from a PrV-infected HeLa cell lysate. This reaction was incubated overnight, at 4 °C and so indicates elevated levels of biotinylation in samples that were incubated for longer.

In this chapter, we aimed to design a plasmid construct into which we could insert viral sequences of interest and use it to elucidate the role of the inserted viral sequence, *in vivo.* Our studies have highlighted our naivety, even with a known virus with a protein that has a known function, we faced challenges. These challenges are biological in nature; and highlight the short-sightedness of assuming an understanding of proteins, and so, of viral systems. Here, we have managed to develop and validate a tool for the *in vivo* or *in vitro* expression of a protein, and the *in vivo* and *in vitro* biotinylation of proteins of interest. With the understanding from this research, we can optimize the protocols to identify PrV-interacting proteins *in vivo.* In particular, if PrV forms replication complexes within enclosed structures on the mitochondrial membrane, like FHV, then this approach would permit a deeper understanding of where the replication complexes are situated and what host proteins are involved in the formation of these replication complexes.

Chapter 5: Enhancing the detection of PrV replication

The data in this chapter is being prepared for publication by the authors Jarvie, Zana and Jiwaji.

5.1 Introduction

Detergents

Detergents are amphipathic compounds that consist of a polar head group and a hydrophobic chain or tail (Seddon *et al*., 2004). The structure of the detergent molecules imbue them with unique properties; one of these properties is that they often form micelles spontaneously in the presence of an aqueous solution (Seddon *et al*., 2004). Detergents are used in a variety of biological applications including polyacrylamide electrophoresis (PAGE), membrane permeabilization, membrane dissolution, lipid raft extraction and protein crystallisation. Detergents can be classified into four distinct groups: ionic, non-ionic, bile acid salts and zwitterionic. Ionic detergents contain a charged head group that can either be anionic (e.g. SDS) or cationic (e.g. CTAB). Bile acid salts are also ionic detergents but contain a rigid steroidal backbone and as a result have a polar and apolar face instead of a defined head group (Linke, 2009; Seddon *et al*., 2004). Non-ionic detergents have an uncharged hydrophilic head and are considered to have mild and non-denaturing effects (Seddon *et al*., 2004). Zwitterionic detergents exhibit a combination of ionic and non-ionic properties. They contain both positive and negatively charged groups but have no net charge; and they are able to break protein-protein interactions (Seddon *et al*., 2004; Serva.de, 2019).

Detergents are defined by four main characteristics (Seddon *et al*., 2004; Serva.de, 2019): critical micelle concentration (CMC), aggregation number, average micellar weight and hydrophilic:lipophilic balance (HLB). The CMC is the minimum concentration of detergent required for individual detergent molecules to cluster together to form micelles. An average CMC is often reported as the formation of micelles is affected by conditions such as pH, ionic strength, temperature and the presence of proteins, lipids or other detergents. The CMC value decreases with the length of alkyl chains and increases with the number of double bonds and branch points present in the detergent. The aggregation number is the number of individual detergent molecules required to form a micelle and the average micelle weight is determined by multiplying the aggregation number by the molecular weight of the detergent. This gives an indication on the size of micelle formed by the detergent. The HLB value is the measurement of the relative hydrophobicity of a detergent and gives an indication of the detergent's ability to solubilize membrane proteins (Linke, 2009). The value is numerically calculated based on the structure of the detergent and can range from 0 to 40 (Linke, 2009). Most non-ionic detergents remain within the HLB range of 0-20. Non-ionic detergents with HLB values below

10 are not water soluble and are often used as anti-foaming agents; detergents greater than 16 are used as stabilizers and non-ionic detergents with values between 12 and 20 are preferred for non-denaturing solubilization.

Lipid rafts and detergent – resistant membranes

Lipid rafts are heterogenous and highly dynamic structures that are enriched with sterol and sphingolipid molecules (Pike, 2009; van Gestel *et al*., 2016). These regions are reported to enable the compartmentalization of cellular processes (Pike, 2009; van Gestel *et al*., 2016). Lipid rafts are often referred to as detergent resistant membranes, when in fact, detergent resistant membranes are artefacts and do not occur naturally in the cell (Brown, 2006; Lingwood and Simons, 2007).

Lipid rafts are often utilized by viruses for entry (*Papovaviridae*, *Reoviridae*, *Picornaviridae* and the *Adenoviridae*; Suzuki and Suzuki, 2006), in virus trafficking within the cell (*Filoviridae*, *Paramyxoviridae*; Bavari *et al*., 2002; Suzuki and Suzuki, 2006) and in virus budding (*Orthomyxoviridae*; Barman and Nayak, 2007). PrV replication accessory protein, p40, has been reported to associate with detergent resistant membranes in insect cell lines (Short *et al*., 2013). This report, in addition to our inability to express and detect p40 *in vivo* in HeLa cells, prompted us to investigate the effect of detergent treatment on PrV-infected mammalian cells. We hypothesized that the detergent would enter the cell and alter the intracellular membrane structures as well as the lipid rafts and in turn, have an effect on the PrV replication complexes. We expected enhanced solubilisation of p40 and wanted to use this to study PrV infection of mammalian cell lines.

5.2 Results

5.2.1 The selection of the detergents

A range of detergents were selected based on their biochemical characteristics (Figure 5.1), which are summarised in Table 5.1. SDS and CTAB are both ionic detergents that are composed of a charged, polar head group and non-polar tails. SDS is an anionic detergent and denatures proteins by breaking down the intra- and intermolecular interactions, generally resulting in irreversible denaturation (Seddon *et al*., 2004). CTAB is a cationic detergent and is often used in DNA isolation and to permeabilise cells (Serva.de, 2019). Non-ionic detergents, Tween 20, digitonin, saponin, NP-40 and Triton X-100 have uncharged, hydrophilic head groups (Seddon *et al*., 2004). Digitonin is a steroidal saponin that was isolated from *Digitalis purpurea* (foxglove). Saponin is found in a number of plant families (Podolak *et al*., 2010). Both of these plant-isolated detergents have complex chemical structures (Figure 5.1); the level of purity between preparations varies and as a result these

two detergents may contain other, unknown compounds. NP-40 was originally produced by Shell Chemical Company but is no longer produced (Sinha *et al*., 2017). There are now commercially available substitutes available, but they are much more potent and require dilutions (Sinha *et al*., 2017). NP-40, Tween 20 and Triton X-100 are all mixtures of compounds and so, there remains some variability between preparations as well as their biochemical characteristics. The ionic detergents (SDS and CTAB) are considered more effective in disrupting protein-protein interactions whereas the non-ionic detergents, such as Triton X-100, are considered better for disrupting lipid-protein and lipid-lipid interactions as they are considered to be mild and non-denaturing detergents (Seddon *et al*., 2004).

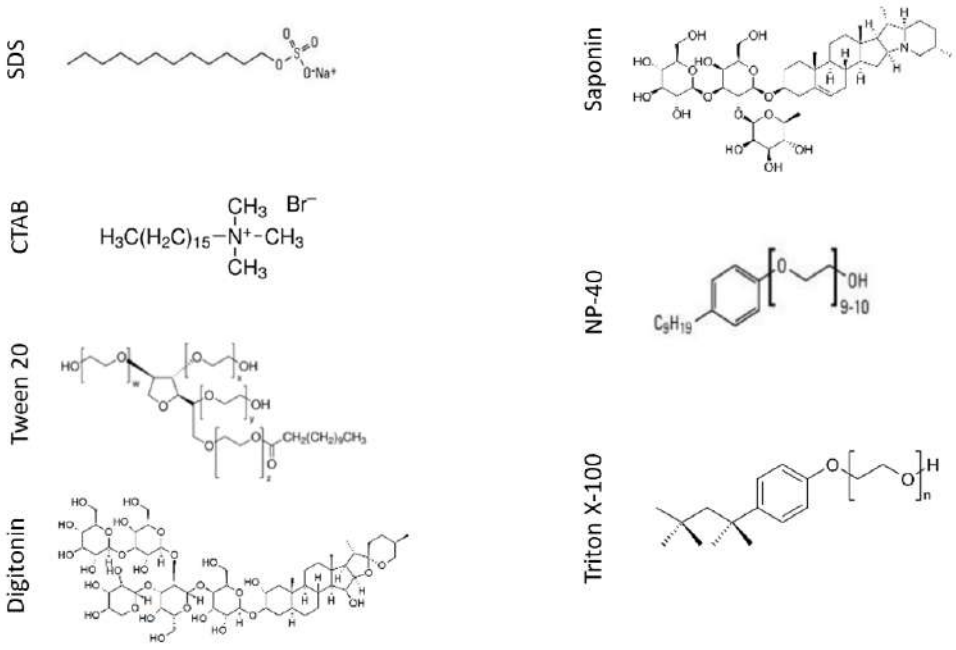


Figure 5.1: Chemical structures of the detergents.

Table 5.1: Biochemical characteristics of the detergents.[1]

Detergent		Nature	MW anhydrous	CMC (mM) at 25 °C	Aggregation number	Average micellar weight	HLB
SDS	Sodium dodecyl sulfate	Anionic	288.4	7-10	62	18 000	40
CTAB	Cetyl methyl ammonium bromide	Cationic	364.5	0.92	170	62 000	15.8
Tween 20	Polyethylene glycol sorbitan monolaurate	Non-ionic	1228	0.059	~58	70 000	16.7
Digitonin		Non-ionic	1229.3	0.67 – 0.73	60-70	70 000	0.4
Saponin		Non-ionic	634.851	0.8-1.2	~50	30 000	14.3
NP40 substitute	Nonylphenyl-polyethylene glycol	Non-ionic	573 for n=8	0.05-0.3	~132	~90 000	~13.1
Triton X-100	Polyethylene glycol tert-octylphenyl ether	Non-ionic	625 (average)	0.2-0.9	100-155	80 000	13.5

[1] Information derived from Lozsan *et al*. (2017); Mitra and Dungan (1997); Pan *et al*. (2009); Pasquali *et al*. (2008); Serva.de (2019); Seddon *et al*. (2004) and Sinha *et al*. (2017).

5.2.2 The effect of detergent treatment on PrV-infected mammalian cells

PrV-infected HeLa cells were plated on glass coverslips and permitted to settle overnight. The live cells were treated with S buffer containing 0.1 % of the following detergents (Figure 5.2); SDS, CTAB, Tween 20, digitonin, saponin, NP-40 and Triton X-100. Detergent-treated cells were prepared for confocal microscopy; the protocol included fixation, permeabilization and detection using dsRNA- and PrV-specific antibodies. The cells were visualized on the confocal microscope; the level of dsRNA and p40 fluorescence in the S buffer-treated cells was normalized to 100 % and the other samples were compared to the buffer-treated cells (Appendix A3, Figure A3).

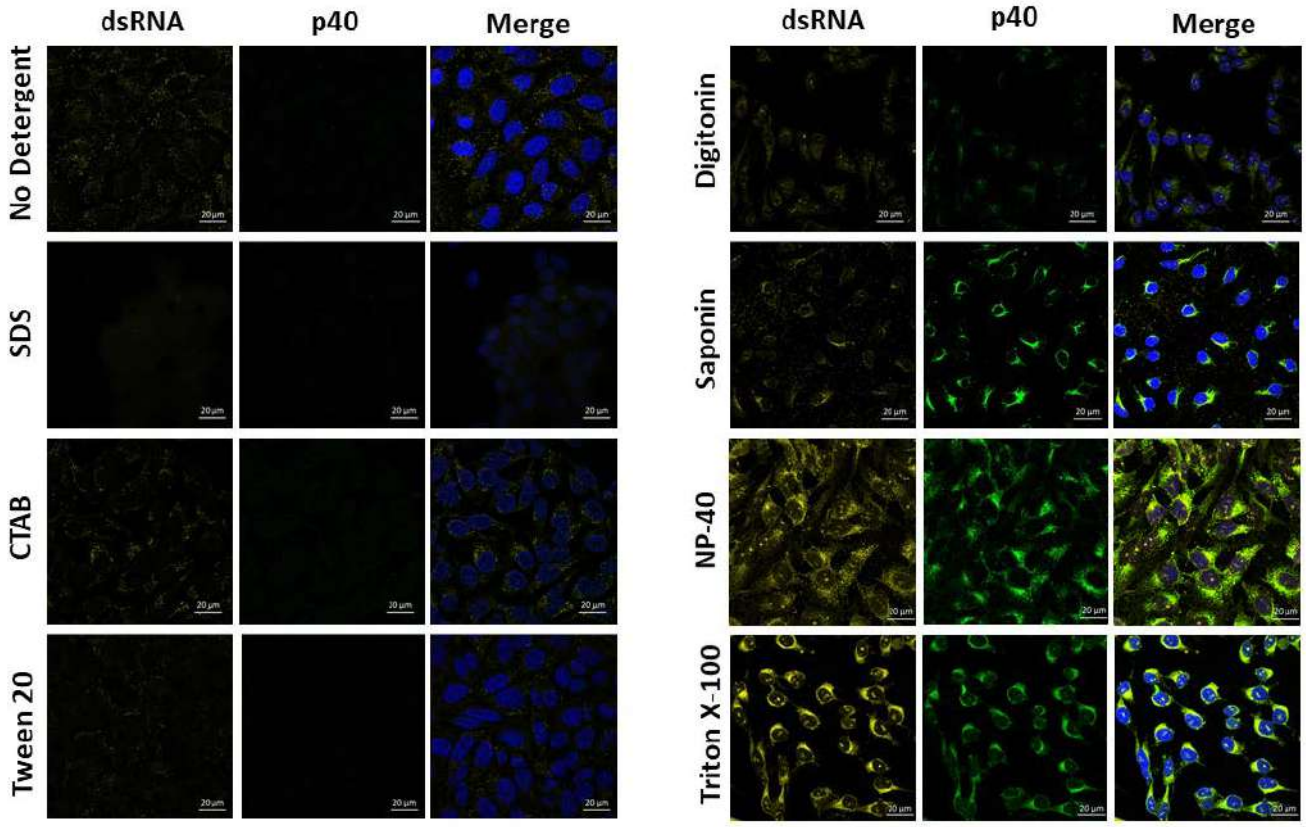


Figure 5.2: PrV-infected HeLa cells treated with a range of detergents. HeLa cells were plated on glass coverslips and permitted to settle overnight. The cells were treated with S buffer containing 0.1 % of each detergent (SDS, CTAB, Tween 20, Digitonin, Saponin, NP-40 and Triton X-100) for 15 minutes at 37 °C. The panel referred to as no detergent represent the S buffer-treated control cells. The cells were permitted to recover in fresh medium for 15 minutes at 37 °C and then subsequently prepared for confocal microscopy. The cells were fixed with 4 % paraformaldehyde and then probed with the following antibodies: IgG-biotin anti-p40 with streptavidin AF 488, anti-dsRNA (mouse) and goat anti-mouse AF 546. The nuclei were stained with DAPI. The scale bar represents 20 µm.

When the cells were treated with the ionic detergents (SDS and CTAB), they become viscous. It was concluded that the treatment with SDS and CTAB had resulted in the denaturation of the cellular DNA (Tan and Yiap, 2009). This was corroborated by the visibly compromised nuclei and the reduced staining with DAPI (Figure 5.2). SDS treatments showed no detectable fluorescence of dsRNA or p40 when compared to the untreated cells. This was further supported when the fluorescence was quantified and found that SDS treatment resulted in 94 % dsRNA and 114 % p40 fluorescence (Appendix A3, Figure 3A). Despite the visibly compromised nuclei when the cells were treated with CTAB, the dsRNA fluorescence detected increased to 228 % while the level of p40 detected remained relatively unchanged at 124 % (Appendix A3, Figure 3A). When the cells were treated with Tween 20, there was no visible increase in the dsRNA nor p40 fluorescence and this was supported when the fluorescence was quantified (117 % dsRNA and 104 % p40; Appendix A3, Figure 3A). The treatment of the PrV-infected HeLa cells with digitonin resulted in a two-fold increase in both dsRNA and p40 (210 % and 193 %, respectively); on the other hand, saponin-treated cells resulted in a visibly noticeable increase in p40 fluorescence (Figure 5.2). When quantified, the saponin treatment resulted in the increase of dsRNA fluorescence to 184 % and p40 fluorescence to 548 % (Appendix A3, Figure 3A). What was interesting to note, digitonin and saponin have comparable CMC values of 0.67 – 0.73 and 0.8 – 1.2, respectively but differ greatly in the HLB value (Table 5.1). Digitonin has a HLB value of 0.4 which makes it a hydrophobic detergent, whereas saponin has a HLB value of 14.3 which means that it is an effective non-denaturing solubilizing detergent.

Treatment of the PrV-infected HeLa cells with NP-40 and Triton X-100 resulted in a dramatic increase in fluorescence for both dsRNA and p40 (Figure 5.2). When the fluorescence was quantified, NP-40 resulted in 408 % and 631 % increase in dsRNA and p40, respectively, when compared to the untreated control (Appendix A3, Figure 3A). The cells treated with both NP-40 and Triton X-100, dsRNA and to a lesser extent, p40, was detected in the nucleoli of the cells. The DAPI fluorescence was enhanced when the cells were treated with both saponin and Triton X-100 (Figure 5.2). Triton X-100 has been reported to preserve nuclear complexes as well as permeabilise both faces of the nuclear membrane (Griffis *et al*., 2003; Scheer *et al*., 1976); this would enable greater accessibility for DAPI to stain the genomic DNA. There have been no reports of saponin resulting in a similar effect, but the data presented in Figure 5.2 suggests that saponin enables increased staining of the nuclei through increased permeabilization of the nuclear membrane.

NP-40 and Triton X-100 both have low CMC values, 0.05 – 0.3 and 0.2 – 0.9, respectively (Table 5.1); Tween 20 also has a low CMC value of 0.059 but differs greatly in chemical

structure when compared to NP-40 and Triton X-100, which both have a less complex and a more linear structure (Figure 5.1). The HLB value of NP-40 and Triton X-100 are 13.1 and 13.5, respectively, which means that they are both effective non-denaturing and solubilising detergents. The detection of elevated levels of dsRNA- and p40-specific fluorescence led us to query whether the same would occur in other mammalian cell lines.

5.2.3 Triton X-100 enhances the detection of dsRNA in MCF-7 and HEK293 cells

We had hypothesized that the addition of a detergent to PrV-infected HeLa cells would result in the enhanced detection of the PrV-specific proteins. Our data supported our hypothesis and we observed a dramatic increase in dsRNA- and p40-specific fluorescence after NP-40 and Triton X-100 treatment of PrV-infected HeLa cells. It is interesting that a detergent with a low CMC value, high HLB value and a linear structure (e.g. Triton X-100 and NP-40) resulted in the enhanced fluorescence whereas another detergent with a similar CMC and HLB value but a more bulky structure did not induce the same effect (Tween 20, Table 5.1).

To determine whether we could reproduce the enhanced detection of PrV-specific proteins and dsRNA in other mammalian cell lines, PrV-infected MCF-7 and HEK293 cells were treated with Triton X-100 and Tween 20, two detergents with similar characteristics but different structures and differing effects when used to treat PrV-infected HeLa cells. MCF-7 and HEK293 cells were plated on glass coverslips and permitted to settle overnight. The cells were then treated with 0.1 % Tween 20 or Triton X-100 for 15 minutes. The cells were permitted to recover in fresh growth medium for 15 mins and after the recovery period, the cells were fixed with paraformaldehyde and prepared for confocal microscopy (Figure 5.3, uninfected control cells included in Appendix A6). The cells were probed with antibodies directed against dsRNA to assess the effect of the detergent treatment on the PrV replication complexes.

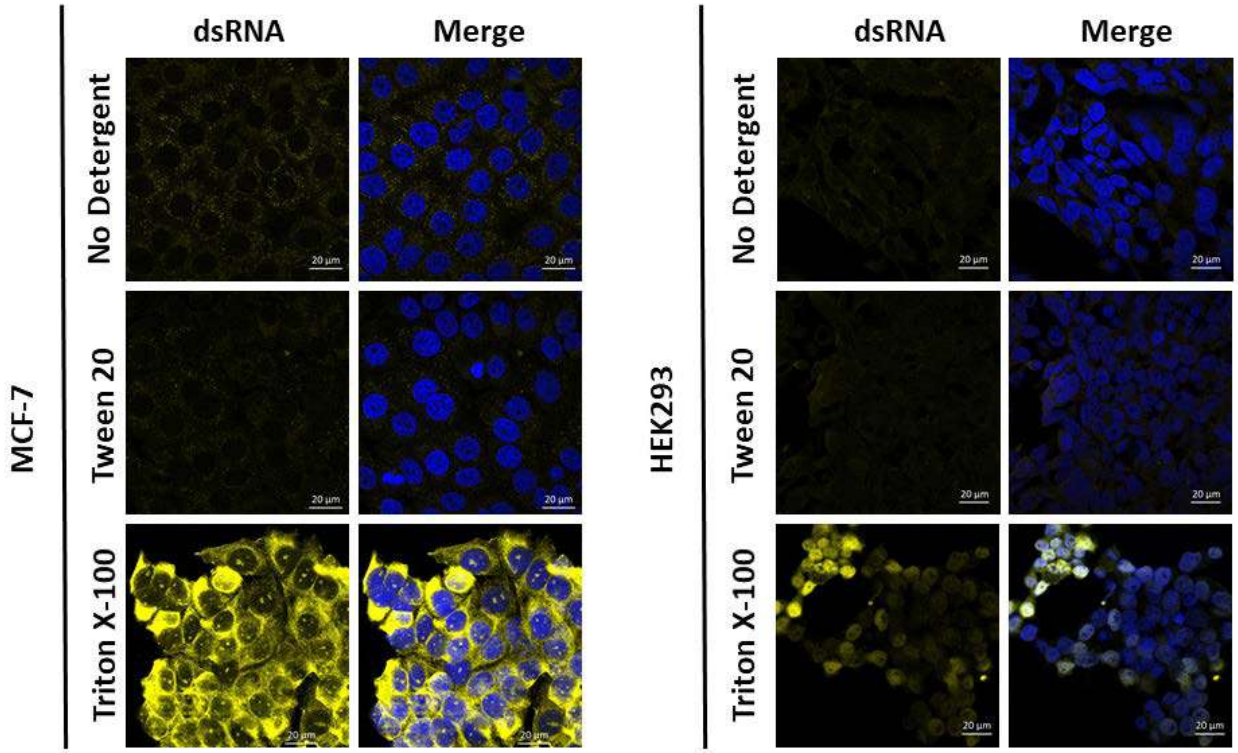


Figure 5.3: Analysis of PrV-infected MCF-7 and HEK293 after detergent treatment. MCF-7 and HEK293 cells were plated on glass coverslips and permitted to settle overnight. The cells were then treated with 0.1 % Tween 20 or Triton X-100 (in S buffer) for 15 minutes at 37 °C. The cells were permitted to recover in fresh medium for 15 minutes at 37 °C and were then subsequently prepared for confocal microscopy. The cells were probed with anti-dsRNA (mouse) with goat anti-mouse AF 546. The nuclei were stained with DAPI and the scale bar represents 20 µm.

The treatment of both MCF-7 and HEK293 cells with Tween 20 resulted in no significant change in the level of dsRNA detected in the PrV-infected cells when compared to the buffer-treated control cells (no detergent; Figure 5.3). When the cells were treated with Triton X-100, the detection of dsRNA was dramatically increased in both MCF-7 and HEK293 cell lines. The data presented here (Figure 5.3) corroborates the data collected in PrV-infected HeLa cells (Figure 5.2). It is also interesting to note that the localisation of dsRNA to the nucleolar region of the nuclei of Triton X-100-treated PrV-infected HeLa cells (Figure 5.2) was also observed in the PrV-infected MCF-7 cells. While we did not observe the same phenomenon in Triton X-100-treated HEK293 cells, we attributed this lack to poor sample preparation. The HEK293 cells could not be retained on the glass coverslips after treatment with Triton X-100; the few cells that remained after treatment were visualised but they were collected at the edges. Most of the cells had sloughed off the glass coverslip during the treatment with Triton X-100 or during the subsequent recovery period.

5.2.4 Enhanced detection of dsRNA fluorescence is dependent on the detergent concentration

We have reported that the treatment of PrV-infected mammalian cells with 0.1 % Triton X-100 resulted in an increase in the detection of dsRNA (Appendix A3, Figure A3, Figure 5.2 and 5.3) and p40 (Appendix A3, Figure A3 and 5.2) in HeLa cells. We also observed that the treatment of cells with detergent at a concentration of 0.1 % resulted in a large number of the cells lifting off the glass coverslip. This was particularly noticeable when we worked with HEK293 cells but was observed to a lesser degree with HeLa and MCF-7 cells as well. This led us to query what effect the detergents had on the PrV-infected mammalian cells and whether the effect was dependent on the concentration of the detergent. For this reason, we treated PrV-infected HeLa cells with a range of Triton X-100 concentrations (Figure 5.4). We subjected these detergent-treated cells to a resazurin-based cell viability assay and also prepared cells for confocal microscopy (Figure 5.4).

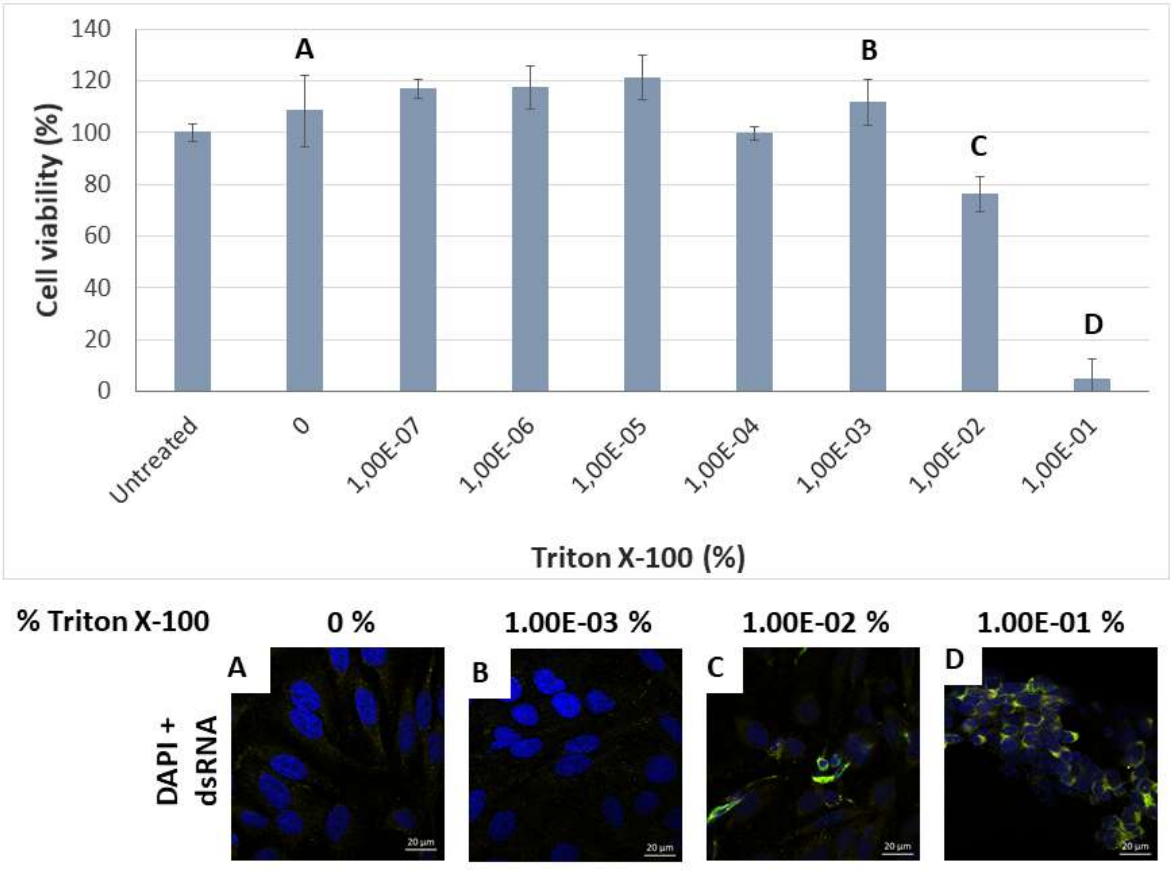


Figure 5.4: Analysis of cell viability after treatment with a range of Triton X-100 concentrations. PrV-infected HeLa cells were treated with Triton X-100 at a range of concentrations for 15 minutes and then permitted to recover in fresh medium for 15 minutes. The cells were treated with either S buffer (0) or with increasing concentrations of Triton X-100, with the maximum at 0.1 %. The cell viability of the treated cells was measured after 12 hours using a resazurin-based *in vitro* toxicology assay kit. The viability of the untreated cells

was normalized to 100 %. Parallel samples were also prepared for confocal microscopy, as described previously.

The buffer-treated cells (0) were normalized to 100 %; the viability of cells treated with increasing concentrations of Triton X-100 were compared to the untreated cells. We observed that at 0.1 %, the detergent-treated cells were not viable (Figure 5.4). As the concentration of Triton X-100 decreased, the cell viability increased until it was comparable to the S buffer-treated HeLa cells. When the Triton X-100-treated cells were examined by confocal microscopy, the cells treated with 0.1 % Triton X-100 showed elevated levels of dsRNA- and p40-specific fluorescence (Figure 5.4). The same effect was observed, but to a lesser degree, when the PrV-infected cells were treated with 0.01 % Triton X-100. Below this concentration, Triton X-100 did not have a detectable impact on the detection of PrV fluorescence in infected HeLa cells (Figure 5.4). We therefore observed a clear parallel between the concentration of Triton X-100 and cell viability and between the detergent concentration and the enhanced detection of PrV infection. To evaluate the impact of Triton X-100 on the ultrastructure within PrV-infected HeLa cells, we prepared detergent-treated cells for TEM (Figure 5.5).

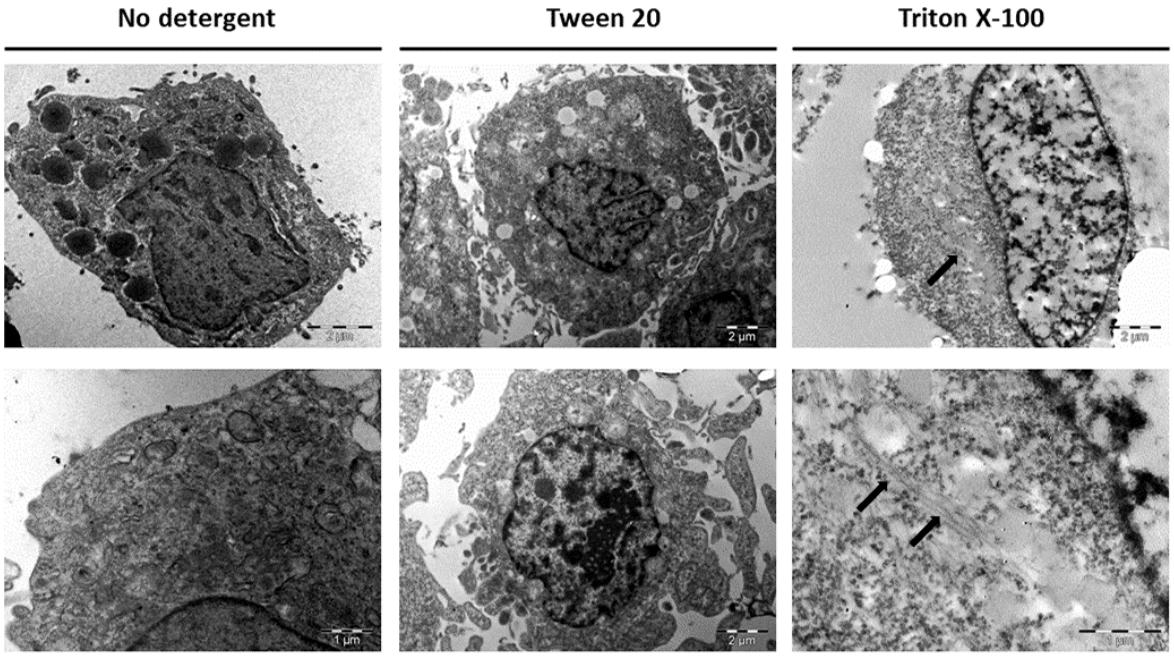


Figure 5.5: TEM analysis of Tween 20 and Triton X-100-treated HeLa cells. The PrV-infected HeLa cells were treated with S buffer (no detergent) or S buffer containing 0.1 % Tween 20 or Triton X-100. The cells were fixed in 2.5 % glutaraldehyde, stained with 1 % osmium tetroxide and dehydrated in ethanol. The cells were then embedded in resin. 100 nM sections of the embedded samples were stained with 5 % uranyl acetate. The sections were visualised using a Zeiss Libra 120. The black arrows in the Triton X-100 treated cells highlight the membrane-like structures.

PrV-infected HeLa cells treated with S buffer (no detergent) remained intact with no noticeable changes to the cellular structures. When the PrV-infected cells were treated with Tween 20, the cells showed some loss in cellular definition but not significantly when compared to the control cells. In contrast, treatment of PrV-infected HeLa cells with Triton X-100 resulted in the complete loss of cytosolic definition. We could identify lipid droplets in the cytosolic region of Triton X-100-treated cells. In addition, there were long strings of membrane-like structures visible. Some of these membranous structures were in proximity to the nucleus but most were within the cytosolic space. In the Triton X-100-treated cells, the nucleus looked swollen and the nuclear material appeared to be fragmented. This would explain lack of cell viability when 0.1 % Triton X-100-treated cells were subjected to the resazurin-based cell viability assay.

The membrane-like structures, highlighted by the black arrows, detected in the cytosolic region of the Triton X-100-treated cells were particularly interesting (Figure 5.5). It is possible that these may be what authors refer to as detergent resistant membranes. van Gestel *et al*. (2016) reported the isolation of detergent resistant membranes and the structures present in their TEM images do bear some resemblance to the structures we detect in PrV-infected HeLa cells. With more definitive research to quantify the lipid loss before and after Triton X-100 treatment, it is not possible to conclusively state that the structures we observed represent detergent resistant membranes.

5.2.5 Further investigation into the effect of detergent treatment on PrV

Thus far, we have demonstrated a link between the treatment of PrV-infected mammalian cells with Triton X-100 and the enhanced detection of PrV infection (Figures 5.2 and 5.3). Treatment with Triton X-100 also resulted in a decrease in cell viability (Figure 5.4) as well as a loss in cellular definition (Figure 5.5). Considering these factors, we were interested in understanding how the treatment of PrV-infected cells with Triton X-100 resulted in the enhanced detection of dsRNA and p40 in detergent-treated cells. One hypothesis was that detergent-treatment resulted in the disruption of virus replication complex-containing membranous structures and so resulted in the exposure of the entities of interest, namely dsRNA and p40, and so the enhanced detection.

To determine if this was the case, PrV-infected HeLa cells were transfected with a plasmid encoding the pEGFP-N1 protein. These cells were treated with S buffer or 0.1 % Tween 20 or Triton X-100 and permitted to recover. Total RNA was extracted from the treated cells and used to prepare cDNA. The cDNA was used in PCR reactions with primers to specifically amplify p40, EGFP and enolase (Figure 5.6). EGFP is soluble fluorescent protein that does

not show a specific localization pattern within the cell and is often used in protein tracking and protein localization studies (Seibel *et al.,* 2007). Enolase is a key enzyme in the glycolytic pathway and has been reported to be distributed throughout the cytosol of HeLa cells (Johnstone *et al*., 1992). These two RNAs and proteins serve as controls within the experiment; neither of these RNAs or proteins are reported to be associated with intracellular membranes.

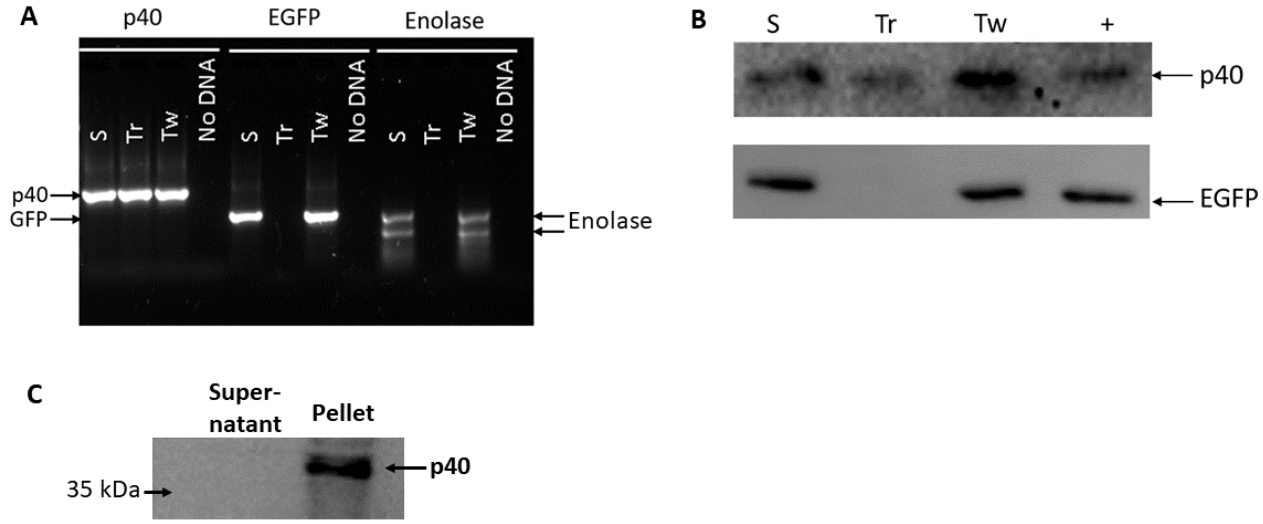


Figure 5.6: Effect of Triton X-100-treatment on the detection of PrV RNA and protein. (A) PrV-infected HeLa cells were transfected with pEGFP-N1 and then treated with either S buffer (S), 0.1 % Triton X-100 (Tr) or 0.1 % Tween 20 (Tw). The total RNA was extracted and random hexamers were used in reverse transcription reactions. The cDNA was PCR-amplified using primers specific to p40 (JRS79 and JRS80), EGFP (EGFP_F and EGFP_R) and enolase (Enolase_F and Enolase_R). **(B)** The proteins isolated from PrV-infected and pEGFP-N1 transfected HeLa cells were separated by SDS-PAGE and analysed using western blot analysis. The membranes were probed with IgG-biotin anti-p40 and anti-GFP (mouse). The positive control, represented by a +, is a lysate from MG8 cells that were persistently infected with PrV. **(C)** PrV-infected HeLa cells were lysed and the soluble (supernatant) and insoluble (pellet) fractions were collected. The fractions were analysed by western blot and the membrane was probed with IgG-biotin anti-p40 antibodies.

Total cDNA prepared from EGFP-expressing and PrV-infected HeLa cells and treated with S buffer or Tween 20 was probed for enolase, EGFP or p40. All three transcripts were detected in cDNA prepared from S buffer- or Tween 20-treated HeLa cells. In contrast, while p40 was detected in EGFP-expressing and PrV-infected HeLa cells treated with Triton X-100, no transcripts were detected for enolase or EGFP (Figure 5.6A). Both EGFP and enolase mRNA are translated in the cytosol; these results suggest that the treatment with Triton X-100 resulted in the loss of cytosolic mRNA. This data suggest that in contrast to the cytosolic mRNAs, the PrV mRNA was protected from the Triton X-100 treatment.

When proteins from the EGFP-expressing and PrV-infected HeLa cells were subjected to western analysis (Figure 5.6B), a similar trend was observed. p40 protein was detected in all the samples, irrespective of the treatment. In contrast, EGFP was not detected in the Triton X-100-treated cells. Again, PrV p40 appears to be protected during the treatment with Triton X-100 whereas the soluble and non-membrane associated EGFP was lost as a result of the treatment with Triton X-100. Interestingly, the concentration of p40 detected in the Tween 20-treated cells was higher than the levels that were present in S buffer and Triton X-100-treated cells (Figure 5.6B). This is in contradiction to the immunofluorescence data (Appendix A3, Figure A3, 5.2 and 5.3), which suggests that treatment with Tween 20 does not enhance the detection of p40. It is possible that this difference represents inter-experimental variability. Alternatively, the handling, processing and analysis of detergent-treated samples poses some technical challenges when performing western analysis. It is possible that this could have resulted in the variation observed. Based on the data presented in the previous chapter, it also highlights that antibodies have different sensitivities when using confocal microscopy and western analyses and this could be another example of this differential technique-dependent detection of the same target.

To further investigate whether the lack of detection of p40 during western blot was due the association of p40 with cellular membranes, PrV-infected HeLa cells were lysed and then separated into soluble (supernantant) and insoluble (pellet) fractions via centrifugation (Figure 5.6C). The supernantant and pellet were analysed by western blot and the membranes were probed with anti-p40 antibodies. It was found that p40 protein was only detected in the pellet fraction and not in the supernatant, thus confirming that the lack of detection of p40 was due to the cell debris (pellet) fraction being discarded.

Taken together, these data indicate that the treatment of EGFP-expressing and PrV-infected HeLa cells with Triton X-100 results in intracellular changes that cause the loss of cytosolic RNA and proteins and the retention of PrV proteins and dsRNA. If considered with the data collected on the TEM, the enhanced detection is a result of intracellular changes, and possibly in the membranes within the PrV-infected HeLa cells.

5.2.6 Triton X-100 treatment of stable BirA-p40 expressing HeLa cells

Earlier in this document, we generated a stable BirA-p40 expressing HeLa cell line and demonstrated that we were unable to efficiently detect the BirA-p40 protein using western analysis. We attributed this lack of detection to two key factors; firstly, that p40 was membrane-associated and therefore was not readily available for analysis. And secondly, that the

antibodies we were using exhibited varying sensitivity and specificity in different techniques, namely confocal microscopy and western analysis. With the ability to enhance the detection of PrV-specific proteins in infected HeLa cells, we treated stable BirA-p40 expressing HeLa cells with 0.1 % Triton X-100 and prepared the cells for confocal microscopy (Figure 5.7). The cells were probed with antibodies targeted towards the myc epitope tag and towards p40.

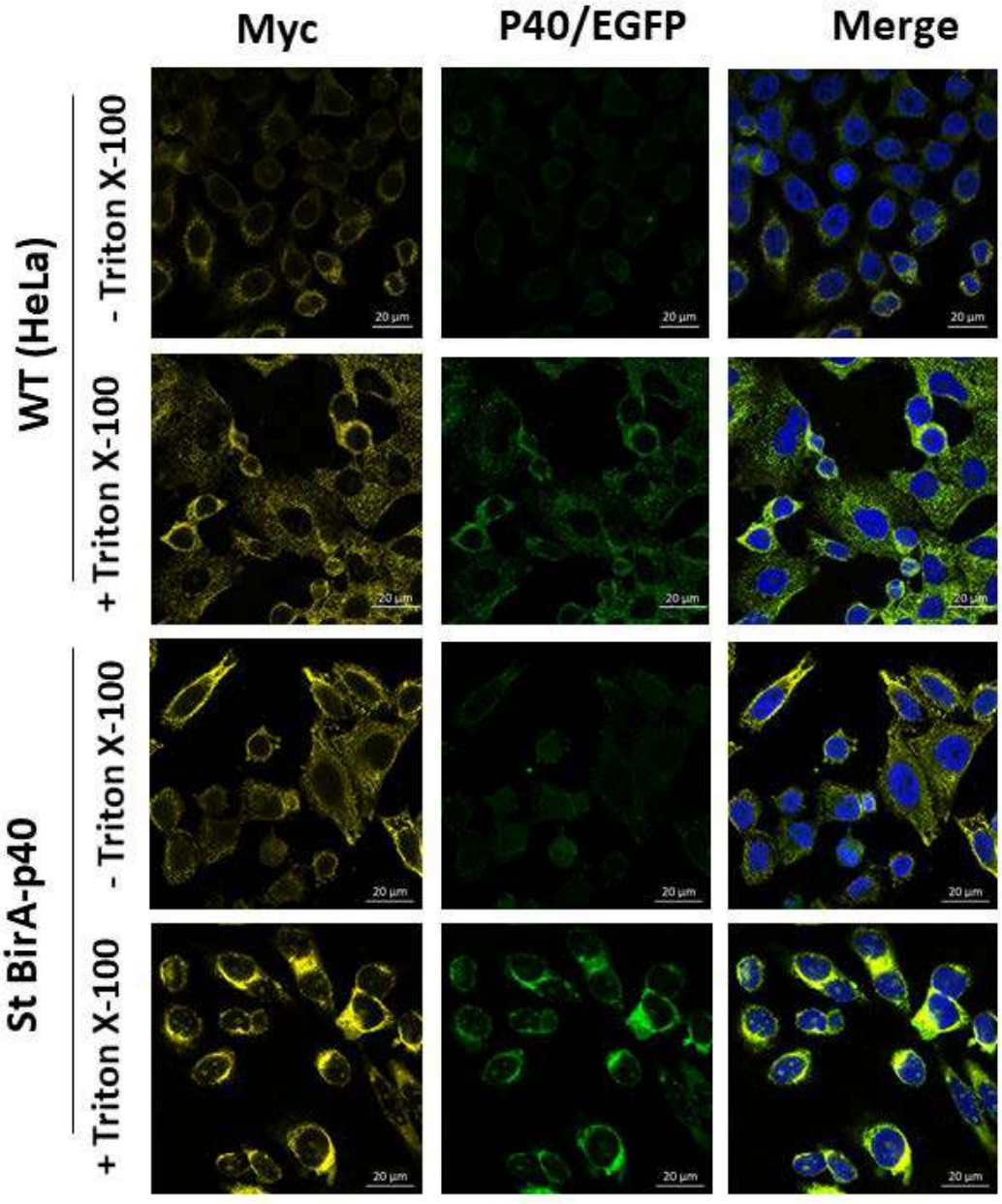


Figure 5.7: Treatment of a stable BirA-p40-expressing HeLa cell line with Triton X-100. PrV-infected HeLa cells (WT HeLa) and stable BirA-p40 (St BirA-p40) expressing cell line were treated with S buffer (- Triton X-100) or S buffer containing 0.1 % Triton X-100 (+ Triton X-100) for 15 minutes. The cells were permitted to recover for a further 15 minutes before they were prepared for confocal microscopy. The cells were probed with anti-myc (mouse) and

IgG-biotin anti-p40 primary antibodies and goat anti-mouse AF 546 and streptavidin AF 488 secondary antibodies. The nuclei were stained with DAPI and the scale bar represents 20 µm.

HeLa wild-type cells, infected with PrV, showed negligible levels of p40 in the untreated cells and enhanced detection of p40 when cells were treated with 0.1 % Triton X-100 (Figure 5.7). Wild-type HeLa cells also showed fluorescence when they were probed with anti-myc primary antibodies; this signal was due to the endogenous myc protein from which the myc epitope tag was derived (Brizzard, 2008). Surprisingly, we detected enhanced levels of fluorescence signal attributed to myc when HeLa cells were treated with 0.1 % Triton X-100 (Figure 5.7). This was unexpected and led us to query the localization of myc in the cell. It was interesting to note that c-myc, from which the sequence of the epitope tag is derived, is predicted to be membrane associated (Proteinatlas.org., 2019).

The stable BirA-p40-expressing cell line expressed both myc epitope-tagged BirA-p40 as well as EGFP. When the cells were treated with S buffer, in the absence of detergent, the fluorescence in the myc channel represents that of endogenous myc as well as the myc-BirA-p40 protein and the fluorescence in the p40/EGFP channel represents the EGFP generated from the chromosomally integrated myc-BirA-p40 DNA (Figure 5.7). The fluorescence detected in the AF 546 channel was higher in the stable BirA-p40 cell line compared to the wild-type HeLa cell line. This indicated that the stable cell line was expressing higher levels of the myc-tagged BirA-p40 protein. Treatment of the stable BirA-p40-expressing HeLa cell line with Triton X-100 results in the enhanced detection in both the AF 546 channel as well as in the AF 488 channel (Figure 5.7). We observed previously that Triton X-100 treatment causes the loss of the EGFP protein from the detergent treated cell (Figure 5.6), therefore the fluorescence in the AF 488 channel can be attributed to PrV p40. The fluorescence in the AF 488 channel overlaps exactly with the fluorescence in the AF 546 channel, indicating the presence of myc-BirA-p40 co-localizing with PrV p40 (Figure 5.7). As the antibodies are recognizing a myc-BirA-p40 and p40, it was reasonable for the signals to overlap. It is however interesting to observe how the signal in the two channels was focused in a specific region of the cell (Figure 5.7).

The co-localized fluorescence in the AF 488 and AF 546 channels was particularly interesting; the structures that the fluorescence formed appeared to be almost ‘nest-like’. Stable BirA-p40 expressing HeLa cells were prepared for TEM to determine whether the ‘nest-like’ structures could be visualized and detected with anti-p40 antibodies (Figure 5.8). HeLa cells (WT) and stable birA-p40 expressing cells were treated with 0.1 % Triton X-100 for 15 minutes and then permitted to recover before preparation for TEM. The cells were fixed, stained and embedded

in resin. Sections were cut from the embedded cells were incubated with IgG-biotin anti-p40 antibodies and anti-biotin quantum dots before visualisation.

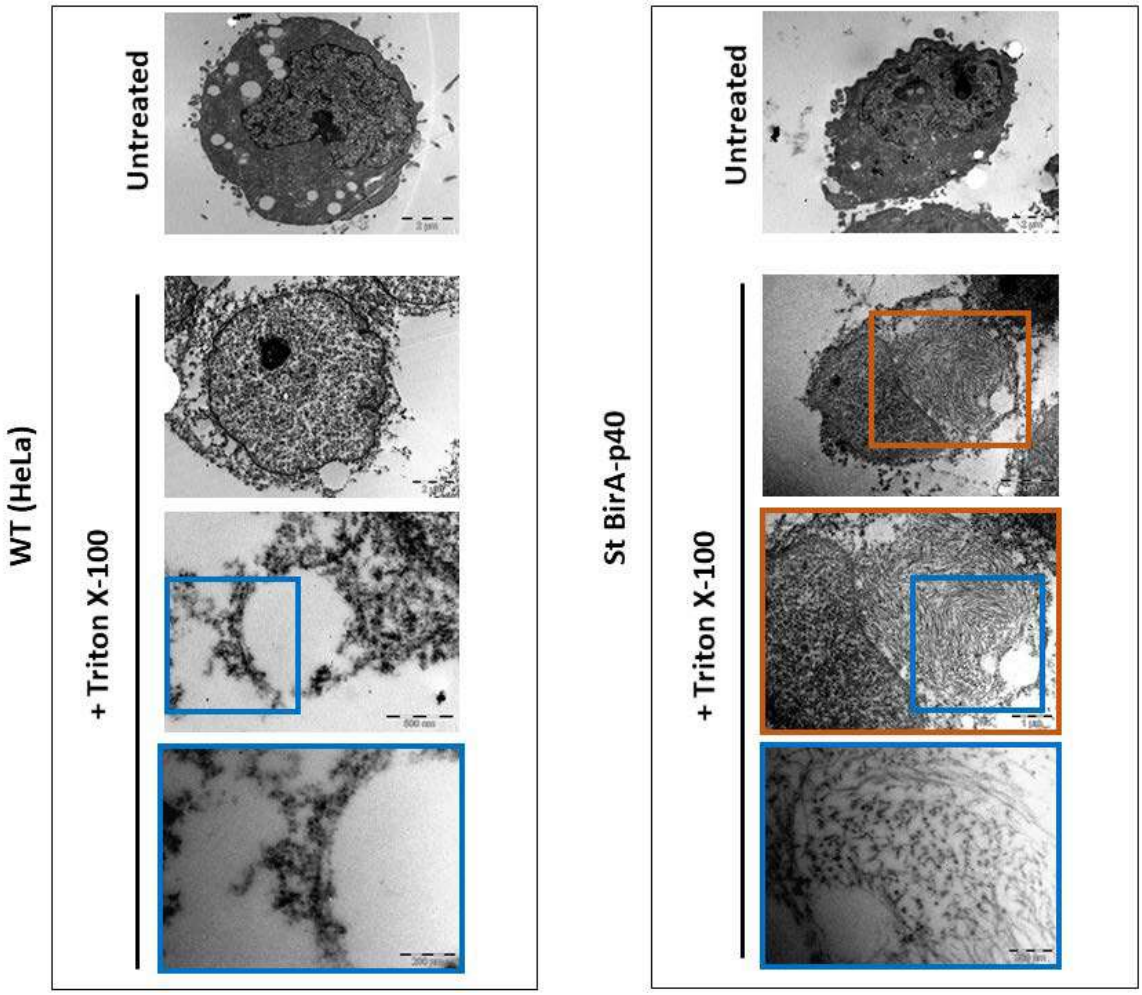


Figure 5.8: TEM analysis of stable BirA-p40 expressing HeLa cells treated with Triton X-100. Wild-type and stable BirA-p40 expressing HeLa cells were treated with 0.1 % Triton X-100 for 15 minutes. The cells were prepared as described in the methods and then embedded in resin. Slices 40 µm thick were cut on an ultramicrotome and stained with uranyl acetate. These slices were probed with IgG-biotin anti-p40 antibodies paired with streptavidin-colloidal gold conjugate.

HeLa cells and stable BirA-p40-expressing cells treated with S buffer showed structural definition within the cells (Figure 5.8). After the treatment with Triton X-100, both cell types lost cellular definition, and the nucleus appeared swollen and the nuclear contents looked fragmented (Figure 5.8). There were some interesting structures that looked membranous in the cells treated with 0.1 % Triton X-100 (Figure 5.8). There were more of these structures present in the BirA-p40-expressing cell line than in the wild-type HeLa cells. These structures looked ‘nest-like’ and were reminiscent of the structures detected on the confocal microscope (Figure 5.7). The cells embedded in resin were probed with IgG-biotin anti-p40 primary antibodies and with streptavidin-colloidal gold conjugate as the secondary antibody. Here the

colloidal gold was expected to aggregate in regions where the primary antibody was localized and highlight areas of p40 in the HeLa cell. We detected colloidal gold precipitation in the region around the membranous structures in both wild-type and stable BirA-p40-expressing HeLa cells (Figure 5.8).

These data led us to hypothesize that the membranous structures detected in HeLa cells, be they wild-type or BirA-p40-expressing cells, were 'detergent resistant membranes' and that the 'nest-like' structures we detected were a visual representation of these detergent resistant structures (highlighted by the blue square, Figure 5.8). The fact that the colloidal gold-labelled antibodies aggregated in the vicinity of these membranous structures indicated that these structures contained PrV p40, and that the higher levels of colloidal gold in the BirA-p40 expressing HeLa cell line were due to the higher levels of BirA-p40 in the membranes of these stable p40 expressing HeLa cells.

5.3 Discussion

The enhanced detection of dsRNA and p40 in detergent-treated PrV-infected cells could represent increased levels of replication or alternatively enhanced detection. Our data demonstrate that treatment with detergents with a low CMC value, high HLB value and a linear structure (namely Triton X-100 and NP-40) resulted in the enhanced detection of PrV replication. This is the first report where the treatment of persistently infected mammalian cells with detergents results in the enhanced detection of viral replication, *in vivo.*

The critical factor here is that the PrV-infected cells were treated with detergent prior to fixation and permeabilization steps, during the preparation of samples for confocal microscopy. Fixation for confocal microscopy entails the addition of paraformaldehyde, which cross-links proteins within the cell (Thavarajah *et al*., 2012). Permeabilization of the cell membrane occurs when 0.1 % Triton X-100 is added to fixed cells. However, once the cells are fixed, the possibility of intracellular structural rearrangements is low. This then means that the detergent-induced rearrangements are critical for the enhanced detection of PrV replication within mammalian cells. We hypothesize that these structural rearrangements result in the exposure of epitopes that are recognised by the antibodies, resulting in enhanced detection.

In insect cells, PrV replication is reported to be associated with detergent resistant membranes (Short *et al*., 2013). It is important here to highlight the difference between detergent resistant membranes and lipid rafts. Lipid rafts are most commonly defined as lipid-ordered micro-domains (between 10 and 200 nm) that are highly dynamic and are enriched with cholesterol and sphingolipids (Pike, 2009; van Gestel *et al*., 2016). These micro-domains can be isolated

using a treatment of 1 % Triton X-100 at 4 °C (van Gestel *et al*., 2016). It is difficult to separate the terms lipid rafts and detergent resistant membranes; lipids rafts are often detergent resistant membranes due to their biochemical composition (Heerklotz, 2002; Wang and Schey, 2015). However, as mentioned previously, detergent resistant membranes are artefacts and do not occur naturally in the cell. It is also important to note that the bioinformatic analysis of p40 indicates it is an integral membrane protein and associated with the mitochondrial membrane. It is therefore possible that a robust association between p40 and the mitochondrial membrane in mammalian cells may result in the detergent resistant phenotype observed in the PrV-infected HeLa cells.

Bioinformatic analyses predicted transmembrane domains within PrV p40 and the association of this protein with membranes within HeLa cells. The observation, using confocal microscopy and TEM, that increasing the level of p40 protein within the cell resulted in the increased formation of what we refer to as 'nest-like' structures led us to link the increased levels of p40 with the membranous structures within the mammalian cells. This then leads us to ask the question- is p40 associating with detergent resistant membranes and stabilizing these membranes? Alternatively, does the presence of p40 within the membranes make the membrane detergent resistant? Acknowledging that the detergent resistance is a characteristic and not in itself a defining feature, it does then highlight that the role of p40 in the replication complex may be the formation of a stable structure within which PrV replication can occur. Our hypotheses are informed by the current understanding of FHV and the mechanisms FHV employs for replication.

In our study, we use the term detergent resistant membranes to refer to structures that form within the mammalian cells when they are treated with 0.1 % Triton X-100 at 37 °C for 15 minutes. Triton X-100 was used at 37 °C by Short *et al*. (2016) when they isolated FHV-infected mitochondria from *Drosophila* cells to develop a cell-free replication system. Short *et al*. (2016) demonstrated that when 1 % NP-40 or Triton X-100 was added to viral replication complexes, the complexes were disrupted as evidenced by a reduction in viral replication and an increase in nuclease-dependent degradation of the dsRNA replication intermediate. For FHV, the membranous structures provided an enclosed environment for positive-sense RNA virus replication and protection from cellular antiviral defence mechanisms. It is tempting to speculate about the structure of the PrV replication complex on the mitochondrial membrane. Our attempts to visualise the structure have not been successful however we hypothesize that PrV forms a funnel shaped invagination in the mitochondrial membrane, similar to the structure formed by FHV, and that this invaginated structure provides the protected environment for

viral replication. Based on our observations, it is tempting to speculate that p40 is important in the formation of the invaginated structure in the membrane.

We were most excited to visualize PrV-specific proteins in the nucleolus of the nuclei of PrV-infected HeLa and MCF-7 cells, when these cells were treated with 0.1 % Triton X-100. This is the first observation of PrV proteins within the nucleolus of the mammalian cells and it could have important ramifications when considering PrV replication in mammalian cells. The nucleolus is a highly dynamic structure within the nucleus and is involved in ribosome subunit biogenesis, the mediation of cell stress responses and the regulation of cell growth (Hiscox, 2007). There have been several studies that have reported virus-induced modifications of nucleolar structure and composition; these changes result in the interference with fundamental global cellular processes including cell cycle regulation, apoptosis and intracellular trafficking (Salvetti and Greco, 2014). Most DNA viruses and retroviruses replicate in the nucleus of the host cell and so the viral proteins are expected to target the nuclear and nucleolar structures (Hiscox, 2007). In contrast, most RNA viruses replicate in the cytoplasm of the infected host cell however the gene products of some viruses have been found to localise to the nucleolus (Salvetti and Greco, 2014). Examples include the viral capsid or nucleoproteins of coronaviruses (Hiscox *et al*., 2001), arteriviruses (Rowland *et al*., 1999), alphaviruses (Michel *et al*., 1990) and flaviviruses (Balinsky *et al*., 2013; Wang *et al*., 2002). The non-structural proteins of Dengue virus (NS5) and HCV (NS5B) have also been shown to accumulate in the nucleolus of virus-infected cells (Fraser *et al*., 2016; Hirano *et al*., 2003). Dengue virus NS5 accumulates in the nucleolus in response to a change in the extracellular pH (Fraser *et al*., 2016) whereas HCV NS5B binds to nucleolin, a nucleolar protein, which results in its relocation to the cytoplasm; this process results in enhanced viral replication (Hirano *et al*., 2003). The hijacking of the nucleolus has been employed by plant viruses, including the umbraviruses (Kim *et al*., 2004; Taliansky and Robinson, 2003) and maize fine streak virus (Tsai *et al*., 2005). Viral proteins that are trafficked to the nucleolus are thought to employ a form of molecular mimicry, where the viral protein contains motifs that resemble nucleolar localisation signals (Hiscox, 2007). There have been no reports of a conserved nucleolar trafficking signal in viral proteins (Hiscox, 2007; Fraser *et al*., 2016). It is however possible that viral proteins may be trafficked to the nucleolus when they associate with other nucleolar proteins (Hiscox, 2007). The detection of dsRNA and p40 in the nucleolus of the PrV-infected mammalian cells may indicate how PrV establishes a persistent infection in mammalian cells.

Chapter 6: Final discussion and conclusions

Emerging viruses

There are several factors which influence the emergence of new diseases and the adaptation of already known viruses. Viral diversity is far higher than is currently known. This is supported by the report of Shi *et al.* (2016). It is therefore naïve to believe that the next viral pandemic will be caused by a known virus. The increased contact between humans in previously isolated populations increases the risk of disease transmission and the emergence of new viral epidemics (Carrasco-Hernandez *et al*., 2017). The influence of climate change has been linked to the emergence of disease and this concept has been interrogated and discussed in The Stockholm Paradigm (Brooks *et al*., 2019; Cable, 2020). When considering the emergence of a new disease, the pathogen needs to adapt to the host and the ecological niche, through factors such as gene mutation, genetic recombination and reassortment (Alcaïs *et al.,* 2009; Mouchet and Carnevale, 1997; Nii-Trebi, 2017). Environmental factors such as deforestation, expansion and modernization of agricultural practices, and natural disasters have led to changes in the microbial niches and have fuelled adaptation (Neiderud, 2015; Nii-Trebi, 2017; Tong *et al*., 2015). Sociodemographic factors such as the ever-increasing population density, increased global travel, the killing and eating of wild animals for their meat as well as conflict and social instability in war-stricken areas have increased the prevalence of viral outbreaks (Nii-Trebi, 2017). In conflict areas such as Syria, which had been polio-free for over 10 years, there has been the re-emergence of poliovirus; this has been attributed to a reduction in the rate of vaccination (Akil and Ahmad, 2016). In 2012, vaccination rates fell from 91 % to an estimated 68 % (Akil and Ahmad, 2016). Ebola virus outbreaks have been linked to the contact and consumption of Ebola virus-infected animals such as chimpanzees and duikers (Leroy *et al*., 2004; Cantlay *et al*., 2017).

Bats are considered a particularly important natural reservoir. This is because they are natural reservoirs for a variety of viruses and often, they do not show signs of infection, but they can transmit these viruses to susceptible human hosts (Calisher *et al*., 2006; Leech and Baker, 2017). Deadly zoonotic viruses such as paramyxoviruses, lyssaviruses, coronaviruses and filoviruses have been isolated from bats (Calisher *et al*., 2006; Leech and Baker, 2017). In 2012, Ge *et al*. performed metagenomic sequencing of bat guano and reported that bats contained a large number of novel viruses, dominated by densoviruses, dicistroviruses, coronaviruses, parvoviruses and tobamoviruses. We must mention here, Kemenesi *et al*. (2016) isolated the PrV genome from bat guano in Hungary. Advances in detection and control strategies will always be a step behind, and the emergence of each new disease brings a unique set of challenges (Firth, 2014; Nii-Trebi, 2017).

Antiviral therapies

Over the past 50 years, only 90 antiviral drugs have been approved for the treatment of 9 human diseases, namely: HIV, HCV, HBV, human Influenza, RSV, HCMV, HSV, VZV and HPV (De Clercq and Li, 2016). The positive-sense RNA viruses that are known to establish persistent infections in humans and the antiviral drugs available to treat these viruses are summarised in Table 6.1. There are 10 positive-sense RNA viruses that are known to establish persistent infections in humans. Of these 10, HCV is the only virus that has antiviral drugs approved for treatment. Japanese encephalitis virus, poliovirus and rubella virus (with measles and mumps - MMR vaccine) are the only persistent, positive-sense RNA viruses that have approved vaccines available. This highlights the desperate need for the generation of antiviral drugs and vaccines.

Table 6.1: Positive-sense RNA viruses that establish persistent infections in humans and the antiviral drugs available (Table adapted from Randall and Griffin (2017) and De Clercq and Li (2016)).

Virus	Antiviral drugs available
Noroviruses (Norwalk virus)	No
Hepatitis C virus	Yes, 20 *
Zika virus	No
West Nile virus	No
Japanese encephalitis virus	No (vaccine available)
Poliovirus	No (vaccine available)
Coxsackie virus	No
Rhinovirus	No
Rubella virus	No (vaccine available)
Chikungunya virus	No

* Telaprevir, Boceprevir, Simeprevir, Asunaprevir, Vaniprevir + ribavirin + PegIFNα-2b, Paritaprevir, Grazoprevir, Sofosbuvir + ribavirin, Sofosbuvir + ribavirin + PegIFNα, Daclatasvir + asunaprivir, Ledipasvir + sofosbuvir, Sofosbuvir + simeprevir, Ombitasvir + dasabuvir + paritaprevir + ritonavir, Ombitasvir + paritaprevir + ritonavir, Daclatasvir + sofosbuvir, Elbasvir + grazoprevir, ribavirin, Peglated interfon alpha 2b, Interferon alfacon 1, Peglated interfon alpha 2b + ribavirin, Peglated interfon alpha 2a.

Mutation rates in RNA viruses

Due to the population size, the short generation times as well as the error-prone RdRp, RNA viruses have a high mutation rate; which translates to the continuous production of viral variants (Cook *et al*., 2013; Kane and Golovkina, 2010; Randall and Griffin, 2017; Shi *et al*.,

2016). The viral variants result in the generation of a quasi-species which adds to the challenge of developing antiviral drugs and the treatment of RNA viruses (Lauring and Andino, 2010; Sklan *et al*., 2009). The major form of treatment for these viral infections is the prescription of anti-inflammatories such as corticosteroids which treats only the symptoms, not the disease (Randall and Griffin, 2017). Protection from identification by the host antiviral response would also suggest that the viruses are also protected from antiviral drugs that target viral replication. Antiviral treatments would therefore need to target aspects of the viral lifecycle that are not protected.

Most of our knowledge about viral evolution is based on the study of cultured viruses (Ge *et al*., 2012; Shi *et al*., 2016). It is important to note that many viruses are unable to replicate in cell culture. Challenges include a lack of susceptible host cell lines, low virus titres and the toxicity of environmental samples (Ge *et al*., 2012). In cultured viruses, the variability between strains of the same virus can cause difficulties when studying virus replication biology and pathology. RNA viruses have a high mutation rate, even within a species, viral sequence can show between 10 and 30 % sequence variability at a nucleotide level (Firth, 2014). For example, Zika virus, a positive-sense, RNA virus has evolved dramatically from the 2007 outbreak on the Yap Island, Microindonesia when compared to the more recent outbreak in the Americas in 2015. Zika virus was considered an inconsequential flavivirus that caused mild symptoms. This virus has evolved such that the 2015 virus causes a more severe disease and results in severe congenital defects, primarily microcephaly in children (Liu *et al*., 2019; Shan *et al*., 2016; Wang *et al*., 2017). The escalation in the severity of the disease highlights the increased need for systems and tools to study and understand viruses.

Host switching and cross-species/Kingdom transfer

Most viruses are thought to have co-evolved with their host species as the success of viral replication is complex and requires many interactions with the host (Bandin and Dopazo, 2011). The co-evolution between virus and host leads to species specificity and in turn makes interspecies transfer difficult (Bandin and Dopazo, 2011). Natural host switches are believed to be rare; however, there are several families of viruses that have been reported to have a broader host range, for example *Orthomyxoviridae*, *Flaviviridae, Bunyaviridae* and *Togaviridae* (Jiwaji *et al*., 2019; Figure 6.1).

The ability of a virus to move from one host into a novel species is described as host shifting and is often associated with changes in the viral genome sequence (Jiwaji *et al*., 2019; Longdon *et al*., 2014). These alterations often come at a cost to the viral fitness; it is remarkable that PrV purified from persistently infected insect cells can infect and replicate in

plants and then go on to establish a productive infection in mammalian cell cultures (Jiwaji *et al*., 2019). The ability to infect hosts belonging to different Kingdoms would imply that there are similar receptors for binding and entry between insect and mammalian cells. In addition, there needs to be a comparable site of replication in insect, mammalian and plant cells. Jiwaji *et al.* (2019) published a very elegant summary of the virus families known to infect plants, vertebrates and invertebrates. In this figure (Figure 6.1), it shows that some viruses are capable of infecting both plants and invertebrates or invertebrates and vertebrates but only viruses belonging to the *Carmotetraviridae* and *Nodaviridae* families have been shown to infect plant, invertebrate and vertebrate. In this document, we report that PrV replicates in association with the mitochondria in HeLa cells. FHV also establishes replication complexes in association with the mitochondria and is the only other known virus to replicate in insect, plant and mammalian cell line (Miller *et al*., 2001). We propose that it is the fact that there are comparable sites of replication that permit both FHV and PrV to demonstrate their broad host range.

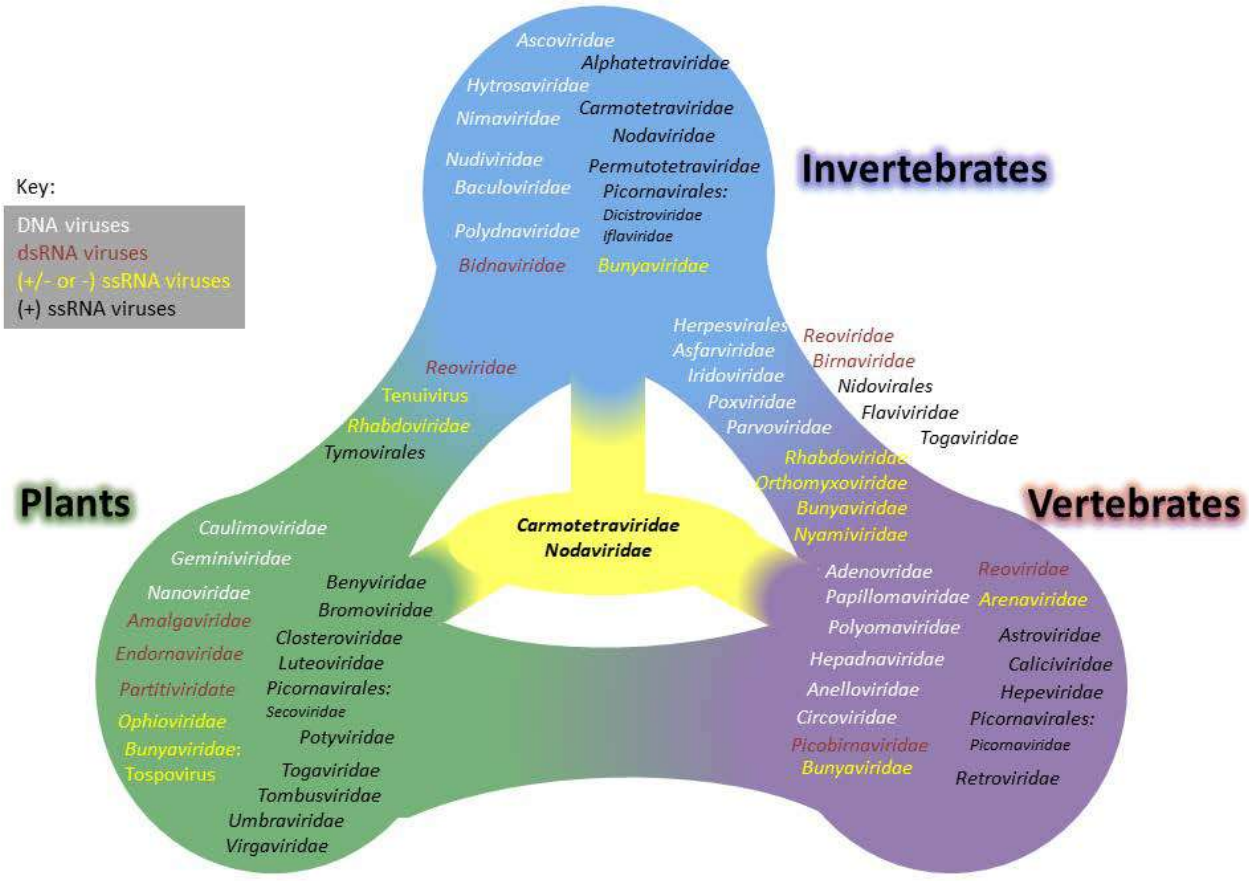


Figure 6.1: A summary of the known virus families and the associated hosts. The virus families that infect and replicate in plants, vertebrates and invertebrates are shown in green, purple

and blue, respectively. The virus families that have been reported to infect more than one Kingdom are included in the joining branch. The virus families known to replicate in all three Kingdoms have been included in the middle, yellow group. This figure has been adapted from Jiwaji *et al*. (2019).

The location of viral replication

Most positive-sense RNA viruses replicate in association with membranes; and viral infection results in extensive reorganisation of intracellular membranes to facilitate viral replication and to avoid detection by the host antiviral response (Martinez-Turino and Hernandez, 2012). Viruses such as Dengue virus and HCV replicate in association with modified ER membranes (Gosert *et al*., 2003; Heaton and Randall, 2011). Members of the *Nodaviridae* and *Tombusviridae* have been reported to replicate in association with mitochondrial membranes, causing ultra-structural changes to create replication factories (Gómez-Aix *et al*., 2015; Martínez-Turiño and Hernández, 2012; Miller *et al*., 2001). We hypothesize that PrV, like FHV, forms invaginations in the mitochondrial membrane to generate viral replication factories that are protected from detection. This hypothesis is supported by previous research performed by Jarvie (2017). Jarvie (2017) adapted DRACO, a **D**ouble-stranded **R**NA **A**ctivated **C**aspase **O**ligomerizer reported by Rider *et al*. (2011) to target PrV. DRACO combined a dsRNA detection domain with an apoptosis-inducing domain. When dsRNA was detected in a cell, DRACO would selectively induce apoptosis without causing an inflammatory response (Guo *et al*., 2015; Rider *et al*., 2011). DRACO had been reported to be effective against a wide variety of RNA viruses (Guo *et al*., 2015; Rider *et al*., 2011) but DRACO was not effective against PrV. Treatment of a PrV-infected HeLa cell did not cause apoptosis and our analyses indicated that DRACO was able to detect dsRNA, but not able to detect replicating PrV dsRNA in the cell. We hypothesized that this was due to the sequestering of dsRNA within replication structures thus preventing their detection by DRACO (Jarvie, 2017, Unpublished).

PrV as a model system

The ability of PrV to cross Kingdom boundaries (Jiwaji *et al*., 2019) and establish a persistent infection in cell culture (Jiwaji *et al*., 2016; Pringle *et al*., 2003) provided the unique opportunity to use PrV to develop a biological system to study viral replication biology, viral persistence and the mechanisms involved in cross Kingdom and species transfer. Preliminary studies have used PrV-infected mammalian cells as a screening platform for antiviral drug activity (Duba, 2018, Unpublished). PrV-infected HeLa cells were treated with Favipiravir (an anti-Influenza drug) and Ribavirin (an anti-HCV drug); both drugs inhibited PrV replication (Duba, 2018, Unpublished). This highlights the value of PrV as a model system and it is for this reason that it was selected in this study. It must also be highlighted that the data generated in this study

provides preliminary insights into PrV specifically as well as the potential for PrV-based applications.

Our studies with PrV have resulted in a deeper understanding of the challenges that face molecular virologists interested in the study of virus biology. In this final chapter, we would like to bring together aspects raised through the study for deeper discussion.

The bioinformatics analysis of viral genomes

There are few bioinformatic tools that are available to study viral genomes. We used two online ORF prediction sites; GeneMarkS and SoftBerry to analyse the PrV genome. In addition to the identification of putative proteins predicted and reported by Walter *et al.* (2008), the analyses highlighted the presence of ORFs on the negative-sense strand of the PrV genome (Figure 3.3); Gene 1 (8.4 kDa) was identified by GeneMarkS and Gene E (20.3 kDa) by SoftBerry, and the two ORFs were in different locations. It is possible that these are inaccurate predictions and we have no biological data to indicate that proteins of 8.4 or 20.3 kDa are being produced. In fact, until recently, there have been no reports of coding ORFs on the negative-sense strand of positive-sense single-stranded RNA viruses (Dinan *et al*., 2019).

In 2013, Cook *et al*. identified two mosquito-associated narna-like viruses after deep sequencing mosquitoes and chironomids collected from natural environments. Analysis of these genomes showed that they were single stranded, with a single ORF that spanned most of the length of the genome; based on homology, this ORF was inferred to be the RNA dependent RNA polymerase (Cook *et al*., 2013; Dinan *et al*., 2019). Further analysis by Dinan *et al*. (2019) indicted that the narna-like virus also contained a reverse frame ORF, and this was thought to represent a genuine protein coding sequence.

Narnaviridae are a family of viruses with positive-sense single-stranded RNA genomes which are non-encapsidated. These viruses were originally described as infecting fungi, but transcriptomic analyses have identified related viruses in a diverse range of organisms (Dinan *et al*., 2019). There are two genera, *Mitovirus* and *Narnavirus*. Mitoviruses replicate in association with the mitochondria whereas Narnaviruses replicate exclusively in the cytosol with no reported organelle-association (Dinan *et al*., 2019; Hillman and Cai, 2013). Most positive-sense RNA viruses form replication factories in association with organelle membranes, resulting in extensive re-organisation of the membranes to protect the dsRNA replication intermediate from detection by the host antiviral response (Dinan *et al*., 2019; Martinez-Turino and Hernandez, 2012). The formation of these protective structures may also function as a mechanism to separate viral translation from viral replication; the negative strand

is kept within the structure while the positive strand is exported and translated by the ribosomes in the cytosol (Dinan *et al*., 2019). It is possible that PrV may also employ this division of tasks; replication occurs in association with the mitochondria however the mRNA is translated in the cytoplasm of the mammalian cells (Figure 3.8 and 3.9). This scenario makes the presence of functional ORFs on the negative strand unlikely, but not impossible. Narnaviruses, in contrast to most positive-sense RNA viruses, do not associate with cellular membranes when replicating; this factor may have contributed to the evolution of negative strand translation (Dinan *et al*., 2019; Fujimura *et al*., 2005). To determine whether the ORFs on the negative strand are produced, we need to perform deep protein mass spectrometry and identify the presence of putative proteins of interest.

"Hidden" ORFs are very short coding regions found on the genomes of positive-sense RNA viruses; these ORFs overlap previously identified ORFs and they are frequently expressed via non-canonical translation mechanisms (Dinan *et al*., 2019; Firth, 2014). Hidden genes have been identified in many viruses, for example Turnip yellows virus (family *Luteoviridae*, genus *Polerovirus*; Smirnova *et al*., 2015); encephalomyocarditis virus (family *Picornaviridae,* genus *Cardiovirus;* Napthine *et al*., 2017); some *Enteroviruses* (family *Picornaviridae;* Lulla *et al*., 2019). Firth (2014) developed an algorithm (synplot2) to use multiple genomesof viral protein coding sequences to search for synonymous sites and identify "hidden" ORFs. It would be interesting to perform this analysis on PrV; it may yield some information about the unexpected proteins predicted by GeneMarkS and SoftBerry (Figure 3.3) and the unexpected bands detected in the western blot analyses (Figure 3.5 and 3.6).

The study of virus replication in vivo

In 2012, Roux *et al*. described BioID, an *in vivo* approach to proximity labelling of proteins using biotin and a promiscuous biotin ligase enzyme. This system has been used to successfully study viruses including HIV-1 (Le Sage *et al*., 2015; Ritchie *et al*., 2015) and herpesviruses (Lajko *et al*., 2015). The system was effective when applied to insoluble and membrane associated proteins, and we adapted BioID to study PrV proteins. The PrV accessory protein p40, was the most comprehensively studied PrV protein and was selected as the focus of this study. The intention of this study was to identify the proteins that interact with the PrV replication proteins, and so elucidate the mechanisms of PrV replication in mammalian cells. The BirA-p40 system was used *in vivo* both in transient and in stable protein expression situations. In addition, the system was applied in an *in vitro* context as well. One of the critical challenges in this project was the discrepancy between the detection of BirA-p40 during western blot analyses (Figure 4.5) compared to during confocal microscopy (Figure 4.6). When cells are prepared for confocal microscopy, the cells are fixed in the native

conformation and the addition of the permeabilization buffer permits the entry of antibodies, which can then bind to the protein of interest. During western analyses, the addition of SDS resulted in the irreversible denaturation of the protein which altered protein conformation and epitope availability. These results highlight the importance of antibody validation in a context and application specific manner (Älgenäs *et al*., 2014; Uhlen *et al*., 2016). The variation in detection of the same protein using different techniques is concerning and posed a great challenge when studying PrV. The challenges with protein detection using western blot analyses resulted in the inability to perform the intended experiments and analysis *in vivo*.

The BioID system was not as effective for the study of PrV as desired. The accessory protein p40 associated strongly with detergent resistant membranes and this is thought to have prevented the promiscuous biotinylation of proximal proteins. The over-expression of BirA-p40 in HeLa cells that were treated with detergent and prepared for TEM (Figure 5.8) showed that p40 was localised to 'nest-like' structures within the cytosol and we hypothesize that the association of p40 with these membranes results in the membranes being more resistant to the detergent treatment.

The BioID system could be useful for *in vivo* studies if we could develop biotin molecules that are fluorescently labelled at the correct site. At present, all fluorescently tagged biotin molecules studied would lose the fluorescent moiety when the biotin was activated by BirA. Once available, the fluorescently-tagged biotin would permit the visualisation of the protein-protein interactions using confocal microscopy and live cell imaging. It would also be interesting to label viral RNA using the primary amines, which would permit tracking the PrV genome from the replication factory and into the viral particle.

The development of super-resolution fluorescence microscopy, such as single-molecule localisation microscopy (SMLM), total internal reflection fluorescence microscopy (TIRF-M) and spinning disk confocal microscopy (SDCM), provide new opportunities to visualise viral replication *in vivo.* These super-resolution techniques, referred to as nanoscopy as they can achieve resolution down to 10 nm, allows for the visualisation of subviral structures as well as host-virus interactions through live cell imaging (Hanne *et al*., 2016; Sakin *et al*., 2016). These techniques are currently limited when attempting to visualise live interaction events *in vivo;* the fluorophore needs to enter the cell and bind to the target protein and the fluorophores needs to be photostable. Many of the fluorophores that are currently available are not able to enter the cell (Sakin *et al*., 2016). One alternative is to use autofluorescent protein tags (such as GFP) however fusion proteins brings their own challenges when studying protein interactions (Sakin *et al*., 2016). We performed live cell imaging on PrV-infected HeLa cells

that were treated with Triton X-100 and stained with MitoTracker. The cells were visualised over a 15 minute period (Appendix A4, Figure A4). As time progressed, the stained mitochondria were observed to move towards the nuclear area and to form what appeared to be the 'nest-like' structures observed using TEM and confocal microscopy in Chapter 5. Unfortunately, we were unable to label the PrV proteins with PrV-specific antibodies in live cells therefore we were unable to co-localise the nest-like structures with PrV replication-associated proteins *in vivo*.

The detection of virus replication in mammalian cells

In this document, we reported the development of a technique that enhances the detection of viruses. The challenges of detection include a lack of technical skills required to perform screening assays and the high cost of samples preparation. The problem is serious in areas that have high prevalence of viral infection but low resource availability (El Ekiaby *et al.,* 2010). There have recently been technological developments which enable the amplification and detection of multiple pathogens within a single sample (Ali *et al*., 2017). In 2017, Ali *et al*. used multiplex PCR for the simultaneous detection of multiple viruses in clinical samples in blood screening facilities. The authors combined nucleic acids, purification by magnetic separation, with chemiluminescent detection to detect multiple viruses including HIV-1, HBV and HCV in a sample after PCR amplification. The process could be automated and used in high throughput applications (Ali *et al*., 2017), however it does not permit the study of viral replication *in vivo.*

In this study, we presented data that showed that when PrV-infected mammalian cells, which appeared to have low levels of replicating virus, were treated with non-ionic, linear detergents (namely, Triton X-100 and NP-40) there was dramatic increase in the detection of both dsRNA and p40. We showed that in the case of PrV, the use of detergents enhanced the detection of viruses that were replicating persistently albeit at low levels within the cells. We propose that the treatment of virus-infected mammalian cells with Triton X-100 may represent a cost-effective tool to detection viral infection, provided there is available access to a confocal microscope. This is particularly important because detection is the first and critical step when studying virus replication *in vivo.* Once detected, the identification and study of viral replication permit the development of antiviral drugs and assays to screen for antiviral drugs.

Anticipating the depth of viral diversity

The development of high throughput sequencing technologies and metagenomic tools has opened new avenues for the identification and study of viral diversity (Ge *et al*., 2012; Hall *et al*., 2014; Shi *et al*., 2016). Previously, the study of viral sequences was accomplished through

cloning and Sanger-based sequencing of randomly amplified nucleic acids (Breitbart *et al*., 2003; Hall *et al*., 2014). The advancements in sequencing technologies enables the generation of millions of sequences reads without any prior knowledge of the sample (Hall *et al*., 2014). In 2016, Shi *et al*. published a study that provided a new perspective on viral diversity. The authors performed a large-scale metatranscriptomic survey of a diverse range of invertebrate taxa (over 220 invertebrate species) and reported the discovery of 1 445 RNA viruses. This approach is relatively unbiased as there was no enrichment of viral particles through filtration, centrifugation or nuclease treatment (Shi *et al*., 2016). The identification of these new viruses filled gaps within the RNA virus phylogenetic trees and revealed interesting evolutionary histories that are characterised by co-divergence and host-switching. The study by Shi *et al.* (2016) highlights the current lack of knowledge and understanding of the virome. With increasing studies, there have been reports of viruses that share similarity to PrV, for example Reuter *et al*. (2019). These studies indicate that the depth of viral diversity will identify new viruses that share similar characteristics to PrV and provide new opportunities for understanding viral diversity and evolution.

Concluding remarks

In this study, we optimistically intended to unravel PrV replication and report where PrV was replicating and what cellular proteins the PrV replication-associated proteins were interacting with. Our study highlighted our lack of understanding about viruses in general, and PrV specifically. For a virus with three known proteins, and some putative proteins, the replication biology is very complex.

It is interesting to note the differences in the site of virus replication between the insect cells (Short *et al*., 2013) and the mammalian cells. In the insect cells, PrV established replication complexes at the secretory vesicles and Golgi apparatus (Short *et al*., 2013) whereas here, we report that in HeLa cells PrV replicates at the mitochondria. It is also interesting to note that VCAP was readily detected in insect cells (Jiwaji *et al*., 2019; Walter *et al*., 2010) compared to this study where VCAP was only detected in mammalian cells undergoing apoptosis. It would be very interesting to sequence the genomes of PrV particles from insect, mammalian and plant cells to evaluate whether there have been significant changes in the PrV genome sequence, and if so, where these changes have been.

It is tempting to facetiously name this report “Several ways of how not to study Providence virus replication in mammalian cells”. Despite the challenges faced in this study, the findings in this report have advanced our understanding of PrV replication biology *in vivo* and have

highlighted avenues for further research. For this reason, we have entitled this work “Unravelling the replication biology of Providence virusin a cell culture-based model system”.

To end, the virus that has been the focus of this study was named for the town where the laboratory that developed the MG8 cell line was located (Pringle *et al.*, 2003). The name of the virus invites a wry smile because Providence as a noun has a second, and deeper, meaning. The Cambridge Dictionary (2019) defines the word as “an influence that is not human in origin and is thought to control people’s lives”.

Appendices

Appendix A1

	Analytical Report
Report to	**Rachel Jarvie** **Department of Biochemistry and Microbiology** **Rhodes University**
Document Number	**RJ_2018-09**
Quote number **Version**	**1.0**
Status	**Final**
Date	**3 October 2018**

Samples Submitted

Ten (11) gel slices were submitted labelled as indicated in Table 1.

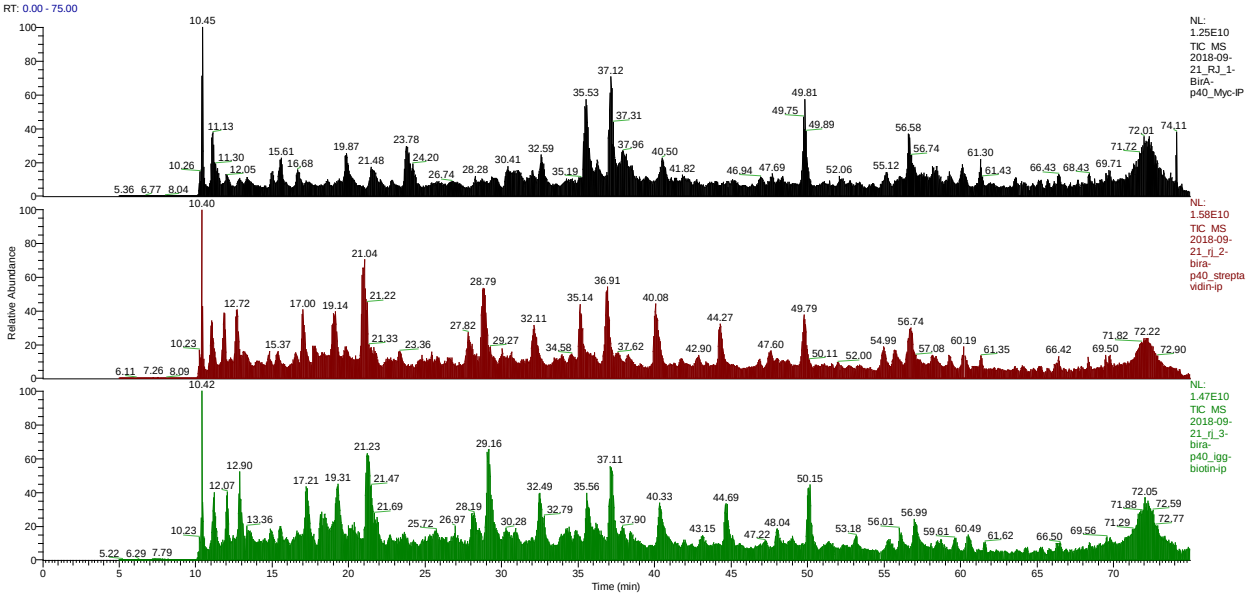
Table 1: Samples submitted

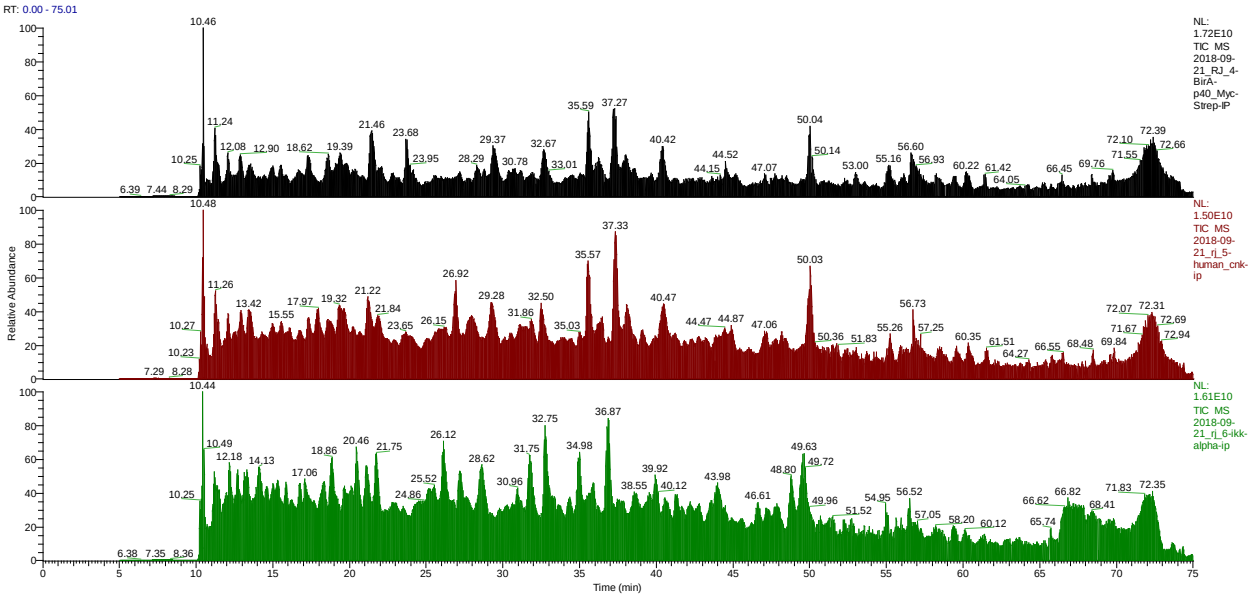
Sample number	Sample ID
1	BirA-p40 (Myc IP)
2	BirA-p40 (Streptavidin IP)

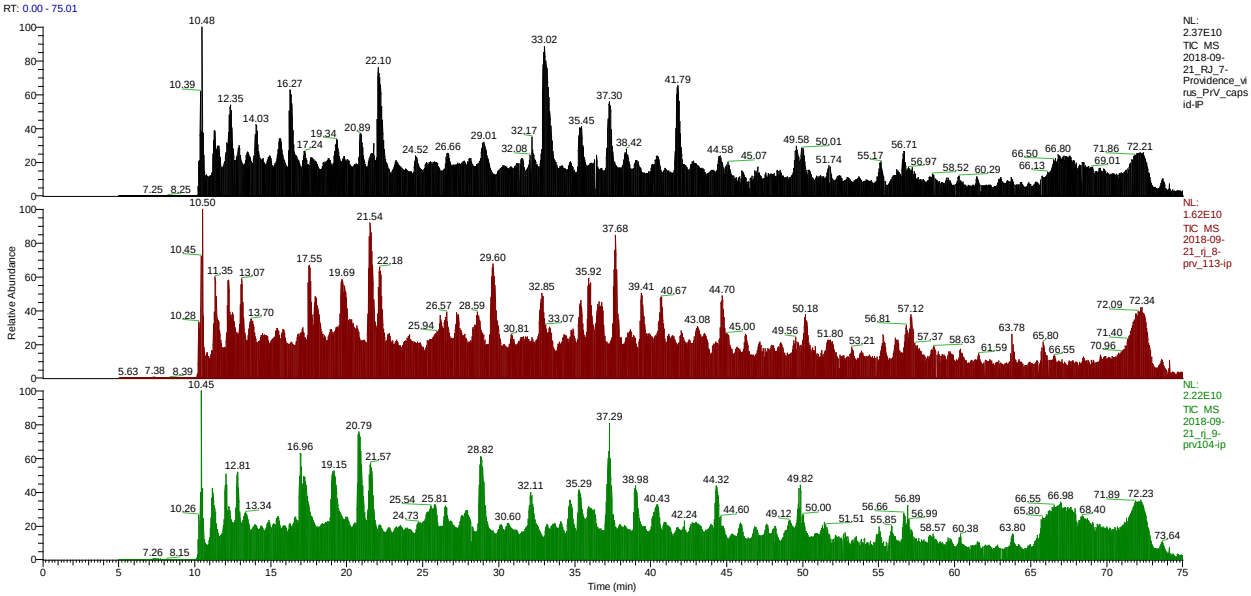
3	BirA-p40 (IgG-Biotin IP)
4	BirA-p40 (Myc and Strep IP)
5	Human CNK IP
6	IKKalpha IP
7	Providence virus (PrV) capsid IP
8	PrV p113 IP
9	PrV p104 IP
10	PrV p40 IP
11	PrV p17 IP

Analysis Details

The samples were digested with trypsin and the total ion chromatograms (TIC's) indicated successful (Figure 1).




RT: 0.00 - 75.01
NL: 1.72E10 TIC MS 2018-09-21_RJ_4-BirA-p40_Myc-Strep-IP
NL: 1.50E10 TIC MS 2018-09-21_rj_5-human_cnk-ip
NL: 1.61E10 TIC MS 2018-09-21_rj_6-ikk-alpha-ip
Relative Abundance
Time (min)


RT: 0.00 - 75.01
NL: 2.37E10 TIC MS 2018-09-21_RJ_7-Providence_virus_PrV_capsid-IP
NL: 1.62E10 TIC MS 2018-09-21_rj_8-prv_113-ip
NL: 2.22E10 TIC MS 2018-09-21_rj_9-prv104-ip
Relative Abundance
Time (min)

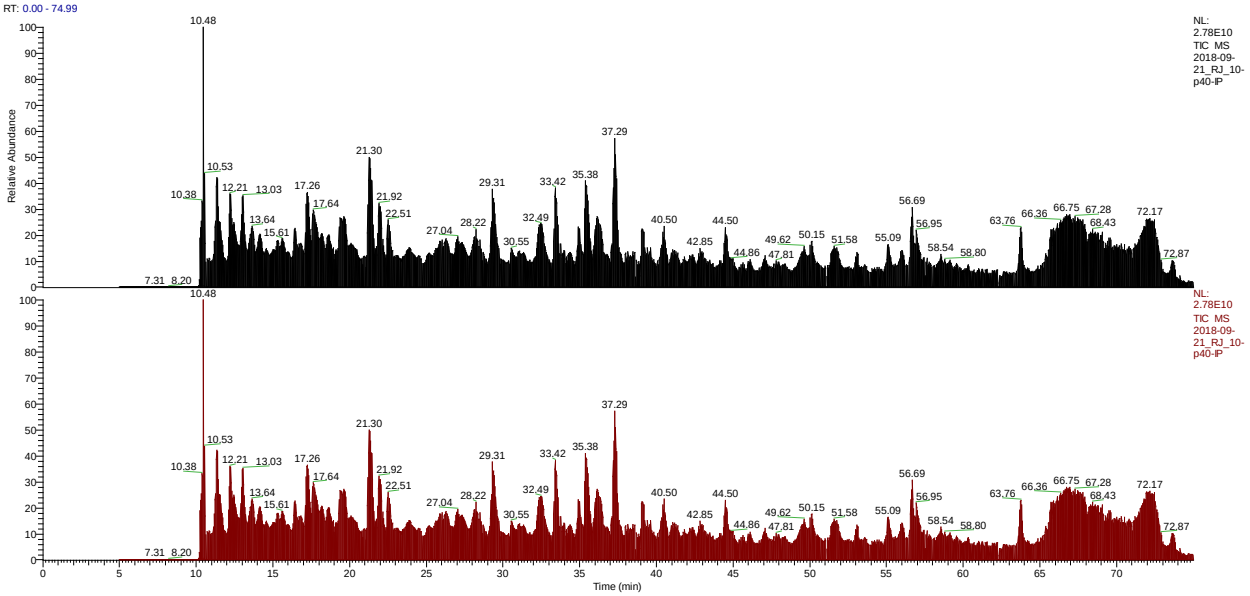

RT: 0.00 - 74.99
NL: 2.78E10 TIC MS 2018-09-21_RJ_10-p40-IP
NL: 2.78E10 TIC MS 2018-09-21_RJ_10-p40-IP
Relative Abundance
Time (min)

Figure 1: Total ion chromatograms of the samples submitted.

Data base interrogation yielded multiple proteins in all the samples. A total 1555 proteins with 2 peptides or more and a protein FDR of 1% was detected across all the samples. Table 2 provides a breakdown of the protein ID distribution. The identity of the proteins are contained in the files 2018-09-21_(Sample ID).sf3 and 2018-09-21_RJ_Combined as well as the corresponding Excel files.

Table 2: Number of proteins identified per sample.

Sample number	Sample ID	Number of protein identified
1	BirA-p40 (Myc IP)	890
2	BirA-p40 (Streptavidin IP)	1016
3	BirA-p40 (IgG-Biotin IP)	1040
4	BirA-p40 (Myc and Strep IP)	1177
5	Human CNK IP	1410
6	IKKalpha IP	1420
7	Providence virus (PrV) capsid IP	1230
8	PrV p113 IP	1307
9	PrV p104 IP	1349
10	PrV p40 IP	1423
11	PrV p17 IP	1365

Experimental Procedures

In-gel Digest

All reagents are analytical grade or equivalent. Gel slices supplied were destained in an Eppendorf 1.5 mL tube with 200 mM NH_4HCO_3:Acetonitrile 50:50 (Sigma) until clear. Samples were dehydrated and desiccated before reduction with 2 mM triscarboxyethyl phosphine (TCEP; Fluka) in 25 mM NH_4HCO_3 for 15 minutes at room temperature with agitation. Excess TCEP were removed and the gel pieces again dehydrated. Cystein residues were thiomethylated with 20 mM *S*-Methyl methanethiosulfonate (Sigma) in 25 mM NH_4HCO_3 for 30 minutes at room temperature. After thiomethylation the gel pieces were dehydrated and washed with 25 mM NH_4HCO_3 followed by another dehydration step. Proteins were digested by rehydrating the gel pieces in trypsin (Pierce) solution (20ng/□□L) and incubating at 37 °C overnight. Peptides were extracted from the gel pieces once with 50µL water and once with 50% acetonitrile. The samples were dried down and resuspended in 30□L 2% acetonitrile:water; 0.1% FA

Residual digest reagents were removed using an in-house manufactured C_{18} stage tip (Empore Octadecyl C_{18} extraction discs; Supelco). The samples were loaded onto the stage tip after activating the C_{18} membrane with 30□□L methanol (Sigma) and equilibration with 30□L 2% acetonitrile:water; 0.05% TFA. The bound sample was washed with 30□L 2% acetonitrile:water; 0.1% TFA before elution with 30 □L 50% acetonitrile:water 0.05% TFA. The eluate was eavaporated to dryness. The dried peptides were dissolved in 2% acetonitrile:water; 0.1% FA for LC-MS analysis.

Liquid chromatography

Dionex nano-RSLC

Liquid chromatography was performed on a Thermo Scientific Ultimate 3000 RSLC equipped with a 5mm x 300□m C_{18} trap column (Thermo Scientific) and a Pepmap 25cmx75□m 2□m particle size C_{18} column (Thermo Scientific) analytical column. The solvent system employed was loading: 2% acetonitrile:water; 0.1% FA; Solvent A: 2% acetonitrile:water; 0.1% FA and Solvent B: 100% acetonitrile:water. The samples were loaded onto the trap column using loading solvent at a flow rate of 15□□L/min from a temperature controlled autosampler set at 7C. Loading was performed for 5 min before the sample was eluted onto the analytical column. Flow rate was set to 350nL/minute and the gradient generated as follows: 5.0% -30%B over 60 min using Chromeleon non-linear gradient 6; 30% -50% B from 60 -70. Chromatography was performed at 40°C and the outflow delivered to the mass spectrometer through a stainless steel nano-bore emitter.

Mass spectrometry

Mass spectrometry was performed using a Thermo Scientific Fusion mass spectrometer equipped with a Nanospray Flex ionization source. The sample was introduced through a stainless steel emitter. Data was collected in positive mode with spray voltage set to 1.8kV and ion transfer capillary set to 280°C. Spectra were internally calibrated using polysiloxane ions at m/z = 445.12003 and 371.10024. MS1 scans were performed using the orbitrap detector set at 120 000 resolution over the scan range 350-1650 with AGC target at 3 E5 and maximum injection time of 40ms. Data was acquired in profile mode.

MS2 acquisitions were performed using monoisotopic precursor selection for ion with charges +2-+7 with error tolerance set to +/- 10ppm. Precursor ions were excluded from fragmentation once for a period of 60s. Precursor ions were selected for fragmentation in HCD mode using the quadrupole mass analyser with HCD energy set to 32.5%. Fragment ions were detected in the orbitrap mass analyzer set to 30 000 resolution. The AGC target was set to 5E4 and the maximum injection time to 80ms. The data was acquired in centroid mode.

Data Analysis

The raw files generated by the mass spectrometer were imported into Proteome Discoverer v1.4 (Thermo Scientific) and processed using both Sequest and Amanda algorithm. Database interrogation was performed against a concatenated database created using the Uniprot *Homo sapiens* reviewed database, Uniprot viral database and the cRAP contaminant database. Semi-tryptic cleavage with 2 missed cleavages was allowed for. Precursor mass tolerance was set to 10ppm and fragment mass tolerance set to 0.05 Da. Demamidation (NQ), oxidation (M) and acetylation of protein N-terminal was allowed as dynamic modifications and thiomethyl of C as static modification. Peptide validation was performed using the Target-Decoy PSM validator node. The results files were imported into Scaffold 1.4.4 and identified peptides validated the Peptide and Protein Prophet algorithms included in Scaffold.

Deviations

No deviations were recorded

Enquiries

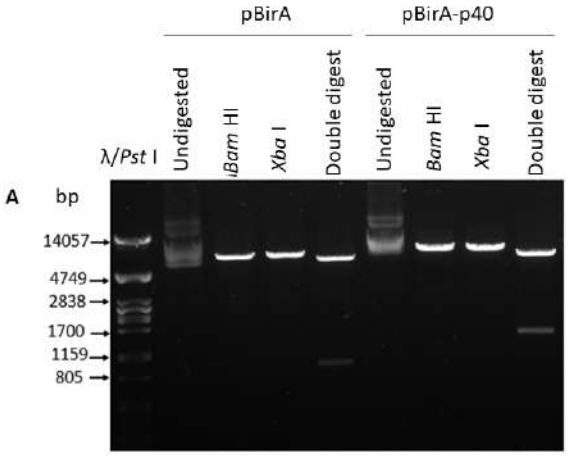
Your senior analyst for this analysis is Maré Vlok; do not hesitate to contact him on (021) 938 9469 or marevlok@sun.ac.za for additional discussion or information on this report.

Data Repository

All the data generated, including raw files, will be available on the SU proteomics server for 1 week after delivery of this report. After this period all data will be permanently removed from our storage system. **Please download and verify the data using the tool available in the tools folder within this time frame.** For bigger projects data can be supplied on external hard drives. The data will be stored in a folder named according to project name.

The proteomics ftp server can be accessed using an ftp client such as FileZilla (https://filezilla-project.org/download.php).

Appendix A2



B

	Single Digest	Double Digest
BirA	7200 bp	890 and 6300 bp
BirA-p40	8300 bp	1400 and 6900 bp

Figure A2: Validation of pBirA and pBirA-p40 constructs using restriction enzyme digests. The plasmids were digested with either *Bam* HI, *Xba* I or both restriction enzymes (double digest). (A) The restriction digests were analysed on a 1 % (w/v) agarose gel containing SYBR Safe. (B) A summary of the bands visualised on the gel. Both *Bam* HI and *Xba* I resulted in a single band of approximately 7200 bp (pBirA) and 8300 bp (pBirA-p40). The double digest resulted in two bands of 890 and 6300 bp for pBirA and 1400 and 6900 bp for pBirA-p40.

Appendix A3

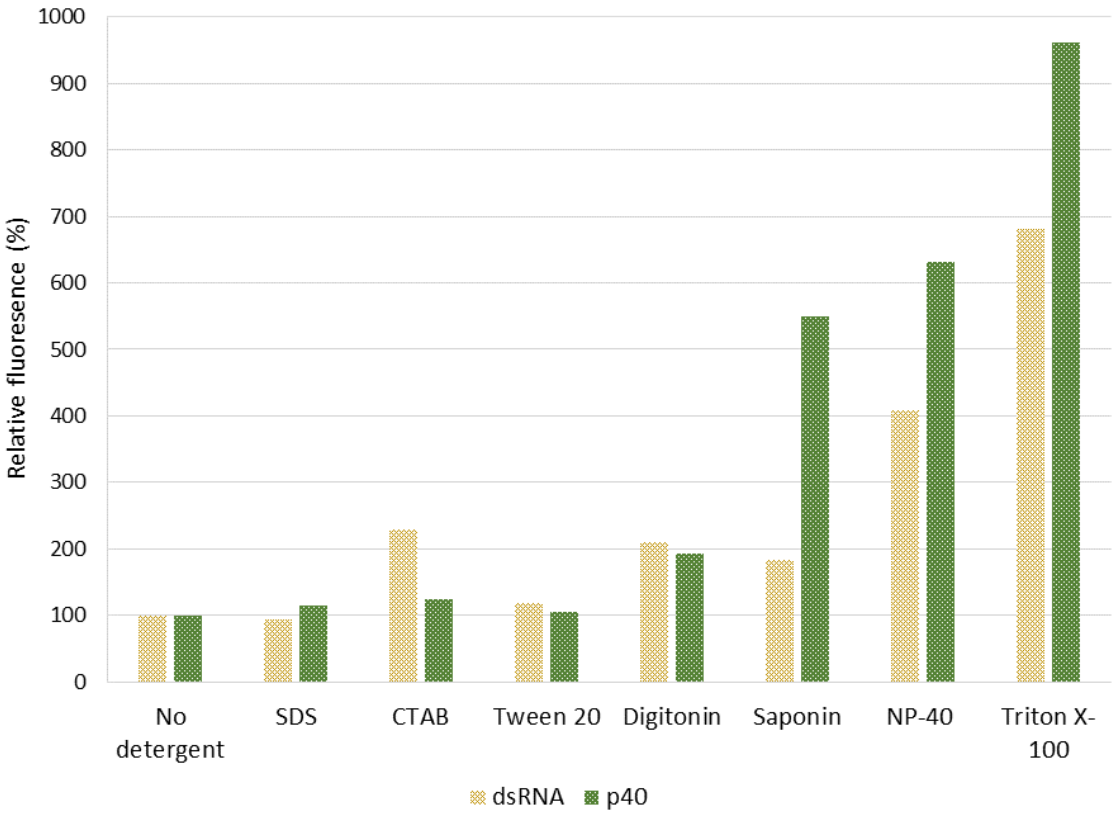


Figure A3: Relative fluorescence attributed to dsRNA and p40 in PrV-infected HeLa cells that were treated with detergents. The level dsRNA and p40 fluorescence of the no detergent treated (S Buffer-treated) cells was normalized to 100 %. All other treatments are reflected in comparison to the S buffer-treated cells.

Appendix A4

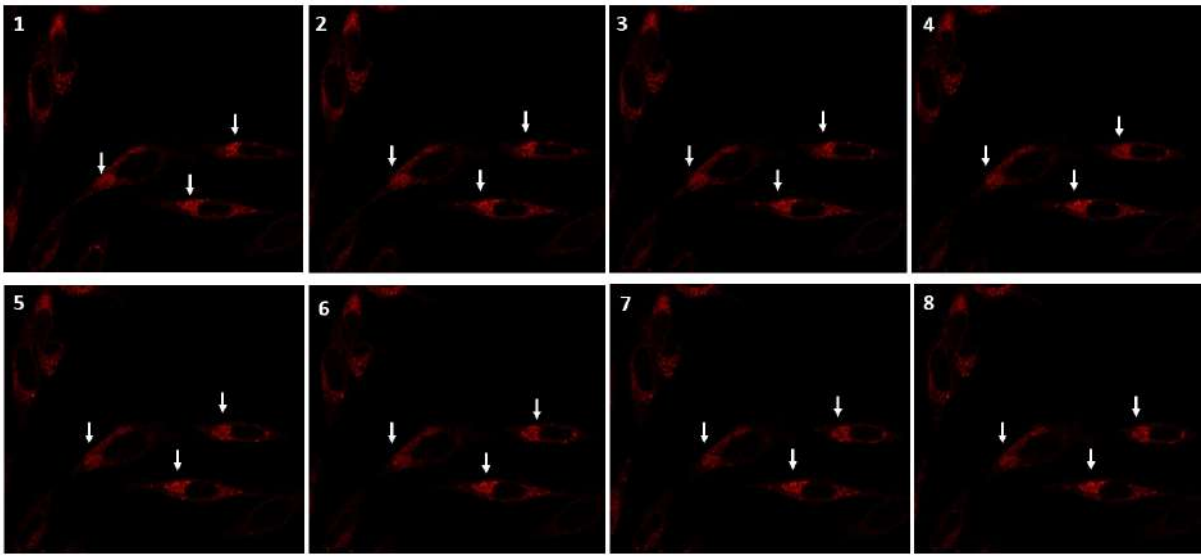


Figure A4: Live cell imaging of MitoTracker-stained HeLa cells treated with Triton X-100. PrV-infected HeLa cells were plated in Ibidi chambers and permitted to settle overnight. The cells were stained with MitoTracker 4 hours prior to detergent treatment. The cells were treated with 0.1 % Triton X-100 and then visualised over a period of 15 minutes.

Appendix A5

PrV antibodies and how they were generated:

1. P113 and p17:
   Both anti-p113 and anti-p17 were peptide antibodies generated by Mpho Peters. The information gathered about these antibodies was via personal communication as a thesis has yet to be submitted. The specific details of why and how the antibodies were generated and the efficacy of the antibodies are therefore unavailable. The anti-p17 antibody was raised in rabbits and recognises the amino acids between 1 and 130. The anti-p113 was raised in mice and recognises the amino acid sequence between 595 and 611.

2. P40:
   The anti-p40 antibody was generated by creating a GST fusion protein. GST was fused to the N terminus of the p40/p104 amino acid sequence (248 to 331). This protein was expressed in *E. coli* BL21(DE) cells and affinity purified. The purified protein was used to generate anti-p40 antibodies raised in rabbits. The antibody was found to be highly non-specific and so the rabbit-specific epitopes were masked by biotinylation to improve the specificity. (Generated by Cheryl Walter).

3. P104:
   The region, 889 to 902 was selected and the synthesized peptide was used to generate an antibody raised in rabbits. However, the antibody was found to have a low affinity and could only be used for immunoprecipitation reactions, not immunofluorescence microscopy. (Generated by Ritah Nakayinga).

4. VCAP:
   The PrV particles were purified. The whole virus was used to generate anti-capsid antibodies which were raised in rabbits. The antibody recognises and binds to the whole virus particle and the 60 kDa mature capsid protein. (Generated by Cheryl Walter).

Appendix A6

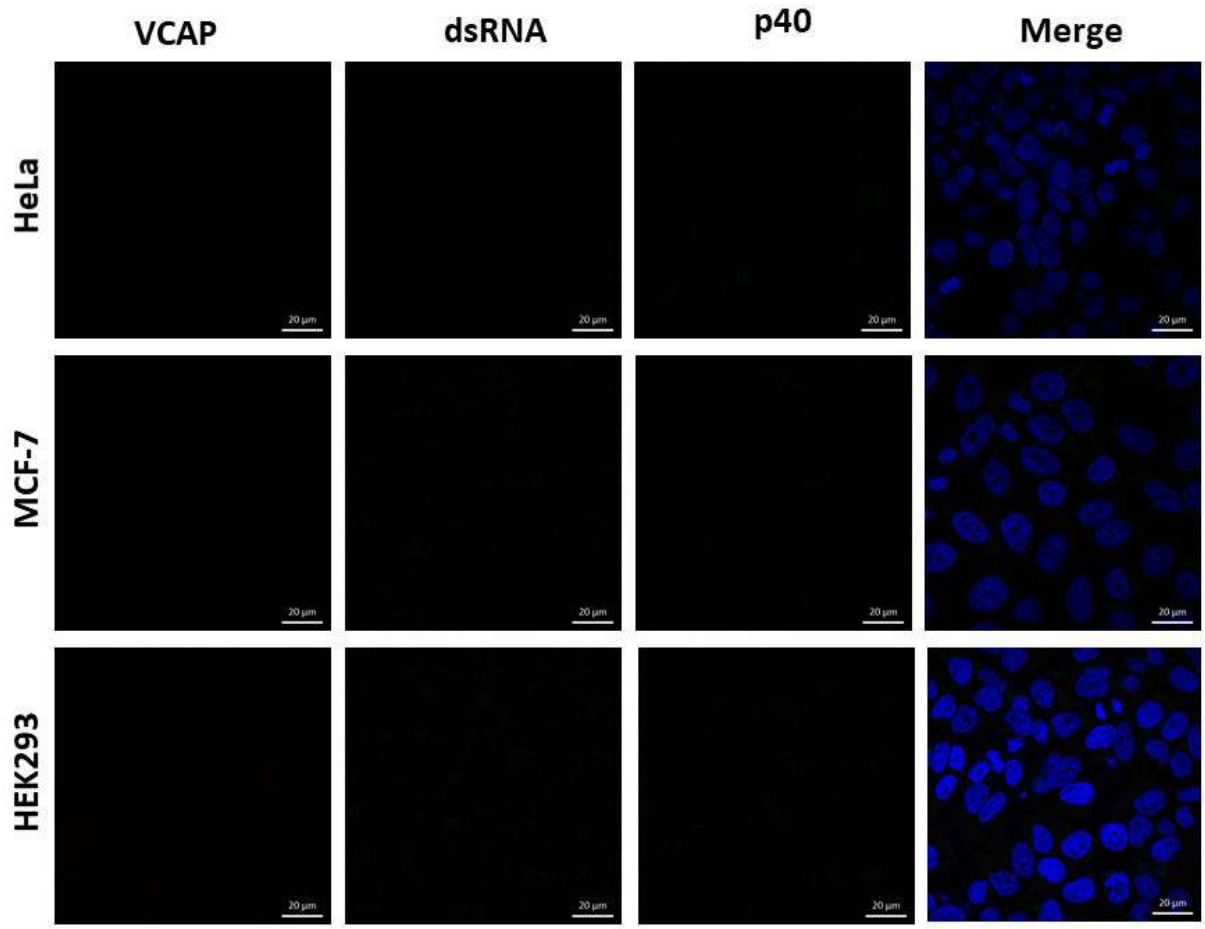


Figure A6: Uninfected HeLa, MCF-7 and HEK293 cells analysed by confocal microscopy. The cells were fixed with 4% paraformaldehyde and then prepared for immunofluorescence analysis. The cells were probed with anti-VCAP (rabbit), anti-dsRNA (mouse) and IgG-biotin anti-p40 primary antibodies and goat anti-rabbit AF 633, goat anti-mouse AF 546 and streptavidin AF 488 secondary antibodies. The nuclei were stained with DAPI.

Appendix A7

Table A7: The PrV protein peptides identified by mass spectrometry

File Name	Protein(s)	Peptide	Longest Amino Acid Sequence
PrV Capsid IP	PrV_B_F1.93	YPPPSSILI	PSSI
PrV Capsid IP	PrV_B_F3.81	KQQITRRKK	QQIT
PrV Capsid IP	PrV_B_F3.81	KQQITRRKK	QQIT
PrV Capsid IP	PrV_T_F2.47	LIIWIT	IIWI
PrV Capsid IP	PrV_T_F2.74	LEALTIQAESSSVWRSMVTSTQCGIIPPQ	IIPP

File Name	**Protein(s)**	**Peptide**	**Longest Amino Acid Sequence**
PrV p113 IP	PrV_B_F1.80	PPFSPTAAPYPVTIVLSPT	VTIV
PrV p113 IP	PrV_B_F3.11	SCRIIIDLV	IIID
PrV p113 IP	PrV_B_F3.14	LYVIEILR	VIEI
PrV p113 IP	PrV_B_F3.81	KQQITRRKK	QQIT
PrV p113 IP	PrV_T_F1.19	NLLLTTIGGPKEIT	PKEI
PrV p113 IP	PrV_T_F1.26	LPAVVCLPDQ	PAVV
PrV p113 IP	PrV_T_F2.47	LIIWIT	IIWI
PrV p113 IP	PrV_T_F3.56	KVPPVTPVATSMLLPVMALVASSIQPSTAPL	SSIQ
File Name	**Protein(s)**	**Peptide**	**Longest Amino Acid Sequence**
PrV p104 IP	PrV_B_F3.11	SCRIIIDLV	IIID
PrV p104 IP	PrV_B_F3.14	LYVIEIIR	IEII
PrV p104 IP	PrV_B_F3.14	LYVIEILR	VIEI
PrV p104 IP	PrV_B_F3.18	LLLLGCR	LLLLGCR
PrV p104 IP	PrV_B_F3.81	KQQITRRKK	QQIT
PrV p104 IP	PrV_T_F1.26	LPAVVCLPDQ	PAVV
PrV p104 IP	PrV_T_F2.47	LIIWIT	IIWI
File Name	**Protein(s)**	**Peptide**	**Longest Amino Acid Sequence**
PrV p17 IP	PrV_B_F1.5	DPQWIQR	QWIQ
PrV p17 IP	PrV_B_F1.94	VCATIIPLLQ	ATII
PrV p17 IP	PrV_B_F3.11	SCRIIIDLV	IIID
PrV p17 IP	PrV_B_F3.14	LYVIEIIR	IEII
PrV p17 IP	PrV_B_F3.14	LYVIEILR	VIEI
PrV p17 IP	PrV_B_F3.14	LYVIEIIR	IEII
PrV p17 IP	PrV_B_F3.14	LYVIEILR	VIEI
PrV p17 IP	PrV_B_F3.81	KQQITRRKK	QQIT
PrV p17 IP	PrV_T_F1.26	LPAVVCLPDQ	PAVV
PrV p17 IP	PrV_T_F1.26	FIRVICC	IRVI
PrV p17 IP	PrV_T_F2.47	LIIWIT	IIWI
PrV p17 IP	PrV_T_F3.56	KVPPVTPVATSMLLPVMALVASSIQPSTAPL	SSIQ
File Name	**Protein(s)**	**Peptide**	**Longest Amino Acid Sequence**
PrV p40 IP	PrV_B_F1.8	LPIISGLASQYSTD	PIIS
PrV p40 IP	PrV_B_F2.53	TLTLPPGLPVVLFPALFPSQIAAGCGCN	IAAG

PrV p40 IP	PrV_B_F3.14	LYVIEIIR	IEII
PrV p40 IP	PrV_B_F3.14	LYVIEILR	VIEI
PrV p40 IP	PrV_B_F3.14	LYVLELLR	YVLELLR
PrV p40 IP	PrV_B_F3.14	LYVIEIIR	IEII
PrV p40 IP	PrV_B_F3.14	LYVIEIIR	IEII
PrV p40 IP	PrV_B_F3.14	LYVIEILR	VIEI
PrV p40 IP	PrV_B_F3.81	KQQITRRKK	QQIT
PrV p40 IP	PrV_T_F1.26	LPAVVCLPDQ	PAVV
PrV p40 IP	PrV_T_F1.26	FIRVICC	IRVI
PrV p40 IP	PrV_T_F2.21	YYPVLTPGK	PVLTPGK
PrV p40 IP	PrV_T_F2.47	LIIWIT	IIWI
PrV p40 IP	PrV_T_F3.56	KVPPVTPVATSMLLPVMALVASSIQPSTAPL	SSIQ

References

Alcaïs, A., Abel, L., & Casanova, J. L. (2009). Human genetics of infectious diseases: between proof of principle and paradigm. *The Journal of clinical investigation*, *119*(9), 2506-2514.

Ali, Z., Wang, J., Tang, Y., Liu, B., He, N., & Li, Z. (2017). Simultaneous detection of multiple viruses based on chemiluminescence and magnetic separation. *Biomaterials science*, *5*(1), 57-66.

Akil, L., & Ahmad, H. A. (2016). The recent outbreaks and reemergence of poliovirus in war and conflict-affected areas. *International Journal of Infectious Diseases*, *49*, 40-46.

Allocati, N., Petrucci, A. G., Di Giovanni, P., Masulli, M., Di Ilio, C., & De Laurenzi, V. (2016). Bat–man disease transmission: zoonotic pathogens from wildlife reservoirs to human populations. *Cell death discovery*, *2*, 16048.

Älgenäs, C., Agaton, C., Fagerberg, L., Asplund, A., Björling, L., Björling, E., ... & Wester, K. (2014). Antibody performance in western blot applications is context-dependent. *Biotechnology journal*, *9*(3), 435-445.

Amarasinghe, G. K., Ayllón, M. A., Bào, Y., Basler, C. F., Bavari, S., Blasdell, K. R., ... & Buchholz, U. J. (2019). Taxonomy of the order Mononegavirales: update 2019. *Archives of virology*, *164*(7), 1967-1980.

Andersen, K. G., Shapiro, B. J., Matranga, C. B., Sealfon, R., Lin, A. E., Moses, L. M., ... & Momoh, M. (2015). Clinical sequencing uncovers origins and evolution of Lassa virus. *Cell*, *162*(4), 738-750.

Atkinson, B., Thorburn, F., Petridou, C., Bailey, D., Hewson, R., Simpson, A. J., ... & Aarons, E. J. (2017). Presence and persistence of Zika virus RNA in semen, United Kingdom, 2016. *Emerging infectious diseases*, *23*(4), 611.

Bah, E. I., Lamah, M. C., Fletcher, T., Jacob, S. T., Brett-Major, D. M., Sall, A. A., ... & Bausch, D. G. (2015). Clinical presentation of patients with Ebola virus disease in Conakry, Guinea. *New England Journal of Medicine*, *372*(1), 40-47.

Balinsky, C. A., Schmeisser, H., Ganesan, S., Singh, K., Pierson, T. C., & Zoon, K. C. (2013). Nucleolin interacts with the dengue virus capsid protein and plays a role in formation of infectious virus particles. *Journal of virology*, *87*(24), 13094-13106.

Ball, L. A., & Johnson, K. L. (1999). Reverse genetics of nodaviruses. In *Advances in virus research* (Vol. 53, pp. 229-244). Academic Press.

Bandín, I., & Dopazo, C. P. (2011). Host range, host specificity and hypothesized host shift events among viruses of lower vertebrates. *Veterinary research*, *42*(1), 67.

Barathan, M., Mohamed, R., Yong, Y. K., Kannan, M., Vadivelu, J., Saeidi, A., ... & Shankar, E. M. (2018). Viral persistence and chronicity in hepatitis C virus infection: role of T-cell apoptosis, senescence and exhaustion. *Cells*, *7*(10), 165.

Barman, S., & Nayak, D. P. (2007). Lipid raft disruption by cholesterol depletion enhances Influenza A virus budding from MDCK cells. *Journal of virology*, *81*(22), 12169-12178.

Bartenschlager, R., Baumert, T. F., Bukh, J., Houghton, M., Lemon, S. M., Lindenbach, B. D., ... & Thimme, R. (2018). Critical challenges and emerging opportunities in hepatitis C virus research in an era of potent antiviral therapy: considerations for scientists and funding agencies. Virus research, 248, 53-62.

Bavari, S., Bosio, C. M., Wiegand, E., Ruthel, G., Will, A. B., Geisbert, T. W., ... & Aman, M. J. (2002). Lipid raft microdomains: a gateway for compartmentalized trafficking of Ebola and Marburg viruses. *Journal of Experimental Medicine*, *195*(5), 593-602.

Benelli, G., & Mehlhorn, H. (2016). Declining malaria, rising of dengue and Zika virus: insights for mosquito vector control. *Parasitology research*, *115*(5), 1747-1754.

Besemer, J., Lomsadze, A., & Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic acids research*, *29*(12), 2607-2618.

Bhatnagar, J., Rabeneck, D. B., Martines, R. B., Reagan-Steiner, S., Ermias, Y., Estetter, L. B., ... & Gary, J. (2017). Zika virus RNA replication and persistence in brain and placental tissue. *Emerging infectious diseases*, *23*(3), 405.

Bird, S. W., & Kirkegaard, K. (2015). Escape of non-enveloped virus from intact cells. *Virology*, *479*, 444-449.

Bird, S. W., Maynard, N. D., Covert, M. W., & Kirkegaard, K. (2014). Nonlytic viral spread enhanced by autophagy components. *Proceedings of the National Academy of Sciences*, *111*(36), 13081-13086.

Blair, W., & Cox, C. (2016). Current landscape of antiviral drug discovery. *F1000Research*, *5*.

Boldogh, I., Albrecht, T., & Porter, D. D. (1996). Persistent viral infections. In *Medical Microbiology. 4th edition*. University of Texas Medical Branch at Galveston.

Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P., & Rohwer, F. (2003). Metagenomic analyses of an uncultured viral community from human feces. *Journal of bacteriology*, *185*(20), 6220-6223.

Breman, J. G., Arita, I., Unit, S. E., & World Health Organization. (1980). *The confirmation and maintenance of smallpox eradication* (No. WHO/SE/80.156). Geneva, Switzerland: World Health Organization.

Brizzard, B. (2008). Epitope tagging. *Biotechniques*, *44*(5), 693-695.

Brooks, D. R., Hoberg, E. P., & Boeger, W. A. (2019). *The Stockholm paradigm: Climate change and emerging disease*. University of Chicago Press.Brown, D. A. (2006). Lipid rafts, detergent-resistant membranes, and raft targeting signals. *Physiology*, *21*(6), 430-439.

Bukh, J. (2016). The history of hepatitis C virus (HCV): Basic research reveals unique features in phylogeny, evolution and the viral life cycle with new perspectives for epidemic control. *Journal of hepatology*, *65*(1), S2-S21.

Burke, K. P., & Cox, A. L. (2010). Hepatitis C virus evasion of adaptive immune responses: a model for viral persistence. *Immunologic research*, *47*(1-3), 216-227.

Cable, J. (2020). The Stockholm paradigm. *The Lancet Infectious Diseases*.

Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V., & Schountz, T. (2006). Bats: important reservoir hosts of emerging viruses. *Clinical microbiology reviews*, *19*(3), 531-545.

Callaway, E. (2019). 'Make Ebola a thing of the past': first vaccine against deadly virus approved. [online] Nature.com. Available at: https://www.nature.com/articles/d41586-019-03490-8. [Accessed 23 Nov. 2019].

Cambridge Dictionary. (2019). Available at: https://dictionary.cambridge.org/dictionary/english/providence. (Accessed 4 Dec. 2019).

Cantlay, J. C., Ingram, D. J., & Meredith, A. L. (2017). A review of zoonotic infection risks associated with the wild meat trade in Malaysia. *EcoHealth*, *14*(2), 361-388.

Carrasco-Hernandez, R., Jácome, R., López Vidal, Y., & Ponce de León, S. (2017). Are RNA viruses candidate agents for the next global pandemic? A review. *ILAR journal*, *58*(3), 343-358.

CEPI, (2019). *Priority diseases – CEPI*. [online] Available at: https://cepi.net/research_dev/priority-diseases/. (Accessed 25 Nov. 2019).

Chacinska, A., Koehler, C. M., Milenkovic, D., Lithgow, T., & Pfanner, N. (2009). Importing mitochondrial proteins: machineries and mechanisms. *Cell*, *138*(4), 628-644.

Choi-Rhee, E., Schulman, H., & Cronan, J. E. (2004). Promiscuous protein biotinylation by *Escherichia coli* biotin protein ligase. *Protein science*, *13*(11), 3043-3050.

Coffey, L. L., Beeharry, Y., Bordería, A. V., Blanc, H., & Vignuzzi, M. (2011). Arbovirus high fidelity variant loses fitness in mosquitoes and mice. *Proceedings of the National Academy of Sciences*, *108*(38), 16038-16043.

Cook, S., Chung, B. Y. W., Bass, D., Moureau, G., Tang, S., McAlister, E., ... & Gould, E. A. (2013). Novel virus discovery and genome reconstruction from field RNA samples reveals highly divergent viruses in dipteran hosts. *PloS one*, *8*(11), e80720.

Cozzetto, D., Minneci, F., Currant, H., & Jones, D. T. (2016). FFPred 3: feature-based function prediction for all Gene Ontology domains. *Scientific reports*, *6*, 31865.

Daszak, P., Carroll, D., Wolfe, N., & Mazet, J. (2016). The global virome project. *Int. J. Infect. Dis. S*, *53*, 4-163.

Debing, Y., Jochmans, D., & Neyts, J. (2013). Intervention strategies for emerging viruses: use of antivirals. *Current opinion in virology*, *3*(2), 217-224.

De Clercq, E., & Li, G. (2016). Approved antiviral drugs over the past 50 years. *Clinical microbiology reviews*, *29*(3), 695-747.

de Lamballerie, C. N., Pizzorno, A., Dubois, J., Julien, T., Padey, B., Bouveret, M., ... & Terrier, O. (2019). Characterization of cellular transcriptomic signatures induced by different respiratory viruses in human reconstituted airway epithelia. *Scientific reports*, *9*(1), 1-12.

de Silva, A. M., & Harris, E. (2018). Which dengue vaccine approach is the most promising, and should we be concerned about enhanced disease after vaccination? The path to a dengue vaccine: learning from human natural dengue infection studies and vaccine trials. *Cold Spring Harbor perspectives in biology*, *10*(6), a029371.

Devaux, C. A. (2012). Emerging and re-emerging viruses: A global challenge illustrated by Chikungunya virus outbreaks. *World journal of virology*, *1*(1), 11.

Dinan, A. M., Lukhovitskaya, N. I., Olendraite, I., & Firth, A. E. (2019). A case for a reverse-frame coding sequence in a group of positive-sense RNA viruses. *bioRxiv*, 664342.

Dorrington, R. A., Gorbalenya, A. E., Gordon, K. H. J., Lauber, C., & Ward, V. K. (2011). Family *Tetraviridae*. *Virus Taxonomy*, 1091-1102.

El Ekiaby, M., Lelie, N., & Allain, J. P. (2010). Nucleic acid testing (NAT) in high prevalence–low resource settings. *Biologicals*, *38*(1), 59-64.

Firth, A. E., Wills, N. M., Gesteland, R. F., & Atkins, J. F. (2011). Stimulation of stop codon readthrough: frequent presence of an extended 3′ RNA structural element. *Nucleic acids research*, *39*(15), 6679-6691.

Firth, A. E. (2014). Mapping overlapping functional elements embedded within the protein-coding regions of RNA viruses. *Nucleic acids research*, *42*(20), 12425-12439.

Firth, A. E., & Brierley, I. (2012). Non-canonical translation in RNA viruses. *The Journal of general virology*, *93*(Pt 7), 1385.

Fraser, J. E., Rawlinson, S. M., Heaton, S. M., & Jans, D. A. (2016). Dynamic nucleolar targeting of dengue virus polymerase NS5 in response to extracellular pH. *Journal of virology*, *90*(12), 5797-5807.

Fujimura, T., Solórzano, A., & Esteban, R. (2005). Native replication intermediates of the yeast 20 S RNA virus have a single-stranded RNA backbone. *Journal of Biological Chemistry*, *280*(8), 7398-7406.

García-Castillo, S., Sanchez-Pina, M. A., & Pallas, V. (2003). Spatio-temporal analysis of the RNAs, coat and movement (p7) proteins of Carnation mottle virus in Chenopodium quinoa plants. *Journal of general virology*, *84*(3), 745-749.

Garcia-Sastre, A., & Mena, I. (2013). Novel vaccine strategies against emerging viruses. *Current opinion in virology*, *3*(2), 210-216.

Gatherer, D. (2014). The 2014 Ebola virus disease outbreak in West Africa. *Journal of general virology*, *95*(8), 1619-1624.

Ge, X., Li, Y., Yang, X., Zhang, H., Zhou, P., Zhang, Y., & Shi, Z. (2012). Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *Journal of virology*, *86*(8), 4620-4630.

Geoghegan, J. L., & Holmes, E. C. (2017). Predicting virus emergence amid evolutionary noise. Open biology, 7(10), 170189.

Gire, S. K., Goba, A., Andersen, K. G., Sealfon, R. S., Park, D. J., Kanneh, L., ... & Wohl, S. (2014). Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *science*, *345*(6202), 1369-1372.

Goldberg, T., Hecht, M., Hamp, T., Karl, T., Yachdav, G., Ahmed, N., ... & Bernhofer, M. (2014). LocTree3 prediction of localization. *Nucleic acids research*, *42*(W1), W350-W355.

Goldstein, T., Anthony, S. J., Gbakima, A., Bird, B. H., Bangura, J., Tremeau-Bravard, A., ... & Grodus, M. (2018). The discovery of Bombali virus adds further support for bats as hosts of ebolaviruses. *Nature microbiology*, *3*(10), 1084.

Golumbeanu, M., Desfarges, S., Hernandez, C., Quadroni, M., Rato, S., Mohammadi, P., ... & Ciuffi, A. (2019). Proteo-transcriptomic dynamics of cellular response to hiv-1 infection. *Scientific reports*, *9*(1), 213.

Gómez-Aix, C., García-García, M., Aranda, M. A., & Sánchez-Pina, M. A. (2015). Melon necrotic spot virus replication occurs in association with altered mitochondria. *Molecular Plant-Microbe Interactions*, *28*(4), 387-397.

Goodrum, F., Caviness, K., & Zagallo, P. (2012). Human cytomegalovirus persistence. *Cellular microbiology*, *14*(5), 644-655.

Gordon, K. H., Williams, M. R., Hendry, D. A., & Hanzlik, T. N. (1999). Sequence of the genomic RNA of Nudaurelia β virus (Tetraviridae) defines a novel virus genome organization. *Virology*, *258*(1), 42-53.

Gosert, R., Egger, D., Lohmann, V., Bartenschlager, R., Blum, H. E., Bienz, K., & Moradpour, D. (2003). Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons. *Journal of virology*, *77*(9), 5487-5492.

Gould, E., Pettersson, J., Higgs, S., Charrel, R., & de Lamballerie, X. (2017). Emerging arboviruses: why today?. *One Health*, *4*, 1-13.

Grebely, J., Hajarizadeh, B., & Dore, G. J. (2017). Direct-acting antiviral agents for HCV infection affecting people who inject drugs. *Nature reviews Gastroenterology & hepatology*, *14*(11), 641.

Griffis, E. R., Xu, S., & Powers, M. A. (2003). Nup98 localizes to both nuclear and cytoplasmic sides of the nuclear pore and binds to two distinct nucleoporin subcomplexes. *Molecular biology of the cell*, *14*(2), 600-610.

Guo, C., Chen, L., Mo, D., Chen, Y., & Liu, X. (2015). DRACO inhibits porcine reproductive and respiratory syndrome virus replication in vitro. *Archives of virology*, *160*(5), 1239-1247.

Hadi, C. M., Goba, A., Khan, S. H., Bangura, J., Sankoh, M., Koroma, S., ... & Bausch, D. G. (2010). Ribavirin for Lassa fever postexposure prophylaxis. *Emerging infectious diseases*, *16*(12), 2009.

Hall, R. J., Wang, J., Todd, A. K., Bissielo, A. B., Yen, S., Strydom, H., ... & Peacey, M. (2014). Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *Journal of virological methods*, *195*, 194-204.

Hanne, J., Zila, V., Heilemann, M., Müller, B., & Kräusslich, H. G. (2016). Super-resolved insights into human immunodeficiency virus biology. *FEBS letters*, *590*(13), 1858-1876.

Hanzlik, T. N., & Gordon, K. H. (1997). The tetraviridae. In *Advances in virus research* (Vol. 48, pp. 101-168). Academic Press.

Harrell, L., Melcher, U., & Atkins, J. F. (2002). Predominance of six different hexanucleotide recoding signals 3′ of read-through stop codons. *Nucleic acids research*, *30*(9), 2011-2017.

Harrower, J., Kiedrzynski, T., Baker, S., Upton, A., Rahnama, F., Sherwood, J., ... & Pulford, D. (2016). Sexual transmission of Zika virus and persistence in semen, New Zealand, 2016. *Emerging infectious diseases*, *22*(10), 1855.

Heaton, N. S., & Randall, G. (2011). Dengue virus and autophagy. *Viruses*, *3*(8), 1332-1341.

Heeney, J. L. (2015). Ebola: hidden reservoirs. Nature, 527(7579), 453.

Heerklotz, H. (2002). Triton promotes domain formation in lipid raft mixtures. *Biophysical journal*, *83*(5), 2693-2701.

Heylen, E., Neyts, J., & Jochmans, D. (2017). Drug candidates and model systems in respiratory syncytial virus antiviral drug discovery. *Biochemical pharmacology*, *127*, 1-12.

Hillman, B. I., & Cai, G. (2013). The family Narnaviridae: simplest of RNA viruses. In *Advances in virus research* (Vol. 86, pp. 149-176). Academic Press.

Hirano, M., Kaneko, S., Yamashita, T., Luo, H., Qin, W., Shirota, Y., ... & Murakami, S. (2003). Direct interaction between nucleolin and hepatitis C virus NS5B. *Journal of Biological Chemistry*, *278*(7), 5109-5115.

Hiscox, J. A., Wurm, T., Wilson, L., Britton, P., Cavanagh, D., & Brooks, G. (2001). The coronavirus infectious bronchitis virus nucleoprotein localizes to the nucleolus. *Journal of Virology*, *75*(1), 506-512.

Hiscox, J. A. (2007). RNA viruses: hijacking the dynamic nucleolus. *Nature Reviews Microbiology*, *5*(2), 119.

Hoenen, T., Groseth, A., & Feldmann, H. (2019). Therapeutic strategies to target the Ebola virus life cycle. *Nature Reviews Microbiology*, *17*(10), 593-606.

Hull, R. (2014). Movement of viruses within plants In: Plant Virology (Hull R., editor., ed.).

Irwin, K. K., Renzette, N., Kowalik, T. F., & Jensen, J. D. (2016). Antiviral drug resistance as an adaptive process. *Virus evolution*, *2*(1).

Janeway Jr, C. A., Travers, P., Walport, M., & Shlomchik, M. J. (2001). The structure of a typical antibody molecule. In *Immunobiology: The Immune System in Health and Disease. 5th edition*. Garland Science.

Jarvie, R.A. (2017). Targeted induction of apoptosis in Providence virus-infected mammalian cell cultures. Honours thesis, Rhodes University, South Africa.

Jiwaji, M., Matcher, G. F., de Bruyn, M. M., Awando, J. A., Moodley, H., Waterworth, D., Jarvie, R. A., & Dorrington, R. A. (2019). Providence virus: An animal virus that replicates in plants or a plant virus that infects and replicates in animal cells?. *PloS one*, *14*(6), e0217494.

Jiwaji, M., Short, J. R., & Dorrington, R. A. (2016). Expanding the host range of small insect RNA viruses: Providence virus (Carmotetraviridae) infects and replicates in a human tissue culture cell line. *Journal of General Virology*, *97*(10), 2763-2768.

Johnstone, S.A., Waisman, D.M., & Rattner, J.B. (1992). Enolase is present at the centrosome of HeLa cells. *Experimental Cell Research 202*, 458-463.

Jones, C. T., Patkar, C. G., & Kuhn, R. J. (2005). Construction and applications of yellow fever virus replicons. *Virology*, *331*(2), 247-259.

Kane, M., & Golovkina, T. (2010). Common threads in persistent viral infections. *Journal of virology*, *84*(9), 4116-4123.

Kemenesi, G., Földes, F., Zana, B., Kurucz, K., Estók, P., Boldogh, S., ... & Jakab, F. (2016). Genetic characterization of Providence virus isolated from bat guano in Hungary. *Genome Announc.*, *4*(3), e00403-16.

Kemenesi, G., Kurucz, K., Dallos, B., Zana, B., Földes, F., Boldogh, S., ... & Jakab, F. (2018). Re-emergence of Lloviu virus in Miniopterus schreibersii bats, Hungary, 2016. *Emerging microbes & infections*, *7*(1), 1-4.

Khandia, R., Munjal, A., Dhama, K., Karthik, K., Tiwari, R., Malik, Y. S., ... & Chaicumpa, W. (2018). Modulation of dengue/zika virus pathogenicity by antibody-dependent enhancement and strategies to protect against enhancement in zika virus infection. *Frontiers in immunology*, *9*, 597.

Kim, C. R., Counotte, M., Bernstein, K., Deal, C., Mayaud, P., Low, N., & Broutet, N. (2018). Investigating the sexual transmission of Zika virus. *The Lancet Global Health*, *6*(1), e24-e25.

Kim, S. H., Ryabov, E. V., Brown, J. W. S., & Taliansky, M. (2004). Involvement of the nucleolus in plant virus systemic infection.

Komar, A. A., & Hatzoglou, M. (2011). Cellular IRES-mediated translation: the war of ITAFs in pathophysiological states. *Cell cycle*, *10*(2), 229-240.

Kuno, G. (2001). Persistence of arboviruses and antiviral antibodies in vertebrate hosts: its occurrence and impacts. *Reviews in medical virology*, *11*(3), 165-190.

Kwon, K., & Beckett, D. (2000). Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein Science*, *9*(8), 1530-1539.

Lajko, M., Haddad, A. F., Robinson, C. A., & Connolly, S. A. (2015). Using proximity biotinylation to detect herpesvirus entry glycoprotein interactions: limitations for integral membrane glycoproteins. *Journal of virological methods*, *221*, 81-89.

Lauring, A. S., & Andino, R. (2010). Quasispecies theory and the behavior of RNA viruses. *PLoS pathogens*, *6*(7), e1001005.

Leech, S., & Baker, M. L. (2017). The interplay between viruses and the immune system of bats. *Microbiology Australia*, *38*(1), 30-32.

Leroy, E. M., Rouquet, P., Formenty, P., Souquière, S., Kilbourne, A., Froment, J. M., ... & Zaki, S. R. (2004). Multiple Ebola virus transmission events and rapid decline of central African wildlife. *Science*, *303*(5656), 387-390.

Le Sage, V., Cinti, A., Valiente-Echeverría, F., & Mouland, A. J. (2015). Proteomic analysis of HIV-1 Gag interacting partners using proximity-dependent biotinylation. *Virology journal*, *12*(1), 138.

Lessler, J., & Cummings, D. A. (2016). Mechanistic models of infectious disease and their impact on public health. *American journal of epidemiology*, *183*(5), 415-422.

Li, Z., Pogany, J., Panavas, T., Xu, K., Esposito, A. M., Kinzy, T. G., & Nagy, P. D. (2009). Translation elongation factor 1A is a component of the tombusvirus replicase complex and affects the stability of the p33 replication co-factor. *Virology*, *385*(1), 245-260.

Lian, J. L., Cornman, R. S., Evans, J. D., Pettis, J. S., Zhao, Y., Murphy, C., ... & Zhou, L. (2014). Systemic spread and propagation of a plant-pathogenic virus in European honeybees. *Apis mellifera*, 00898-13.

Linke, D. (2009). Detergents: an overview. In *Methods in enzymology* (Vol. 463, pp. 603-617). Academic Press.

Lingwood, D., & Simons, K. (2007). Detergent resistance as a tool in membrane research. *Nature protocols*, *2*(9), 2159.

Lipsitch, M., Plotkin, J. B., Simonsen, L., & Bloom, B. (2012). Evolution, safety, and highly pathogenic Influenza viruses. *Science*, *336*(6088), 1529-1531.

Lipton, J. H., & McMurray, W. C. (1977). Mitochondrial biogenesis in cultured animal cells I. Effect of chloramphenicol on morphology and mitochondrial respiratory enzymes. *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis*, *477*(3), 264-272.

Liu, J., Dai, S., Wang, M., Hu, Z., Wang, H., & Deng, F. (2016). Virus like particle-based vaccines against emerging infectious disease viruses. Virologica Sinica, 31(4), 279-287.

Longdon, B., Brockhurst, M. A., Russell, C. A., Welch, J. J., & Jiggins, F. M. (2014). The evolution and genetics of virus host shifts. *PLoS pathogens*, *10*(11), e1004395.

Lozsan, A., Rivas, I., Rodriguez, G., Martinez, S., & Pérez, M. Á. (2017). Determination of Surface-Active Characteristics of a Natural Surfactant Extracted from Sapindus Saponaria. *Tenside Surfactants Detergents*, *54*(2), 109-117.

Luis, A. D., O'Shea, T. J., Hayman, D. T., Wood, J. L., Cunningham, A. A., Gilbert, A. T., ... & Webb, C. T. (2015). Network analysis of host–virus communities in bats and rodents reveals determinants of cross-species transmission. *Ecology letters*, *18*(11), 1153-1162.

Luke, G. A., de Felipe, P., Lukashev, A., Kallioinen, S. E., Bruno, E. A., & Ryan, M. D. (2008). Occurrence, function and evolutionary origins of '2A-like'sequences in virus genomes. *The Journal of general virology*, *89*(Pt 4), 1036.

Lulla, V., Dinan, A. M., Hosmillo, M., Chaudhry, Y., Sherry, L., Irigoyen, N., ... & Firth, A. E. (2019). An upstream protein-coding region in enteroviruses modulates virus infection in gut epithelial cells. *Nature microbiology*, *4*(2), 280.

Manhart, W. A., Pacheco, J. R., Hume, A. J., Cressey, T. N., Deflubé, L. R., & Mühlberger, E. (2018). A chimeric Lloviu virus minigenome system reveals that the bat-derived filovirus replicates more similarly to ebolaviruses than marburgviruses. *Cell reports*, *24*(10), 2573-2580.

Marcos, J. F., Vilar, M., Pérez-Payá, E., & Pallás, V. (1999). In Vivo Detection, RNA-Binding Properties and Characterization of the RNA-Binding Domain of the p7 Putative Movement Protein from Carnation Mottle Carmovirus (CarMV). *Virology*, *255*(2), 354-365.

Mazumdar, T. (2019). *Ebola vaccine approved as second jab trialled*. [online] BBC News. Available at: https://www.bbc.com/news/health-50315970#:~:targetText=A%20second%20Ebola%20vaccine%20is,given%20to%20around%20250%2C000%20people. (Accessed 4 Dec. 2019).

Martínez-Turiño, S., & Hernández, C. (2012). Analysis of the subcellular targeting of the smaller replicase protein of Pelargonium flower break virus. *Virus research*, *163*(2), 580-591.

McKee, E. E., Ferguson, M., Bentley, A. T., & Marks, T. A. (2006). Inhibition of mammalian mitochondrial protein synthesis by oxazolidinones. *Antimicrobial agents and chemotherapy*, *50*(6), 2042-2049.

Mendes, A., Vlok, M., Short, J. R., Matsui, T., & Dorrington, R. A. (2015). An encapsidated viral protein and its role in RNA packaging by a non-enveloped animal RNA virus. *Virology*, *476*, 323-333.

Michel, M. R., Elgizoli, M., Dai, Y., Jakob, R., Koblet, H., & Arrigo, A. P. (1990). Karyophilic properties of Semliki Forest virus nucleocapsid protein. *Journal of virology*, *64*(10), 5123-5131.

Midgard, H., Weir, A., Palmateer, N., Re III, V. L., Pineda, J. A., Macías, J., & Dalgard, O. (2016). HCV epidemiology in high-risk groups and the risk of reinfection. *Journal of hepatology*, *65*(1), S33-S45.

Miller, D. J., Schwartz, M. D., & Ahlquist, P. (2001). Flock house virus RNA replicates on outer mitochondrial membranes in Drosophila cells. *Journal of virology*, *75*(23), 11664-11676.

Mitra, S., & Dungan, S. R. (1997). Micellar properties of Quillaja saponin. 1. Effects of temperature, salt, and pH on solution properties. *Journal of Agricultural and Food Chemistry*, *45*(5), 1587-1595.

Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E., & Hayakawa, T. (2000). IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. *Molecular Therapy*, *1*(4), 376-382.

Mohammadi, P., Ciuffi, A., & Beerenwinkel, N. (2015). Dynamic models of viral replication and latency. *Current Opinion in HIV and AIDS*, *10*(2), 90.

Moodley, H. (2019). An investigation into the interaction partners of the scaffold protein human CNK1 in the NF-κB pathway. MSc thesis, Rhodes University, South Africa.

Mouchet, J., & Carnevale, P. (1997). Impact of changes in the environment on vector-transmitted diseases. *Sante (Montrouge, France)*, *7*(4), 263-269.

Murant, A. F., & Kumar, T. L. I. (1990). Different variants of the satellite RNA of groundnut rosette virus are responsible for the chlorotic and green forms of groundnut rosette disease. *Annals of Applied Biology*, *117*(1), 85-92.

Nagy, P. D., & Pogany, J. (2008). Multiple roles of viral replication proteins in plant RNA virus replication. In *Plant Virology Protocols* (pp. 55-68). Humana Press.

Nakayinga, R. (2019). Computational characterization of providence virus non-structural proteins: Evolutionary and functional implications.

Nakayinga, R. (2013). Understanding the replication biology of Providence virus: elucidating the function of the non-structural proteins. PhD thesis, Rhodes University, South Africa.

Napthine, S., Ling, R., Finch, L. K., Jones, J. D., Bell, S., Brierley, I., & Firth, A. E. (2017). Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nature communications*, *8*, 15582.

Negredo, A., Palacios, G., Vázquez-Morón, S., González, F., Dopazo, H., Molero, F., ... & Herrera, J. E. (2011). Discovery of an ebolavirus-like filovirus in europe. *PLoS pathogens*, *7*(10), e1002304.

Neiderud, C. J. (2015). How urbanization affects the epidemiology of emerging infectious diseases. *Infection ecology & epidemiology*, *5*(1), 27060.

Nelson, N. P., Easterbrook, P. J., & McMahon, B. J. (2016). Epidemiology of hepatitis B virus infection and impact of vaccination on disease. *Clinics in liver disease*, *20*(4), 607-628.

Ng, C. Y., Gu, F., Phong, W. Y., Chen, Y. L., Lim, S. P., Davidson, A., & Vasudevan, S. G. (2007). Construction and characterization of a stable subgenomic dengue virus type 2 replicon system for antiviral compound and siRNA testing. *Antiviral research*, *76*(3), 222-231.

Nii-Trebi, N. I. (2017). Emerging and neglected infectious diseases: insights, advances, and challenges. *BioMed research international*, *2017*.

Pan, G., Jia, C., Zhao, D., You, C., Chen, H., & Jiang, G. (2009). Effect of cationic and anionic surfactants on the sorption and desorption of perfluorooctane sulfonate (PFOS) on natural sediments. *Environmental Pollution*, *157*(1), 325-330.

Parkin, D. M. (2006). The global health burden of infection-associated cancers in the year 2002. *International journal of cancer*, *118*(12), 3030-3044.

Parks, C. L. (2017). Replication-Competent Viral Vectors for Vaccine Delivery. In *Human Vaccines* (pp. 25-63). Academic Press.

Parks, C. L., Picker, L. J., & King, C. R. (2013). Development of replication-competent viral vectors for HIV vaccine delivery. *Current opinion in HIV and AIDS*, *8*(5), 402.

Parrish, C. R., Holmes, E. C., Morens, D. M., Park, E. C., Burke, D. S., Calisher, C. H., ... & Daszak, P. (2008). Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol. Mol. Biol. Rev.*, *72*(3), 457-470.

Pasquali, R. C., Taurozzi, M. P., & Bregni, C. (2008). Some considerations about the hydrophilic–lipophilic balance system. *International journal of pharmaceutics*, *356*(1-2), 44-51.

Patterson, J., Sammon, M., & Garg, M. (2016). Dengue, Zika and chikungunya: emerging arboviruses in the New World. *Western Journal of Emergency Medicine*, *17*(6), 671.

Pawlotsky, J. M. (2016). Hepatitis C virus resistance to direct-acting antiviral drugs in interferon-free regimens. *Gastroenterology*, *151*(1), 70-86.

Pfeiffer, J. K., & Kirkegaard, K. (2005). Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS pathogens*, *1*(2), e11.

Pike, L. J. (2009). The challenge of lipid rafts. *Journal of lipid research*, *50*(Supplement), S323-S328.

Podolak, I., Galanty, A., & Sobolewska, D. (2010). Saponins as cytotoxic agents: a review. *Phytochemistry Reviews*, *9*(3), 425-474.

Price, B. D., Roeder, M., & Ahlquist, P. (2000). DNA-directed expression of functional flock house virus RNA1 derivatives in Saccharomyces cerevisiae, heterologous gene expression, and selective effects on subgenomic mRNA synthesis. *Journal of virology*, *74*(24), 11724-11733.

Pringle, F. M., Johnson, K. N., Goodman, C. L., McIntosh, A. H., & Ball, L. A. (2003). Providence virus: a new member of the Tetraviridae that infects cultured insect cells. *Virology*, *306*(2), 359-370.

Priyamvada, L., Quicke, K. M., Hudson, W. H., Onlamoon, N., Sewatanon, J., Edupuganti, S., ... & Ahmed, R. (2016). Human antibody responses after dengue virus infection are highly cross-reactive to Zika virus. *Proceedings of the National Academy of Sciences*, *113*(28), 7852-7857.

Proteinatlas.org. (2019). *Cell atlas - MYC - The Human Protein Atlas*. [online] Available at: https://www.proteinatlas.org/ENSG00000136997-MYC/cell. (Accessed 15 Nov. 2019).

Randall, R. E., & Griffin, D. E. (2017). Within host RNA virus persistence: mechanisms and consequences. *Current opinion in virology*, *23*, 35-42.

Randall, R. E., & Goodbourn, S. (2008). Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures. *Journal of General Virology*, *89*(1), 1-47.

Rato, S., Golumbeanu, M., Telenti, A., & Ciuffi, A. (2017). Exploring viral infection using single-cell sequencing. *Virus research*, *239*, 55-68.

Rider, T. H., Zook, C. E., Boettcher, T. L., Wick, S. T., Pancoast, J. S., & Zusman, B. D. (2011). Broad-spectrum antiviral therapeutics. *PloS one*, *6*(7), e22572.

Ritchie, C., Cylinder, I., Platt, E. J., & Barklis, E. (2015). Analysis of HIV-1 Gag protein interactions via biotin ligase tagging. *Journal of virology*, *89*(7), 3988-4001.

Reuter, G., Várallyay, É., Baráth, D., Földvári, G., Szekeres, S., Boros, Á., ... & Pankovics, P. (2019). Analysis of a novel RNA virus in a wild northern white-breasted hedgehog (Erinaceus roumanicus). *Archives of virology*, 1-7.

Roizman, B., Zhou, G., & Du, T. (2011). Checkpoints in productive and latent infections with herpes simplex virus 1: conceptualization of the issues. *Journal of neurovirology*, *17*(6), 512-517.

Rothman, A. L., & Ennis, F. A. (2016). Dengue vaccine: the need, the challenges, and progress.

Roux, K. J., Kim, D. I., Raida, M., & Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol*, *196*(6), 801-810.

Roux, K., Kim, D., Burke, B. and May, D. (2018). BioID: A Screen for Protein-Protein Interactions. *Current Protocols in Protein Science*, pp.19.23.1-19.23.15.

Rowland, R. R., Kervin, R., Kuckleburg, C., Sperlich, A., & Benfield, D. A. (1999). The localization of porcine reproductive and respiratory syndrome virus nucleocapsid protein to the nucleolus of infected cells and identification of a potential nucleolar localization signal sequence. *Virus research*, *64*(1), 1-12.

Russo, M., Burgyan, J., & Martelli, G. P. (1994). Molecular biology of Tombusviridae. In *Advances in virus research* (Vol. 44, pp. 381-428). Academic Press.

Ryabov, E. V., Oparka, K. J., Santa Cruz, S., Robinson, D. J., & Taliansky, M. E. (1998). Intracellular location of two groundnut rosette umbravirus proteins delivered by PVX and TMV vectors. *Virology*, *242*(2), 303-313.

Sakin, V., Paci, G., Lemke, E. A., & Müller, B. (2016). Labeling of virus components for advanced, quantitative imaging analyses. *FEBS letters*, *590*(13), 1896-1914.

Salvetti, A., & Greco, A. (2014). Viruses and the nucleolus: the fatal attraction. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1842*(6), 840-847.

Scheer, U., Kartenbeck, J., Trendelenburg, M. F., Stadler, J., & Franke, W. W. (1976). Experimental disintegration of the nuclear envelope. Evidence for pore-connecting fibrils. *The Journal of cell biology*, *69*(1), 1-18.

Schneider-Poetsch, T., Ju, J., Eyler, D. E., Dang, Y., Bhat, S., Merrick, W. C., ... & Liu, J. O. (2010). Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nature chemical biology*, *6*(3), 209.

Schuh, A. J., Amman, B. R., & Towner, J. S. (2017). Filoviruses and bats. *Microbiology Australia*, *38*(1), 12-16.

Seddon, A. M., Curnow, P., & Booth, P. J. (2004). Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, *1666*(1-2), 105-117.

Seibel, N.M., Eljouni, J., Nalaskowski, M.M., Hampe, W. (2007). Nuclear localization of enhanced green fluorescent protein homomultimers. *Analytical Biochemistry 368*, 95–99.

Selling, B. H., Allison, R. F., & Kaesberg, P. (1990). Genomic RNA of an insect virus directs synthesis of infectious virions in plants. *Proceedings of the National Academy of Sciences*, *87*(1), 434-438.

Serva.de. (2019). *Classification of Detergents › Detergents › Information Center › SERVA Electrophoresis GmbH*. [online] Available at: https://www.serva.de/enDE/275_Information_Center_Detergents_Classification_of_Detergents.html. (Accessed 8 Nov. 2019).

Shan, C., Xie, X., Muruato, A. E., Rossi, S. L., Roundy, C. M., Azar, S. R., ... & Vasilakis, N. (2016). An infectious cDNA clone of Zika virus to study viral virulence, mosquito transmission, and antiviral inhibitors. *Cell host & microbe*, *19*(6), 891-900.

Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X., ... & Buchmann, J. (2016). Redefining the invertebrate RNA virosphere. *Nature*, *540*(7634), 539.

Short, J. R., & Dorrington, R. A. (2012). Membrane targeting of an alpha-like tetravirus replicase is directed by a region within the RNA-dependent RNA polymerase domain. *Journal of General Virology*, *93*(8), 1706-1716.

Short, J. R., Knox, C., & Dorrington, R. A. (2010). Subcellular localization and live-cell imaging of the Helicoverpa armigera stunt virus replicase in mammalian and Spodoptera frugiperda cells. *Journal of general virology*, *91*(6), 1514-1523.

Short, J. R., Nakayinga, R., Hughes, G. E., Walter, C. T., & Dorrington, R. A. (2013). Providence virus (family: Carmotetraviridae) replicates vRNA in association with the Golgi apparatus and secretory vesicles. *Journal of General Virology*, *94*(5), 1073-1078.

Short, J. R., Speir, J. A., Gopal, R., Pankratz, L. M., Lanman, J., & Schneemann, A. (2016). Role of mitochondrial membrane spherules in flock house virus replication. *Journal of virology*, *90*(7), 3676-3683.

Siddle, K. J., Eromon, P., Barnes, K. G., Mehta, S., Oguzie, J. U., Odia, I., ... & Wohl, S. (2018). Genomic analysis of Lassa virus during an increase in cases in Nigeria in 2018. *New England Journal of Medicine*, *379*(18), 1745-1753.

Sinha, S., Field, J. J., & Miller, J. H. (2017). Use of substitute Nonidet P-40 nonionic detergents in intracellular tubulin polymerization assays for screening of microtubule targeting agents. *Biochemistry and Cell Biology*, *95*(3), 379-384.

Singh, S., Farr, D., & Kumar, A. (2018). Ocular manifestations of emerging flaviviruses and the blood-retinal barrier. *Viruses*, *10*(10), 530.

Sklan, E. H., Charuworn, P., Pang, P. S., & Glenn, J. S. (2009). Mechanisms of HCV survival in the host. *Nature reviews Gastroenterology & hepatology*, *6*(4), 217.

Skuzeski, J. M., Nichols, L. M., Gesteland, R. F., & Atkins, J. F. (1991). The signal for a leaky UAG stop codon in several plant viruses includes the two downstream codons. *Journal of molecular biology*, *218*(2), 365-373.

Smirnova, E., Firth, A. E., Miller, W. A., Scheidecker, D., Brault, V., Reinbold, C., ... & Ziegler-Graff, V. (2015). Discovery of a small non-AUG-initiated ORF in poleroviruses and luteoviruses that is required for long-distance movement. *PLoS pathogens*, *11*(5), e1004868.

Suzuki, T., & Suzuki, Y. (2006). Current topics of lipid dynamics and pathobiology in membrane lipid rafts. *Biol Pharm Bull*, *29*(8), 1538-1541.

Swan, J. S. (2018). Identifying the binding partners of hCNK1 in the NF-κB pathway. Honours thesis, Rhodes University, South Africa.

Taliansky, M. E., & Robinson, D. J. (2003). Molecular biology of umbraviruses: phantom warriors. *Journal of General Virology*, *84*(8), 1951-1960.

Taliansky, M., Roberts, I. M., Kalinina, N., Ryabov, E. V., Raj, S. K., Robinson, D. J., & Oparka, K. J. (2003). An umbraviral protein, involved in long-distance RNA movement, binds viral RNA and forms unique, protective ribonucleoprotein complexes. *Journal of virology*, *77*(5), 3031-3040.

Tan, S. C., & Yiap, B. C. (2009). DNA, RNA, and protein extraction: the past and the present. *BioMed Research International*, *2009*.

Tang, L. S., Covert, E., Wilson, E., & Kottilil, S. (2018). Chronic hepatitis B infection: a review. *Jama*, *319*(17), 1802-1813.

Thavarajah, R., Mudimbaimannar, V. K., Elizabeth, J., Rao, U. K., & Ranganathan, K. (2012). Chemical and physical basics of routine formaldehyde fixation. *Journal of oral and maxillofacial pathology: JOMFP*, *16*(3), 400.

Tong, M., Hansen, A., Hanson-Easey, S., Cameron, S., Xiang, J., Liu, Q., ... & Bi, P. (2015). Infectious diseases, urbanization and climate change: challenges in future China. *International journal of environmental research and public health*, *12*(9), 11025-11036.

Tsai, C. W., Redinbaugh, M. G., Willie, K. J., Reed, S., Goodin, M., & Hogenhout, S. A. (2005). Complete genome sequence and in planta subcellular localization of maize fine streak virus proteins. *Journal of virology*, *79*(9), 5304-5314.

Uhlen, M., Bandrowski, A., Carr, S., Edwards, A., Ellenberg, J., Lundberg, E., ... & Yamamoto, T. (2016). A proposal for validation of antibodies. *Nature methods*, *13*(10), 823.

Uprichard, S. L. (2010). Hepatitis C virus experimental model systems and antiviral drug research. *Virologica Sinica*, *25*(4), 227-245.

van Gestel, R. A., Brouwers, J. F., Ultee, A., Helms, J. B., & Gadella, B. M. (2016). Ultrastructure and lipid composition of detergent-resistant membranes derived from mammalian sperm and two types of epithelial cells. *Cell and tissue research*, *363*(1), 129-145.

Verma, M., Erwin, S., Abedi, V., Hontecillas, R., Hoops, S., Leber, A., ... & Ciupe, S. M. (2017). Modeling the mechanisms by which HIV-associated immunosuppression influences HPV persistence at the oral mucosa. *PloS one*, *12*(1), e0168133.

Vilar, M., Saurı, A., Monné, M., Marcos, J. F., von Heijne, G., Pérez-Payá, E., & Mingarro, I. (2002). Insertion and topology of a plant viral movement protein in the endoplasmic reticulum membrane. *Journal of Biological Chemistry*, *277*(26), 23447-23452.

Walter, C.T. (2008). Establishing experimental systems for studying the replication biology of Providence virus. PhD thesis, Rhodes University, South Africa.

Walter, C. T., Pringle, F. M., Nakayinga, R., de Felipe, P., Ryan, M. D., Ball, L. A., & Dorrington, R. A. (2010). Genome organization and translation products of Providence virus: insight into a unique tetravirus. *Journal of General Virology*, *91*(11), 2826-2835.

Walters, C. E., Meslé, M. M., & Hall, I. M. (2018). Modelling the global spread of diseases: A review of current practice and capability. *Epidemics*, *25*, 1-8.

Wang, Z., & Schey, K. L. (2015). Proteomic analysis of lipid raft-like detergent-resistant membranes of lens fiber cells. *Investigative ophthalmology & visual science*, *56*(13), 8349-8360.

Wang, S. H., Syu, W. J., Huang, K. J., Lei, H. Y., Yao, C. W., King, C. C., & Hu, S. T. (2002). Intracellular localization and determination of a nuclear localization signal of the core protein of dengue virus. *Journal of General Virology*, *83*(12), 3093-3102.

Wang, A., Thurmond, S., Islas, L., Hui, K., & Hai, R. (2017). Zika virus genome biology and molecular pathogenesis. *Emerging microbes & infections*, *6*(1), 1-6.

White, K. A., & Nagy, P. D. (2004). Advances in the molecular biology of tombusviruses: gene expression, genome replication, and recombination. *Progress in nucleic acid research and molecular biology*, *78*, 187-226.

Whitmer, S. L., Ladner, J. T., Wiley, M. R., Patel, K., Dudas, G., Rambaut, A., ... & Knust, B. (2018). Active Ebola virus replication and heterogeneous evolutionary rates in EVD survivors. *Cell reports*, *22*(5), 1159-1168.

WHO, (2019). *Ebola health update - DRC, 2019*. [online] Available at: https://www.who.int/emergencies/diseases/ebola/drc-2019 (Accessed 25 Nov. 2019).

Woolhouse, M., Scott, F., Hudson, Z., Howey, R., & Chase-Topping, M. (2012). Human viruses: discovery and emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1604), 2864-2871.

Xie, X., Zou, J., Shan, C., Yang, Y., Kum, D. B., Dallmeier, K., ... & Shi, P. Y. (2016). Zika virus replicons for drug discovery. *EBioMedicine*, *12*, 156-160.

Yang, X. L., Tan, C. W., Anderson, D. E., Jiang, R. D., Li, B., Zhang, W., ... & Guan, W. (2019). Characterization of a filovirus (Měnglà virus) from Rousettus bats in China. *Nature microbiology*, *4*(3), 390.

Yau, W. L., Nguyen-Dinh, V., Larsson, E., Lindqvist, R., Överby, A. K., & Lundmark, R. (2019). Model System for the Formation of Tick-Borne Encephalitis Virus Replication Compartments without Viral RNA Replication. *Journal of Virology*, JVI-00292.

Yu, C. S., Cheng, C. W., Su, W. C., Chang, K. C., Huang, S. W., Hwang, J. K., & Lu, C. H. (2014). CELLO2GO: a web server for protein subCELlular LOcalization prediction with functional gene ontology annotation. *PloS one*, *9*(6), e99368.

Zumla, A., Rao, M., Wallis, R. S., Kaufmann, S. H., Rustomjee, R., Mwaba, P., ... & Azhar, E. (2016). Host-directed therapies for infectious diseases: current status, recent progress, and future prospects. *The Lancet Infectious Diseases*, *16*(4), e47-e63.

Zuniga, E. I., Macal, M., Lewis, G. M., & Harker, J. A. (2015). Innate and adaptive immune regulation during chronic viral infections. *Annual review of virology*, *2*, 573-597.