Gene Expression Regulation in Pneumoviruses

By William Collier

Thesis submitted for the degree of Doctor of Philosophy at the University of Warwick

Research conducted in the School of Life Sciences, University of Warwick

Submitted
September 2017

Figures

Tables

Acknowledgements

I would like to thank my supervisor Professor Andrew Easton for the help and support he has provided over the last 4 years. Your door has always been open and you were always willing to answer any question no matter how small. You have supported me through every stage of my PhD and I could not ask for a better supervisor. I would also like to thank Dr Phillip Gould for helping me through my PhD and always being available to answer any question. Your unique style of motivation and humour certainly helped me through the harder points during these 4 years and always kept our lab environment a fun place to be in. I would also like to thank past and present members of the Virology group at the University of Warwick, who have always been helpful and were always around for a chat and a beer.

Lastly I would like to thank my friends and family who have supported me both emotionally and financially and who have made these last 4 years possible.

Declaration

I hereby declare that all the work presented in this thesis is my own. As discussed in Chapter 3, raw unprocessed sequencing data acquired from Dr Phillip Gould was processed and analysed on my own. The data had not been altered or edited and was in raw format direct from sequencing. Unless otherwise stated in this thesis, no work has been submitted for a degree at any other institution.

William Collier

Abbreviations

aa: Amino acids

AMP: Ampicillin

APV: Avian metapneumovirus

Asp: Aspartate

BCA: Bicinchoninic acid assay

bp: Base pairs

bRSV: Bovine respiratory syncytial virus

BSA: Bovine serum albumin

CAT: Chloramphenicol acetyl transferase

CDS: Coding sequence

CHV1: Cryphonectria hypovirus 1

CrPV: Cricket paralysis virus

eEF: Eukaryotic elongation factor

eGFP: Enhanced green fluorescent protein

eIF: Eukaryotic initiation factor

EMCV Encephalomyocarditis virus

ELISA: Enzyme-linked immunosorbent assay

eRF: Eukaryotic release factor

F: Fusion

FBS: Foetal bovine serum

FCV: Feline calicivirus

FMDV: Foot and mouth and disease virus

G: Glycoprotein

GE: Gene stop

GFP: Green fluorescent protein

GMEM: Glasgow Modification of Eagle's Medium

GS: Gene start

h26: Helix 26

HCV: Hepatitis C virus

HIV: Human immunodeficiency virus

HMPV: Human metapneumovirus

hpi: Hours post infection

HRP: Horse radish peroxidase

hRSV: Human respiratory syncytial virus

IBV: Infectious bronchitis virus

ICAM: Intracellular adhesion molecules

IPTG: Isopropyl β-D-1-thiogalactopyranoside

IRES: Internal ribosome entry site

L: Large

LB: L-broth

M: Matrix

m^7G cap: 7-methylguanylate cap

MCP-1: Monocyte chemoattractant protein 1

MIP-1α: Macrophage inflammatory protein 1α

MIP-2: Macrophage inflammatory protein 2

MOI: Multiplicity of infection

mRNA: Messenger ribonucleic acid

N: Nucleoprotein

NS1: Non structural protein 1

NS2: Non structural protein 2

nts: Nucleotides

ORF: Open reading frame

P: Phosphoprotein

PAGE: Polyacrylamide gel electrophoresis

PBS: Phosphate buffered saline

Pfu: Plaque forming units

PRF: Programmed ribosomal frameshifting

PTB: Polypyrimidine-tract binding protein

PVM: Pneumonia virus of mice

QC: Quality control

RBS: Ribosome binding site

RdRP: RNA dependent RNA polymerase

RHDV: Rabbit haemorrhagic disease virus

RNP: Ribonucleoprotein

RPF: Ribosome protected fragments

rpm: rotations per minute

rRNA: ribosomal ribonucleic acid

RT: Reverse transcription

RTBV: Rice tungro bacilliform virus

SD: Shine-Dalgarno

sg: subgenomic

SH: Small Hydrophobic

SHAPE: Selective 2´-hydroxyl acylation and primer extension

sORF: Short ORF

STD: Standard deviation

trp: Tryptophan

TURBS: Terminating upstream ribosomal binding site

UTR: Untranslated region

VSV: Vesicular stomatitis virus

v/v: Volume to volume

w/v: Weight to volume

Summary

Members of the *Pneumoviridae* virus family are responsible for severe respiratory tract disease in their hosts. Human respiratory syncytial virus (hRSV) is responsible for over 200,000 deaths worldwide each year and bovine respiratory syncytial virus (bRSV) causes major economic loss to the cattle industry worldwide. The current model for all non-segmented negative-sense single stranded RNA virus gene expression, is that mRNA is generated in a polar gradient, with decreasing levels of mRNA transcribed from genes further along the genome from the 3´ end. With the exception of translation of ORF-2 located on the bicistronic M2 mRNA, translation of *Pneumoviridae* mRNAs is thought to be regulated through the levels of mRNA abundance. Translation of M2 ORF-2 has been characterised as being regulated by the non-canonical mechanism of coupled translation termination/initiation in pneumonia virus of mice (PVM), hRSV and avian metapneumovirus (APV). This mechanism is reliant on a proportion of the elongating ribosome translating the upstream M2 ORF-1, terminating and reinitiating translation of M2 ORF-2. Although the initiation site for M2 ORF-2 is similar in bRSV to other members of this family that use the mechanism of coupled translation, the mechanism has not been characterised.

Using the technique of ribosomal profiling to analyse steady state viral mRNA abundance and viral translation in both hRSV and bRSV-infected cells, it was observed that for certain viral mRNAs, levels of mRNA abundance did not follow the standard polar transcription model. This was characterised by an increase in the levels of mRNA abundance between the mRNA's respective gene and its upstream neighbour. The increase was observed in the same group of mRNAs in both viruses suggesting that factors other than the transcription polar gradient influence levels of viral mRNA abundance. It was also observed that levels of proportional translation did not match the respective proportional levels of mRNA abundance for certain viral mRNAs in both viruses. This would suggest that translation of viral genomes is not primarily controlled by mRNA abundance and instead other translational regulatory factors influence levels of translation.

The mechanism of bRSV M2 ORF-2 translation was also characterised using reporter plasmids assays. It was identified that the mechanism of initiation of translation of M2 ORF-2 used, was not that of coupled translation termination/initiation used by other members of this family. Instead it was observed that translation of M2 ORF-2 used an internal initiation mechanism located inside M2 ORF-1 to initiate translation.

The mechanism of coupled translation termination/initiation used for translation of PVM M2 ORF-2 was also further characterised. It was observed that translation of M2 ORF-2 was reliant on upstream sequence in the M2 ORF-1 sequence. A predicted mRNA secondary structure was identified in this region and when disrupted, inhibited translation of M2 ORF-2. This was similar to the mechanism of coupled translation used in hRSV, suggesting that the mechanism used by this family is reliant on a mRNA secondary structure located upstream of the initiation site.

Chapter 1
Introduction

1.1.1. The *Pneumoviridae* family

Respiratory diseases have been a major cause of mortality among both human and animal populations throughout history. One of the main agents behind these diseases are viruses. Of these viruses, the *Pneumoviridae* family is a major cause of respiratory diseases in a large range of animals: which include birds, rodents, livestock and humans. These viruses tend to have a high morbidity and in certain situations a high mortality.

The *Pneumoviridae* family is part of the order *Mononegavirales*. This order comprises of a non-segmented single stranded negative-sense RNA genome viruses which consist of 10 genes. The *Pneumoviridae* family can be further divided into the *Orthopneumovirus* and *Metapneumovirus* genera (Alvarez et al., 2003; 2016; McCarthy and Goodman, 2010).

As seen in table 1, the genus *Metapneumovirus* contains two species of virus and the *Orthopneumovirus* genus contains three. All of these viruses cause acute respiratory illness in their host species. This thesis will focus on the *Orthopneumovirus* genus.

Orthopneumovirus	*Metapneumovirus*
Human respiratory syncytial virus	Avian metapneumovirus
Bovine respiratory syncytial virus	Human metapneumovirus
Pneumonia virus of mice	

Table 1: **Members of the *Pneumoviridae family***
Species of virus that are found in the two genera of the *Pneumoviridae* family. List according to the International Committee on Taxonomy of Viruses (ICTV 2016)

1.1.2. Human Respiratory Syncytial Virus

Human respiratory syncytial virus (hRSV) was first isolated in chimpanzees in the 1950s (Blount et al., 1956). The virus can be separated into two subgroups (A and B) based on antigenic differences (Fletcher et al., 1997). Both subgroups are responsible for causing lower respiratory tract diseases, such as bronchiolitis and pneumonia in humans, and are capable of causing disease in all age groups. It is estimated that 160,000 people a year die worldwide from hRSV related diseases (Cowton et al., 2006) -of those 60,000 are children (Nair et al., 2010). By the age of three, it is estimated that almost all children have been infected with hRSV (Simoes, 1999). The

symptoms caused by hRSV tend to be mild with symptoms such as cough and rhinitis/coryza. In general, children, the elderly and the immunodeficient have a higher probability of the infection becoming more severe (McNamara and Smyth, 2002). Most infections tend to show mild upper respiratory symptoms. An infection however may become more severe and progress to the lower respiratory tract causing the associated diseases mentioned above. Cases of mortality tend to occur in infants with other mitigating factors such as immunodeficiency, premature birth and cystic fibrosis (Borchers et al., 2013; Cowton et al., 2006; Fleming et al., 2005; Guvenel et al., 2014; Lee et al., 2012; McNamara and Smyth, 2002).

hRSV infects epithelial cells that are found in both the respiratory and nasal tracts. There are generally no major cytopathological effects caused by hRSV during infection, with the virus tending to supress and delay apoptosis in the cell. In severe infection, the virus causes an increase in fluid extravasation into the air spaces found in the lung. Animal models in mice have also shown that the virus decreases sodium transport across epithelial cells leading to a decrease in fluid clearance in the lungs. Severe infection can also lead to a loss in plasma membrane integrity in bronchial cells (Oshansky et al., 2009).

Currently, there is no vaccine for hRSV. Initial attempts at producing an inactivated viral vaccine ended disastrously in the 1960s. The inactivated vaccine used increased the severity of viral infections once a child was immunized leading to higher rates of hospitalization and mortality among the recipients. This was thought to be due to the vaccine increasing the activity of the T-helper 2 cells, which lead to pulmonary eosinophilia (Borchers et al., 2013). As a result over the next few decades, a sharp decrease in research into a suitable vaccine was seen. Recently there has been an increased level of research into an alternative vaccine. These include using both attenuated and recombinant antigen methods to produce a viable vaccine. Research into an attenuated vaccine has focussed on using hRSV mutants, where individual mutated sites and deletions target the Non Structural 2 (Section 1.2.2), Small hydrophobic (Section 1.2.6) and M2 (Section 1.2.10) viral genes (Guvenel et al., 2014). This combination of knockouts has produced an attenuated vaccine that has reached clinical trials. There has also been initial success in the development of recombinant adjuvant vaccines. All surface proteins (fusion (Section 1.2.8),

glycoprotein (Section 1.2.7) and small hydrophobic protein (Section 1.2.6)) have been used to produce antigens which can induce immunity against hRSV (Borchers et al., 2013; Guvenel et al., 2014; Rudraraju et al., 2013).

One of the most effective treatments for hRSV is the use of the humanized monoclonal antibody Palivizumab (Medimmune). This antibody binds to the fusion glycoprotein that protrudes externally from the viral membrane. The antibody is effective at reducing the viral titre, however this does not in turn completely halt the viral infection and monthly doses must be given as a result in order to stop viral dissemination. As a result of this and due to the high cost (~$14K per dose), this treatment is mainly used as a prophylaxis for young children and only given to children that are at high risk of developing a severe infection from the virus. Ribavirin (a nucleoside analogue) is currently the only antiviral molecule that is allowed for human use. However, due to the fear of side effects and the cost to benefit ratio, the drug is not commonly used. There are currently other drugs that are being tested that have not been approved for human use that target viral pathways. This includes research into new prophylaxis treatment for neonates, that provides protection against several viruses in the *Pneumoviridae* family that infect humans (Corti et al., 2013; Gomez et al., 2014).

1.1.3. Bovine Respiratory Syncytial Virus

Bovine respiratory syncytial virus (bRSV) like all members of the *Orthopneumovirus* genus (McCarthy and Goodman, 2010) has the same 10 genes that code for the same 11 proteins. bRSV causes both upper and lower respiratory infections in cattle, with calves under 6 months being most vulnerable. The symptoms caused by this virus can range from dyspnea, tachypnea, ocular secretions, dry muzzle, reduced activity and anorexia. This in turn leads to a loss of activity, which has a huge economic cost. Death from the virus is also not uncommon among cases. The viral infection also makes cattle more susceptible to bacterial infections such as *Mycoplasma bovis* which is the leading cause of cattle mortality in the US (Sacco et al., 2014).

As in hRSV infection, the virus infects epithelial cells found in the respiratory tract. Many of the viral symptoms are caused by the immune response. Factors released from the innate immune response can lead to fever and anorexia due to the release of

proinflammatory cytokines. Inflammation in the lung can also be linked to a T-helper 2 response. Both the innate and adaptive immune responses are linked to the cause of the more severe symptoms in bRSV (Gershwin, 2012).

There are multiple vaccines that are used for bRSV. These usually include a mixture of both inactivated and live attenuated virus. As with the inactivated vaccine for hRSV, some inactivated vaccines for bRSV have also shown to cause an increase in severity of the disease caused by bRSV (Sacco et al., 2014). Current vaccines that are used in the market today are far from perfect as their efficacy is far from ideal. New research has focussed on the use of adjuvant and attenuated vaccines, which have shown better efficacy and also a decrease in vaccine-enhanced disease. Currently there is no effective treatment for bRSV in cattle (Blodörn et al., 2014; Sacco et al., 2014).

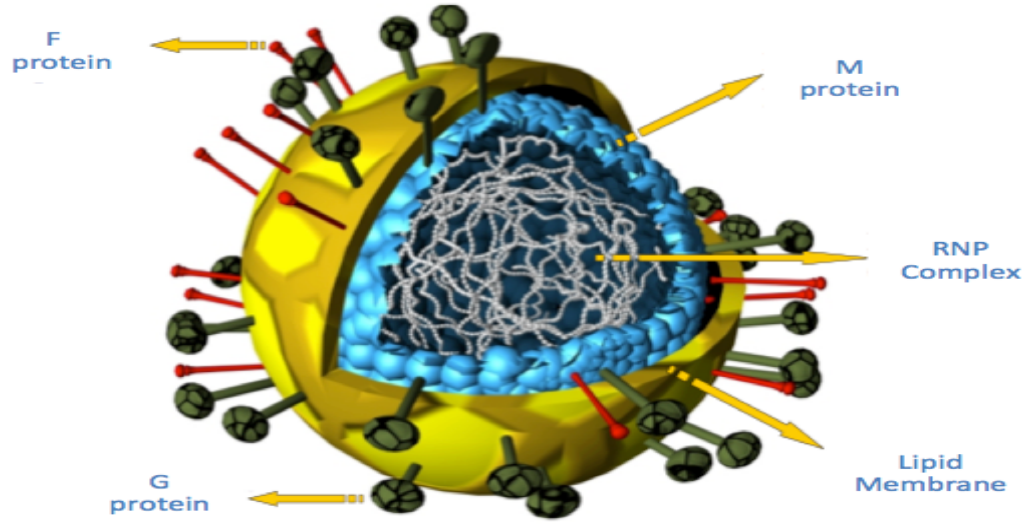
1.1.4. Pneumonia Virus of Mice

Pneumonia virus of mice (PVM) is responsible for causing respiratory disease in mice, with symptoms similar to that of severe disease in hRSV. Death of mice tends to follow quickly after infection. In neonatal mice, there is virtually no inflammation during infection (Rosenberg and Domachowske, 2008). Current research suggests that the virus is not a pathogen to other animals including humans, as the levels of replication are so low. Although there is evidence of neutralizing antibodies found in humans during exposure, it is suggested that this is due to polyreactive neutral antibodies. The virus has been used extensively as a model for the inflammatory response of all viruses in the *Orthopneumovirus* genus. PVM causes a similar inflammatory response compared to that in severe disease in hRSV. Both viral infections cause the release of macrophage inflammatory protein 1α and 2 (MIP-1α and MIP-2) and monocyte chemoattractant protein 1 (MCP-1). The release of these proteins leads to the stimulation and activation of the innate immune cells. During infection some of the viral proteins including NS1 and NS2 are responsible for inhibiting the cellular interferon response to the viral infection (Brock et al., 2012; Dyer et al., 2012; Rosenberg and Domachowske, 2008).

1.2.1. Genome of the *Orthopneumovirus* genus

All three members that belong to this genus have a single negative-sense RNA genome that is ~15kb long. The genome comprises of 10 genes from which 11 proteins are translated. As seen in Fig. 1.1B, all three members of the genus have the same 10 genes in the genome arranged in the same sequential order (Buchholz et al., 1999; Collins et al., 1987; Krempl et al., 2005). The genome is also flanked by a 3′ leader sequence and a 5′ trailer sequence. In hRSV the leader sequence has been reported to be 44 nts in size and the trailer sequence to be 144 nts in size (Mink et al., 1991). In this genus, the largest amount of research has been performed on hRSV. Therefore, when reviewing the genome and its replication cycle the review will mainly focus on hRSV.

A



B

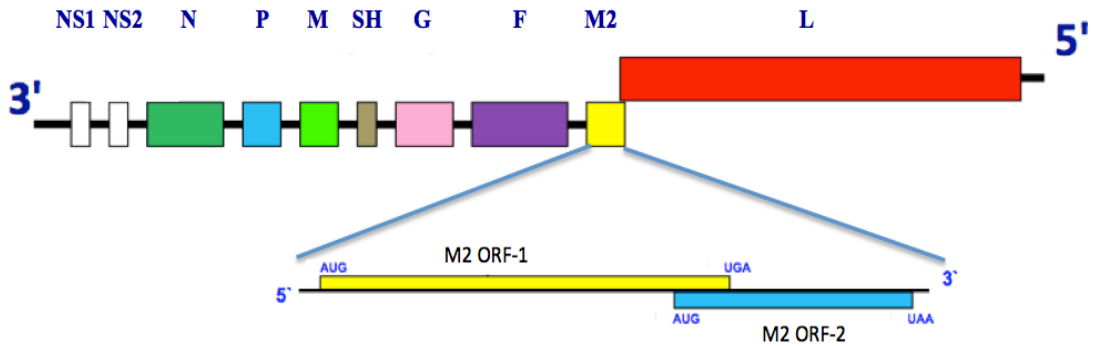


Fig. 1.1: Virion and genome of members of *Orthopneumovirus* family
(A) Virion of members of the *Orthopneumovirus* genus. The ribonucleoprotein (RNP) complex consists of the viral genome encapsidated by the N protein, P protein and L protein. (B) Genome of hRSV including M2 gene transcript transcribed from the M2 gene. Diagram includes both AUG and stop codons of both open reading frames (ORFs) in the M2 gene transcript.

1.2.2. Non structural protein 1 (NS1) and Non structural protein 2 (NS2)

The first two genes from the 3′ end of the *Orthopneumovirus* genome are the non structural 1 NS1 and non structural protein (NS2) genes. These two genes are translated into 139 and 124 amino acid proteins respectively (Johnson and Collins, 1989). The two proteins are not packaged into infectious virion particles and are only present during infection. They are not involved in the replication of the virus, but are however involved in suppressing the host cell response to the virus. The two proteins are involved in blocking type I and type II interferon signalling pathways (Hastie et al., 2012; Lo et al., 2005; Spann et al., 2004). These proteins are also involved in preventing premature apoptosis of the infected cell (Bitko et al., 2007). There is also evidence of involvement of NS1 in suppressing regulatory T cells. These cells are thought to be involved in maintaining self tolerance for the immune system (Yang et al., 2015).

1.2.3. Nucleoprotein (N)

The nucleoprotein (N) encapsidates the RNA genome in the packaged viral particle and encapsidates both genome and antigenome inside the cell during infection. Encapsidation of the genome forms the nucleocapsid, which protects the viral genome during infection from cellular factors such as RNAses. The nucleocapsid also comprises of P and L proteins with the M2-1 protein also associated, which forms the RNP complex. The N protein forms the nucleocapsid into a helix-like structure with variable numbers of N proteins encapsidating the viral RNA in each turn. The N protein also allows for transcription and replication by the L protein by allowing the L protein to bind through interactions with the P protein (Bakker et al., 2013; El Omari et al., 2011; Maclellan et al., 2007).

1.2.4. Phosphoprotein (P)

The phosphoprotein (P) is a 241 amino acid protein that is also associated with the RNP complex. The protein is phosphorylated at multiple positions along the length (Lu et al., 2002) and also forms homotetramer complexes with other P proteins to form a modular structure (Llorente et al., 2006). The protein is known to associate with the N, L and M2-1 proteins (Galloux et al., 2012; Mason et al., 2003). The P protein is essential for successful encapsidation of the viral RNA. It also has been

suggested that the P protein acts as chaperone for the N protein when not associated with the viral genome. This is to prevent the N protein from binding to other cellular RNAs. The protein is an essential co-factor for the L protein in its function as a RNA dependent RNA polymerase (RdRP) (Galloux et al., 2012). This is achieved by specific interactions between the P and N proteins (Lu et al., 2002). The P protein is also known to regulate the M2-2 protein in its function by facilitating the viral polymerase in its switch from transcription to replication (Asenjo and Villanueva, 2016).

1.2.5. Matrix protein (M)

The matrix protein (M) is a non-glycosylated structural protein that is 256 amino acids in size, which is involved in the assembly of the viral complex (Ghildyal et al., 2006). In an infectious virion, the M protein is found as a partial layer between the viral membrane and the ribonucleoprotein (Liljeroos et al., 2013). Its role in assembly has not been well characterized and is thought to aid in preventing L polymerase activity during assembly. The protein forms a bridge between the new viral envelope and the ribonucleoprotein (Ghildyal et al., 2006). It is also a key factor in the formation of viral filaments formed during infection (Mitra et al., 2012).

1.2.6. Small Hydrophobic protein (SH)

The small hydrophobic protein (SH) is a type II membrane protein that is 64 amino acids in size. During infection, there are up to 4 structurally different isomers of the protein expressed (Collins and Mottet, 1993). It is currently not known what fraction of each specific isomer of this protein is expressed during infection. It has been suggested that the protein acts to aid the virus in immune evasion and to aid the virus in virus-mediated cell fusion. It also has been reported that the protein associates with the Golgi apparatus and the plasma membrane (Rixon et al., 2005). The structure has been solved when binding to lipid bilayers with the protein forming a pentameric oligomer (Collins and Mottet, 1993). It also has been found to associate with proteins in antiapoptotic pathways (Li et al., 2015b).

1.2.7. Glycoprotein (G)

The glycoprotein (G) is a type II glycosylated surface membrane protein and varies between 292 and 299 amino acids in size (Teng and Collins, 2002). The G protein

varies highly between strains of virus as it is one of the two viral proteins that induce neutralizing antibodies. The G protein is thought to enable viral-mediated cell fusion by initially binding to glycosaminoglycans and negatively charged carbohydrates. This allows the fusion protein to bind to cellular receptors thereby facilitating virus-mediated cell fusion (Kwilas et al., 2009; Techaarpornkul et al., 2002). The G protein is also released from the cell during the replication cycle where it is thought to act as an immunomodulatory factor (Hendricks et al., 1987).

1.2.8. Fusion protein (F)

The fusion protein (F) is a type I surface membrane protein. The protein is 574 amino acids in size and enables the viral particle to fuse its membrane to the cellular membrane. Upon coming into close proximity to the cell membrane, the fusion protein will insert a number of hydrophobic residues into the cellular membrane. By a process of refolding, the fusion protein brings the cell and viral membranes into close proximity allowing fusion of the membranes. This forms a fusion pore allowing entry of the nucleocapsid into the cytoplasm of the cell. The F protein does not play a significant role in replication inside the cell (González-Reyes et al., 2001; Krzyzaniak et al., 2013; Low et al., 2008; McLellan et al., 2013b; Melero and Mas, 2015; Tayyari et al., 2011).

1.2.9. M2-1 protein

The M2-1 protein is translated from the first open reading frame (ORF) located on the M2 gene transcript. The role of the 194 amino acid protein has been primarily investigated through the use of minigenomes. These studies have elucidated that the M2-1 protein acts as an antitermination factor during transcription of the viral genome (Collins et al., 1996; Fearns and Collins, 1999b; Hardy and Wertz, 1998). The protein prevents termination of the transcribing viral polymerase at both intergenic regions (Collins et al., 1996), as well as at gene stop (GE) sites in the genome (Fearns and Collins, 1999b; Hardy and Wertz, 1998). It has also been found that transcription of longer mRNAs, as well as polycistronic read-through mRNAs are highly dependent on the presence of M2-1 protein. For the transcription of the viral genome, the protein is especially critical in the transcription of gene transcripts past the NS1 and NS2 genes of hRSV. The M2-1 protein however, does not have any specific role in the replication of the viral genome (Fearns and Collins, 1999b). The

M2-1 protein associates itself with the RNP complex during infection, with reported interactions with the N and P proteins. It has also been reported to interact with the M protein although the precise role in the interaction is still unclear (García et al., 1993).

1.2.10. M2-2 protein

The M2-2 protein is translated from the second ORF located on the M2 gene transcript. The role of the 83 amino acid protein is in the switch from transcription to replication of the viral genome. Although the mechanism for this role is unknown, knockout mutants of the M2-2 protein have shown a significant reduction in the levels of genomic viral RNA and also a significant decrease in the levels of viral yield during infection, with as much as a 1000-fold reduction observed. It is unknown whether the switch is due to the M2-2 protein directing synthesis of antigenomic or genomic RNA. However, it has been found that the protein accumulates during hRSV infection and peaks after 12-15 hours post infection and continues to remain constant after this time (Bermingham and Collins, 1999).

1.2.11. Large protein (L)

The large protein (L) is a 2165 amino acid protein and is the final gene in the viral genome (Collins et al., 2013). The L protein acts as an RNA dependent RNA polymerase (RdRP) for the virus and is associated with the RNP complex in the cell and in the virion. The protein during infection will initially individually transcribe viral genes in the viral genome (Noton et al., 2012). The L protein will also cap and polyadenylate the viral mRNA during transcription (Barik, 1993; Liuzzi et al., 2005). During later stages of infection, the L protein switches from transcription to replication by transcribing an antigenome. This acts a template for the L protein to produce multiple copies of the negative-sense genome (Morin et al., 2013). As described in Section 1.2.4, the L protein is reliant on the P protein, which acts as a critical cofactor by interacting with N protein thereby bringing the L protein into contact with the genome (Lu et al., 2002). The N protein must encapsidate both the genome and antigenome throughout the entire replication cycle, therefore the P protein is essential as a cofactor for the L protein for both transcription and replication (Morin et al., 2013). During transcription the local N protein

encapsidating the region that is transcribed must be removed in order for the L protein to access the genome.

1.3.1. Replication cycle

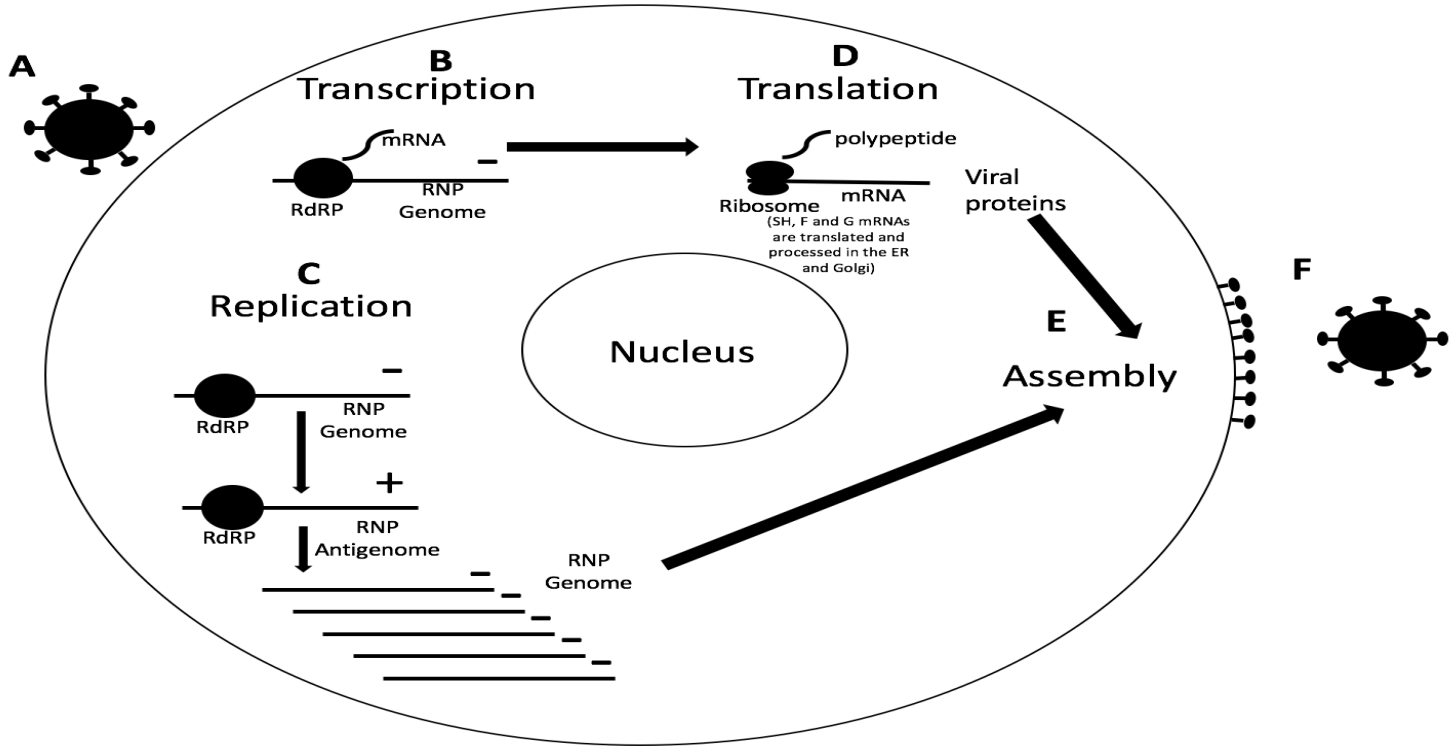


Fig. 1.2 Life cycle of viruses in the *Orthopneumovirus* genus
Life cycle of viruses in the *Othropneumovirus* genus. Transcription, replication, translation and packaging of the viral genome occur in the cytoplasm of the cell. (A) Viral entry. (B) Transcription of viral mRNAs by the RdRP. (C) Replication of the viral genome through the synthesis of an antigenome by the RdRP. The RdRP will use the antigenome as a template to synthesize multiple copies of the genome. (D) Translation of viral mRNAs. (E) Assembly of viral proteins and genome. (F) Budding and release.

1.3.1.1 Host cell entry

For *Orthopneumoviruses* host cell entry (Fig. 1.2A), the G protein initially binds to epithelium cells in order to bring the virus into contact with the cellular membrane. It is generally accepted that the G protein binds to negatively charged carbohydrates such as glycosaminoglycans (Kwilas et al., 2009; Techaarpornkul et al., 2002). There are several reported glycosaminoglycans such as heparan sulphate and chondroitin sulphate that have been reported to bind to the G protein. However with certain examples such as a herapan sulphate, the protein is not present on human airway epithelium cells (Krzyzaniak et al., 2013). It has also been shown that *in vitro*, hRSV can enter the cell without the use of the G protein and also without binding to any glycosaminoglycans using only the F protein (Techaarpornkul et al., 2001).

Upon successful binding of the virus, the F protein is responsible for cell-mediated viral fusion of the viral membrane to the cellular plasma membrane. In the viral membrane the F protein forms a trimeric complex linked by disulphide bonds. This trimeric complex is formed in a conformation called a prefusion complex. When the F protein comes into to close proximity to the membrane, the protein will then insert a stretch of hydrophobic residues into the membrane. This results in a conformational change to form the post fusion complex, which brings the two membranes into close proximity with each other. It is currently thought that in hRSV, there is a trigger that causes the F protein to insert its hydrophobic residues into the membrane (Mastrangelo and Hegele, 2013; McLellan et al., 2013a; McLellan et al., 2013b; Melero and Mas, 2015). It has been suggested that the virus can also bind to intracellular adhesion molecules (ICAM). However, there is increasing evidence that the cellular protein nucleolin acts as a trigger (Behera et al., 2001; Tayyari et al., 2011; Techaarpornkul et al., 2001). The virus uses an endocytotic method called macropinocytosis to insert its genetic material in the form of the RNP into the cell (Krzyzaniak et al., 2013). By bringing the two membranes together, the membranes will then fuse and form a fusion pore that allows the virus to insert its RNP into the cytoplasm of the cell. The exact method of entry into the cell however, is still debated (Melero and Mas, 2015).

1.3.2. Transcription of the *Orthopneumovirus* genome

Once the RNP of the virus is inside the cell, transcription of the genome commences within the cytoplasm in inclusion bodies (Fig. 1.2B) (García et al., 1993). Initiation of transcription occurs at the promoter site in the 44nt leader region which flanks the 3′ end of the viral genome. The site of the promoter at the 3′ end has been suggested to be either at position +1 or +3 in the leader sequence. There are two suggested models on how the L protein initiates transcription at the first gene start (GS) signal of NS1 based on the location of the promoter at either one of the two positions. Once bound to either position +1 or +3, the RdRP will begin to transcribe the leader sequence. In both of these models, there is a checkpoint at position +25 in the leader sequence which causes the termination of transcription of the leader mRNA and dissociation of the nascent mRNA from the polymerase. This allows the RdRP to initiate transcription at the gene start signal for NS1 (Tremaglio et al., 2013). There is also evidence that for successful elongation of transcription, the transcribed mRNA must be guanylated by the polymerase. This is thought to occur when the newly transcribed mRNA is 45 to 50 nucleotides long. Inhibitors used to target this function of polymerase resulted in no full length mRNAs transcribed and instead shorter truncated mRNAs were created that were not guanylated suggesting the importance of guanylation for successful transcription (Liuzzi et al., 2005).

As seen in Fig. 1.3, the L protein transcribes individual mRNAs from each gene beginning at each gene start signal, with synthesis continuing until a gene stop signal is reached. The nascent mRNA is capped during transcription and is polyadenylated and released once the polymerase reaches the gene stop signal. The efficiency of polyadenylation and termination is not 100% efficient. Instead it has been reported that read-through can occur between two genes, with as many as 10% of viral mRNAs transcribed are polycistronic mRNAs caused by readthrough (Tran et al., 2004). For the terminating polymerase reaching the gene stop signal, there are one of two options, the polymerase can either dissociate from the genomic RNA or begin synthesis of the mRNA for the next gene following recognition of the gene start sequence. The dissociated polymerase in order to reinitiate transcription, must bind to the promoter located at the leader sequence found at the 3′ end of the genome or can reinitiate directly at the first gene start signal for the genome at NS1 (Cattaneo et

al., 1987; Collins and Wertz, 1983; Krempl et al., 2002; Plumet et al., 2005; Tran et al., 2004; Tremaglio et al., 2013). This process repeats itself after the termination of transcription at each gene stop signal thereby forming a transcription polar gradient. Larger quantities of mRNAs are therefore synthesized for genes found closer to the 3′ end. This method allows the virus to control the amounts of proteins produced (Cowton et al., 2006). Although transcription can occur *in vitro* without any co-factors, both actin and profilin have been found to act as cellular co-factors for viral transcription (Burke et al., 1998; Burke et al., 2000).

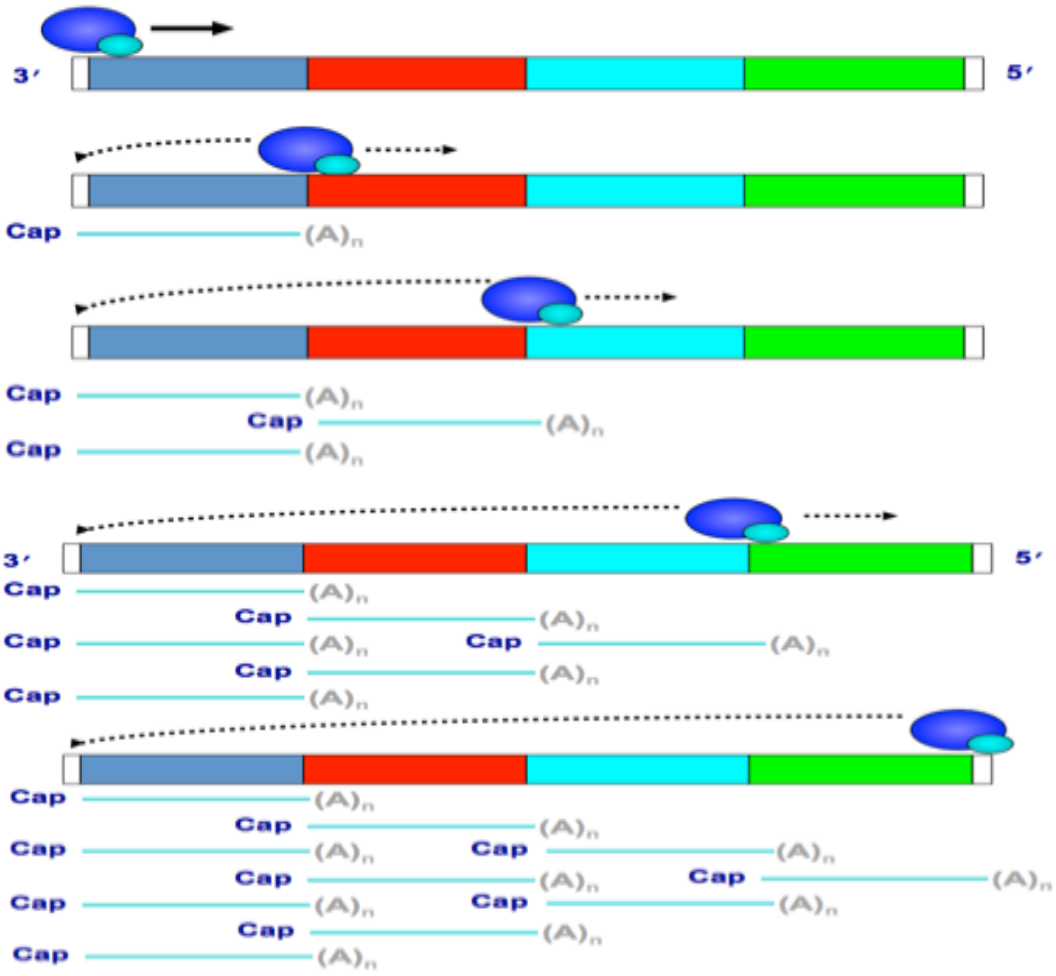


Fig. 1.3: Polar gradient transcription model
Schematic of the polar gradient transcription model. Diagram illustrates possible options for the terminating viral polymerase at each gene stop signal along the genome. The polymerase can either disassociate from the genomic RNA strand and reinitiate transcription at the promoter at 3′ end of the genome or begin transcription of the next gene through the next gene start signal. As a result, a gradient of mRNAs are generated with a larger quantity transcribed for genes closer to the promoter. Genes in the schematic are coloured separately.

Originally evidence for a transcription polar gradient was reported in other *Mononegavirales*. The first example of transcription initiation at the 3′ end of the genome was reported in vesicular stomatitis virus (VSV) using UV irradiation studies (Abraham and Banerjee, 1976). Further evidence of a transcription polar gradient in VSV was reported in Hodges et al., 2012. Before the reclassification of *Orthopneumoviruses* from the *Paramyxoviridae* family, a transcription gradient was reported using UV irradiation in Sendai virus. Similar to VSV, it was reported there was only a single promoter for the polymerase located at the 3′ end of the genome (Glazier et al., 1977). Further evidence for this polar gradient came from the measles virus, which is also a member of the *Paramyxoviridae* family. Contrary to the previous two studies, northern blot analysis was used to quantify the levels of mRNA produced for measles virus (Cattaneo et al., 1987). For hRSV, partial transcription maps have been generated through the use of hybridisation probes (Collins and Wertz, 1983) and minigneomes (Fearns and Collins, 1999b). These provided evidence of the polar gradient model of transcription. However in both examples, the transcription maps were created using only partial number of genes at the 3′ end of the genome (Fearns and Collins, 1999b; Collins and Wertz, 1983). Further evidence for a polar gradient also exists due to a single promoter being used by viruses in this family (Liuzzi et al., 2005).

There is however growing evidence that suggests that the polar gradient model of transcription is not accurate and the gene order of the genome is the not the sole factor controlling the levels of viral mRNA abundance. Transcriptome sequencing of cells infected with hRSV has revealed that certain genes in the genome have higher levels of mRNA abundance than ones further upstream. These data contradicted the polar gradient model, where the level of mRNA abundance is dependent on the position of the gene from the promoter. Contrary to the model, it was reported that the G mRNA had the highest levels of mRNA abundance in the hRSV genome. This gene is located in the centre of the genome and if the polar gradient model were accurate, it would be expected that the NS1 mRNA (closest gene to the promoter) would have the highest mRNA levels. It was also reported that the M mRNA, which is transcribed from the 5th gene from the promoter, also had higher levels than any of the four mRNAs whose genes were located upstream (Aljabr et al., 2016). A

divergence from the polar gradient model has also been reported in quantification of viral mRNA levels (with the exception of the SH gene transcript) in hRSV using Northern Blot analysis. The virus was harvested at a similar time to the previous report between 16-20 hpi. It was reported that mRNA levels did not follow levels predicted by the polar gradient model. However, unlike data acquired from transcriptome sequencing, levels of mRNA abundance from the genes located in the centre of the genome did not significantly alter. If there was an increase in mRNA abundance compared to the upstream gene, it was significantly smaller to what was reported in transcriptome sequencing. There was one large increase in the levels of mRNA abundance reported between the F and M2 genes (Krempl et al., 2002). Further evidence for this model being inaccurate was reported in the transcriptome sequencing of cells infected with hendra virus (*Paramyxoviridae* family). Similar to hRSV, it was reported that the G mRNA had higher mRNA levels than the upstream F mRNA. This difference was not as significant as observed in hRSV (Wynne et al., 2014). Although these data do suggest that the model is inaccurate, the sequential order of the genome still appears to be the defining factor in dictating viral mRNA abundance (Aljabr et al., 2016; Krempl et al., 2002; Wynne et al., 2014).

It could be possible that the transcription polar gradient model may still be accurate even with the data mentioned above. This is due to transcriptome sequencing not accounting for viral mRNA half-life. It could be possible that the levels of mRNA transcribed are relative to the position of the gene in the genome. However different mRNA half-lives could distort the levels of mRNA in the cell when transcriptome sequencing was performed. There is currently no data on mRNA half-lives for members of the *Pneumoviridae* family. It has been reported that in VSV, the half-lives of all viral gene transcripts are similar to the point where they are indistinguishable (Pennica et al., 1979).

The M2-1 protein is also involved in the transcription of the hRSV genome. This protein acts as a factor to inhibit premature termination by associating itself with the P, L and N proteins in inclusion bodies where transcription and replication occur. The precise mechanism in which the M2-1 protein acts as an anti-termination factor is unknown. There are three suggested mechanisms that have been proposed. 1: The protein prevents rehybridisation or secondary structure formation, thereby allowing

elongation to continue (Cartee and Wertz, 2001). 2: By increasing the affinity of the L:RNA interaction. 3: By stopping the release of the L protein once a GE sequence has been reached, therefore allowing the L protein to reinitiate transcription at the next GS signal. (Blondot et al., 2012) NMR and RNA disassociation experiments have revealed that M2-1 protein must initially bind to the P protein in order to bind to the viral RNA. However, once in proximity of the L protein and RNA, the M2-1 protein disassociates from the P protein and associates itself with the L protein. The M2-1 protein also has a high binding affinity to the 3′ end of the leader RNA suggesting that the protein could associate itself with the L protein during initiation. The M2-1 protein is also partially phosphorylated which is thought to aid its function; although the precise role of phosphorylation remains unclear (Blondot et al., 2012; Mason et al., 2003; Tran et al., 2009).

1.3.3. Viral replication

Following transcription, the L protein transcribes a positive-sense copy of the genome known as the antigenome (Fig. 1.4 red strand). This is also encapsidated by the nucleoprotein and subsequently acts as a template for replication, resulting in the production of new negative-sense viral genomes (Fig. 1.4 blue strand). The L protein will also synthesise additional nucleotides to 3′ end of the antigenome. Although the precise function of the addition of these nucleotides is unknown, it is thought that these may act to protect the leader sequence of the genome from RNases.

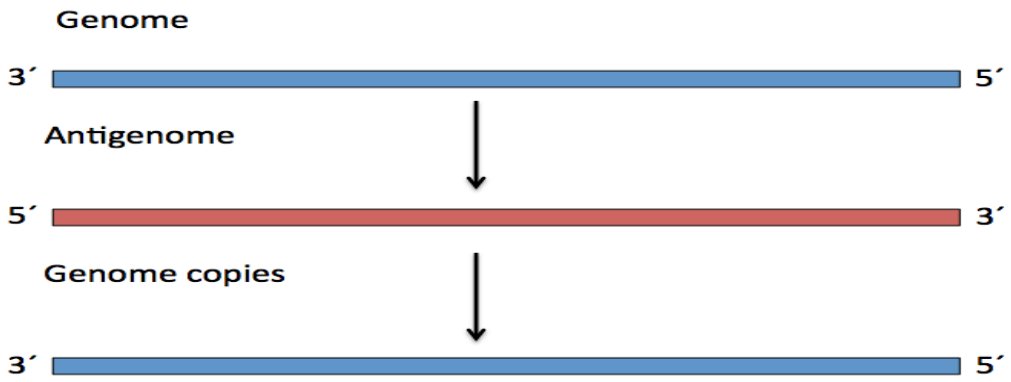


Fig. 1.4: Schematic of *Pneumoviridae genome replication*
Schematic of *Pneumoviridae* genome replication. Viral polymerase will transcribe an antigenome (red) from the genome *(*blue). The viral polymerase will then transcribe multiple copies of the genome (blue) from the antigenome

The switch between transcription and replication is thought to occur between 12-15 hours post infection, with the majority of RdRPs after this point transcribing an antigenome (Bermingham and Collins, 1999). Initiation of replication by the L protein begins with the L protein initiating synthesis at the 3′ promoter found in the leader sequence of the antigenome. Inside the leader sequence, nucleotides 1-13 are the key nucleotides that act to recruit the L protein to initiate RNA synthesis (Noton et al., 2010). Replication is thought to occur in the cytoplasm within inclusion bodies, which consists of the ribonucleoprotein complex (N and P proteins), the L protein and M protein (Grosfeld et al., 1995). Multiple copies of the genome are transcribed from the antigenome which are then encapsidated by the N protein.

The M2-2 protein is thought to aid in the change from transcription to replication. Although there is little research on this process, it has been shown that the level of M2-2 protein has a direct effect on the amounts of viral mRNA produced. The M2-2 protein is expressed in relatively small quantities compared to that of other viral proteins in the genome. Overexpression of the M2-2 protein leads to a large decrease in viral replication (Cheng et al., 2005). It is also important to note that the deletion of the M2-2 gene from the genome does not lead to a halt in replication of the virus but does greatly reduce the ability for the virus to replicate, with the yield of the virus decreasing up to 1000-fold (Bermingham and Collins, 1999; Cheng et al., 2005).

1.3.4. Assembly and budding

Upon termination of replication, inclusion bodies travel to the cell membrane where they join with viral glycoproteins that are bound to the cell membrane (Fig. 1.2E and 1.2F). From here, these proteins assemble into viral filaments, which lead to the formation of the virion. The matrix protein is thought to aid in the maturation of filaments in the late stages of maturation (Mitra et al., 2012; Shaikh et al., 2012). The viral filaments will form when the viral proteins deform the host cell membrane. The M protein is thought to interact with the phospholipids in the plasma membrane in order to achieve this (Money et al., 2009; Shaikh and Crowe, 2013). Following assembly of the viral genome and proteins, the virus will bud from the plasma membrane, this involves a scission event which is directed by cellular proteins (Mitra et al., 2012). The exact mechanism of budding for hRSV is unknown, however Vps-

4 and other proteins such as FIP2, which are commonly used by many other viruses for budding are not used (Utley et al., 2008). This budding event allows the release of newly infectious particles (Bermingham and Collins, 1999; Collins and Graham, 2008; Cowton et al., 2006; Mitra et al., 2012).

1.4.1. Translation of *Orthopneumovirus* viral gene transcripts

All gene transcripts transcribed by the L protein in all members of the *Orthopneumovirus* genus are capped and polyadenylated. With the exception of translation of the M2-2 protein from M2 ORF-2, all gene transcripts initiate translation through the canonical cap-dependent initiation pathway that is used by cellular mRNAs for initiation of translation. This pathway uses the m^7G cap located on the 5′ end of each gene transcript to bind eukaryotic initiation factors (eIF) to the mRNA. The binding of eIFs to the m^7G cap initiates a cascade of recruitment of initiation factors which ultimately leads to the recruitment of the 40S ribosomal complex to the mRNA strand forming the 43S scanning complex. This 43S scanning complex will then scan down the 5′ untranslated region (UTR) until the complex reaches an AUG codon with an optimal Kozak sequence (GCCACCAUGG) (Kozak, 1986, 1997). The 43S scanning subunit will bind to the AUG codon through the Met-$tRNA^{Met}_i$ bound in the subunit. This will then set off a further cascade which will recruit the 60S subunit to the complex and cause the release of bound eIFs to form the 80S subunit. The ribosomes will then begin the elongation phase of translation and synthesize the nascent polypeptide with aid from eukaryotic elongation factors (eEF) to deliver aminoacyl-tRNAs, aid in peptide formation and aid in translocation of the ribosome. Upon reaching a terminating stop codon (UGA,UAA or UAG), the ribosome will release the nascent polypeptide and use eukaryotic release factors (eRF) to initiate the dissociation of the 80S complex and release from the mRNA strand. The elongation and termination phases of translation are thought to be used by all gene transcripts with the exception of termination of translation in M2 ORF-1. A detailed review covering these mechanisms can be found for initiation (Jackson et al., 2010) and elongation/termination (Dever and Green, 2012). In the *Orthopneumovirus* genus, levels of translation with the exception of the M2 ORF-2 are controlled by mRNA abundance, which is dictated by the polar gradient model of transcription.

Regulation of translation in viruses tend to occur in order to maximise the use of their small viral genomes. This can include using polycistronic gene transcripts to translate multiple proteins from a single gene transcript. Viruses access multiple ORFs through employing non-canonical initiation mechanisms. These mechanisms as well as allowing access to multiple ORFs on a gene transcript, allow a virus to tightly regulate levels of translation of specific proteins that may be required at different levels during the entire or specific stages of infection.

Viruses are also capable of disrupting host cell translation by regulating canonical initiation of translation. This allows the virus to maximise viral translation in the cell by utilizing non-canonical initiation mechanisms to continue viral translation, whilst limiting initiation of translation cellular mRNAs. Viruses tend to specifically target translational initiation factors involved with the initiation of the 80S ribosome through the cap-dependent canonical initiation pathway. For example, both adenovirus (Yueh and Schneider, 2000) and vesicular stomatitis virus (VSV) (Connor and Lyles, 2002) reduce the levels of phosphorylated eIF4E through the dephosphorylation of the initiation factor. Phosphorylation of eIF4E increases the binding affinity to the m^7G cap allowing the eIF4F complex to bind to the mRNA. A decrease in the level of phosphorylated eIF4E results in a reduction of eIF4F complexes bound to the mRNA cap, which further prevents the 43S ribosomes from binding to the mRNA.

Other members of the eIF4F complex are also targeted by viruses. During infection in cells by poliovirus, eIF4G is cleaved by the poliovirus 2A protein. eIF4G acts as a scaffold protein in the eIF4F complex by binding to eIF4E and eIF4A. The initiation factor also binds to initiation factors in the 43S complex thereby allowing the ribosome to bind to the mRNA and begin scanning the 5′ UTR. Cleavage of this protein prevents the formation of the eIF4F complex and therefore prevents the 43S ribosome complex from binding to the mRNA (Gradi et al., 1998).

Other viruses also target the tertiary complex. Measles virus reduces canonical initiation by phosphorylating the eIF2α subunit of the eIF2 protein. Although this

only partially inhibits canonical initiation, phosphorylation of eIF2α prevents the exchange of GDP with GTP with the eIF2B protein. The eIF2-GTP complex increases the binding affinity of the initiation factor to Met-tRNA$^{Met}_{i}$. Failure to replace GDP with GTP significantly reduces the affinity, which in turn reduces the levels of tertiary complexes capable of providing Met-tRNA$^{Met}_{i}$ for initiating ribosomes (Sato et al., 2007). Viruses may also reduce the levels of host cellular eIF2α phosphorylation in order to limit the host cell response to infection. hRSV prevents host cellular phosphorylation of eIF2α by using the viral N protein to bind to the protein kinase R which is responsible for host cellular phosphorylation of eIF2α (Groskreutz et al., 2010).

A non-canonical initiation mechanism has been reported in the *Pneumoviridae* for the translation of the M2-2 protein from the second ORF in the M2 gene transcript through the use of coupled translation termination/initiation. This mechanism is used to both access the ORF inside a bicistronic gene transcript and to control the levels of translation for a protein which is required at very precise levels during infection (Ahmadian et al., 2000; Gould and Easton, 2005).

1.4.2. Translation of M2 ORF-2

In all members of the *Pneumoviridae* family, the M2 ORF-2 is located on the bicistronic M2 gene transcript. This gene transcript contains two overlapping open reading frames, with the first (M2 ORF-1) being translated into the M2-1 protein and the second (M2 ORF-2) being translated into the M2-2 protein (Fig. 1.5A). In this family, the open reading frames overlap with varying distances and in different frames to each other. The size and frame depends on the virus in question. In current literature, only three members of the *Pneumoviridae* family (hRSV (Ahmadian et al., 2000), APV (Gould and Easton, 2007) and PVM (Gould and Easton, 2007) have had the initiation mechanism for translation of M2 ORF-2 characterised. For all three members, the non-canonical mechanism of coupled translation termination/initiation has been identified as the mechanism used for the initiation of translation of the M2 ORF-2 on the M2 bicistronic gene transcript. The mechanism is reliant on the translation of the preceding ORF, M2 ORF-1 and relies on the terminating ribosome of the first ORF (M2 ORF-1) reinitiating translation for the second ORF (M2 ORF-

2). The mechanism is not dependent on viral proteins as studies have revealed that transfecting uninfected cells with the M2 gene encoded on an expression plasmid still results in translation of the second ORF (Ahmadian et al., 2000; Gould and Easton, 2007).

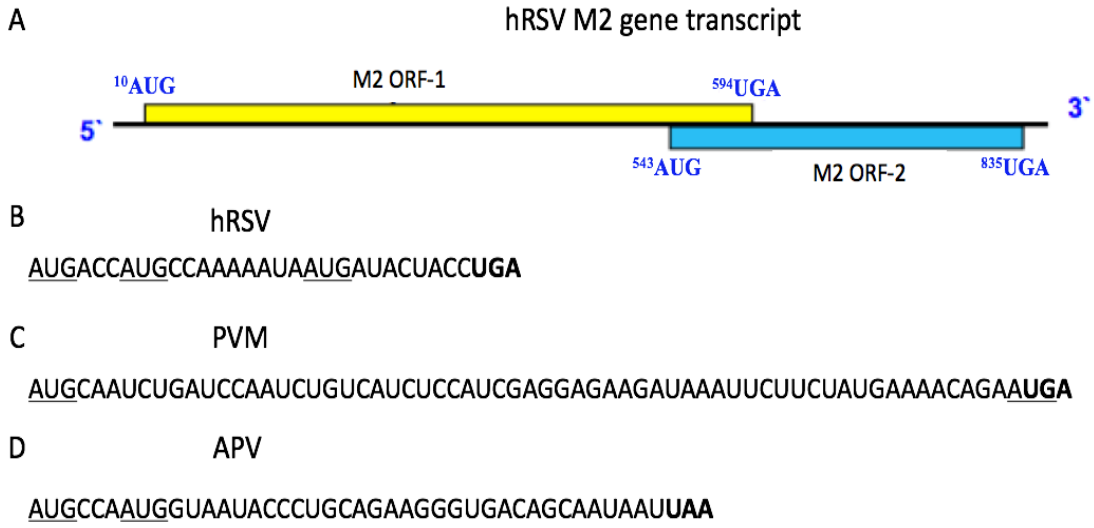


Fig. 1.5: Schematic of hRSV M2 gene transcript and sequences of the M2 ORF-1/ORF-2 overlap regions for members of the *Pneumoviridae* family that use the mechanism of coupled translation
(A) Schematic of the hRSV M2 gene transcript. The black line represents the mRNA strand. Boxes represent ORFs located on the mRNA strand. Start and terminating codons for M2 ORF-1 (yellow) and M2 ORF-2 (blue) are displayed. (B) Sequence of the hRSV M2 ORF-1/ORF-2 overlap region. AUG codons used for initiation of ORF-2 are underlined. The terminating stop codon for ORF-1 is in bold. (C) Sequence of the PVM M2 ORF-1/ORF-2 overlap region. AUG codons used for initiation of ORF-2 are underlined. The terminating stop codon for ORF-1 is in bold. (D) Sequence of the APV M2 ORF-1/ORF-2 overlap region. AUG codons used for initiation of ORF-2 are underlined. The terminating stop codon for ORF-1 is in bold.

In the current literature there has been no other investigation into the mechanism of translation of M2 ORF-2 in bRSV, ovine RSV, caprine RSV and human metapneumovirus (Ahmadian et al., 2000; Gould and Easton, 2007). Ovine RSV and caprine RSV although closely related to hRSV, have yet to be classified into the *Orthopneumovirus* genus by ICTV (ICTV, 2016). Out of the three members that have had the initiation mechanism of M2 ORF-2 translation identified, hRSV is the best characterised mechanism. In hRSV (Fig. 1.5B), M2 ORF-2 is in the second (+1 to M2 ORF-1) reading frame. M2 ORF-1 and M2 ORF-2 overlap each other by 32 nts, this is between the first viable AUG codon that can be used for initiation of translation of M2 ORF-2 and the terminating stop codon of M2 ORF-1. There are three start codons for M2 ORF-2, all of which can be used to initiate translation, with

a slight preference for initiation occurring at the two start codons at the 5′ end of the M2 ORF-1/ORF-2 overlap region (Ahmadian et al., 2000).

The M2 ORF-1/ORF-2 overlap region in PVM (Fig. 1.5C) is significantly larger than hRSV at 67 nts in size. In PVM, M2 ORF-2 is in the third (+2 to M2 ORF-1) reading frame. The overlap region contains two start codons located at either end of the overlap region. Both of these start codons can be used for initiation of M2 ORF-2. Although there is a slight preference for initiation at the AUG start codon at the 5′ end of the overlap. There have been no further studies to characterise the mechanism in PVM (Gould and Easton, 2007). APV, a member of the *Metapneumovirus* genus has an overlap region of 34 nucleotides and like PVM, the M2 ORF-2 is the third reading frame (Fig. 1.5C). APV has two start codons, both located towards the 5′ end of the overlap region. Both of these are used for initiation of M2 ORF-2, although there is a preference for initiation of translation at the AUG located at the 5′ end of the overlap region (Gould and Easton, 2007).

As seen in Fig. 1.6, the mechanism of coupled translation is reliant on the termination of a ribosome that is translating the upstream M2 ORF-1 (Fig. 1.6B). Upon termination of the ribosome (Fig. 1.6C), the nascent M2-1 polypeptide will be released and the majority of ribosomes will disassemble and disengage from the mRNA. However, a small proportion of ribosomes (~3%) (P. Gould and A. Easton, personal communication) will remain attached to the mRNA. These remaining ribosomes will translocate upstream (Fig. 1.6D) until reaching be one of start codons found in the M2 ORF-1/ORF-2 overlap region and reinitiating translation (Fig. 1.6E). Removing the stop codon of M2 ORF-1 completely inhibits the translation of the M2 ORF-2, as the ribosome is unable to terminate translation of M2 ORF-1 and translocate upstream. The mechanism is also distance dependent on the stop codon. Removing the M2 ORF-1 stop codon and reintroducing a stop codon in the same frame as M2 ORF-1 further downstream but still in close proximity to the original M2-1 stop codon, completely inhibits translation of M2 ORF-2 (Ahmadian et al., 2000; Gould and Easton, 2005).

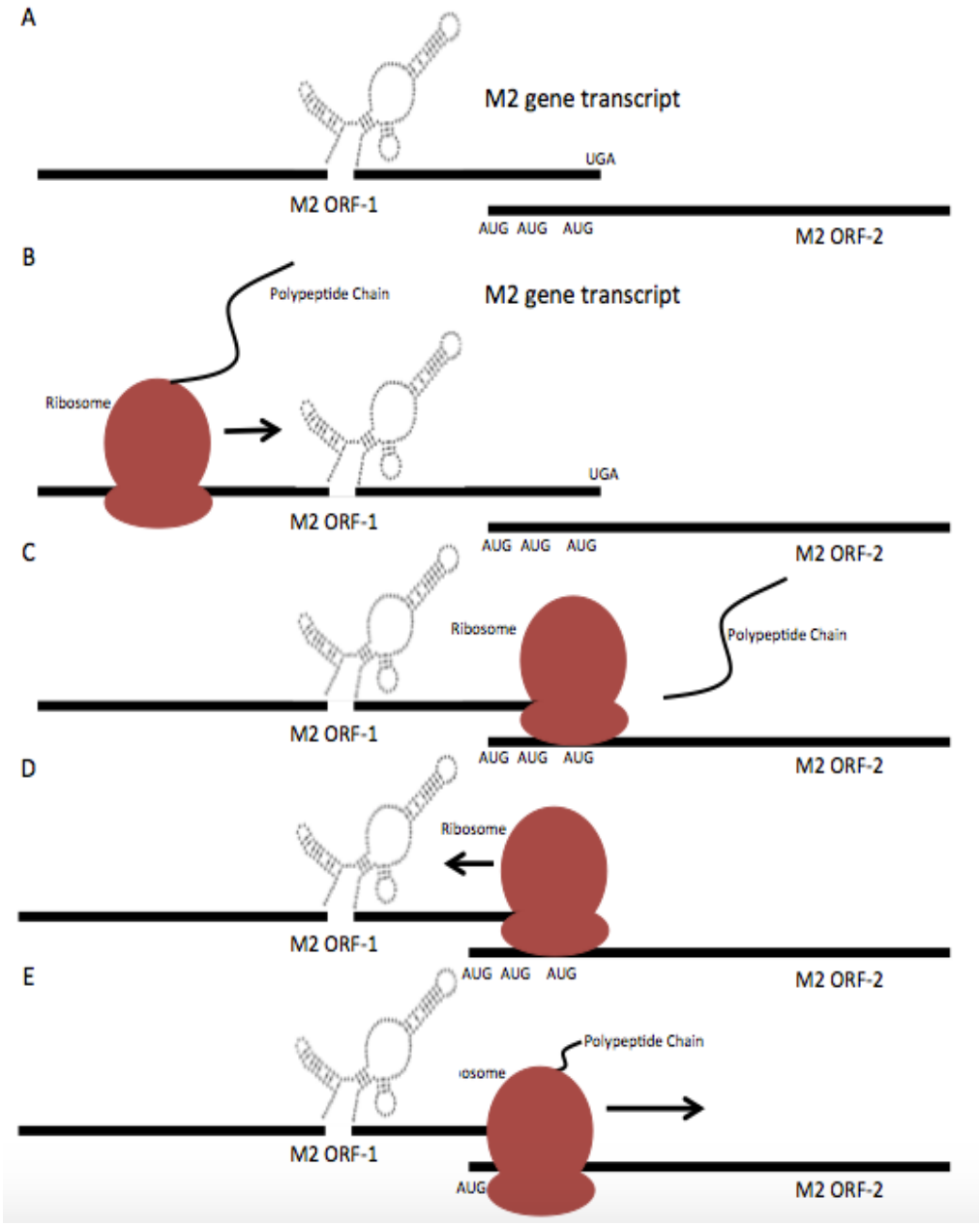


Fig. 1.6: Mechanism of coupled translation termination/initiation for the hRSV M2 ORF-2
(A) Schematic of the hRSV M2 gene transcript, which displays the initiation sites for M2 ORF-2 and terminating stop codon for M2 ORF-1. (B) Translation by ribosome of M2 ORF-1. (C) Termination of translation at the M2 ORF-1 stop codon and release of the M2-1 polypeptide chain by the ribosome. (D) Translocation upstream by the terminating ribosome. (E) Initiation of translation of M2 ORF-2 by the translocating ribosome at one of the three initiation codons located in the M2 ORF-1/ORF-2 overlap region. A representation of the mRNA secondary structure critical for initiation of M2 ORF-2 that is located between position 396 and 511 in the M2 ORF-1 is presented in all panels. The secondary structure is thought to pause elongating ribosomes translating M2 ORF-1 allowing for successful reinitiation of translation for M2 ORF-2 (P.Gould and A.Easton, personal communication)

The sequence of the M2 gene transcript is also significant to the mechanism of coupled translation. For hRSV, sequence 5´ of the M2 ORF-1/ORF-2 overlap region in the M2 ORF-1 is vital for the mechanism. There is however, no vital sequence for

this mechanism 3′ of the overlap region. The M2 ORF-2 sequence 3′ of the M2 ORF-1/ORF-2 overlap region can be replaced or removed and not affect initiation of translation of the second ORF. This characteristic has enabled reporter plasmids to be created that were specifically designed to investigate the mechanism of coupled translation termination/initiation. These reporter plasmids contained an altered M2 gene that would still be transcribed in the same format as the bicistronic M2 gene transcript. In this gene, the M2 ORF-2 sequence past the M2 ORF-1/ORF-2 overlap region was replaced by a reporter gene that codes for the CAT protein. Upon transfection of plasmids into cells and transcription of the gene, the reporter protein was expressed from the second ORF. This was then quantified using an enzyme-linked immunosorbent assay (ELISA) specific for the reporter protein. This enabled quantification of translation of the second ORF and was the primary method used to investigate and characterise this particular mechanism (Ahmadian et al., 2000).

Gould and Easton, 2005 further characterized the role of the M2 ORF-1 sequence upstream of the M2 ORF-1/ORF-2 overlap region by creating chimeric plasmids where increasing proportions of M2 ORF-1 from the 5′ end of gene transcript were deleted and replaced with eGFP coding sequence. These chimeric plasmids used the same reporter system as mentioned above to quantify the expression of M2 ORF-2. Removal of the first 320 nts of the M2 ORF-1 sequence from the 5′ end of the gene transcript slightly affected the efficiency of coupled translation. However, this region was not critical, as removal did not directly inhibit the mechanism. Removal of the next 100 nucleotides markedly lowered translation to 10% in comparison to expression of ORF-2 for plasmids that contained the full length M2 ORF-1 sequence. A removal of a further 20 nucleotides caused complete inhibition of translation of the second ORF. Using this data, primer extension and toe printing assays revealed a large secondary structure found in the region that was critical for the translation of M2 ORF-2 between position 414 and 594 (Gould and Easton, 2005).

This secondary structure was fully identified with use of selective 2′-hydroxyl acylation and primer extension (SHAPE) mapping and was found to consist of four stemloops, which contained three A-rich blocks inside the structure (P. Gould and A. Easton, personal communication). The secondary structure was found to be the primary cause of inhibition of translation of ORF-2 observed in Gould and Easton,

2005. Mutations altering the entire secondary structure completely inhibited the translation of the second ORF. Further mutations revealed that two of the three A-rich blocks had a direct inhibitory effect on the translation of second ORF (P. Gould and A. Easton, personal communication). It was reported that the cellular protein DDX3X (P. Gould and A. Easton, personal communication), an RNA helicase (Garbelli et al., 2011) bound to the A rich blocks. The protein was found to be of crucial importance to the mechanism. Inhibiting the ATPase activity of DDX3X caused complete inhibition of the translation of M2 ORF-2. Although DDX3X is vital for this mechanism, it is still unknown how this protein contributes to this mechanism (P. Gould and A. Easton, personal communication).

Although the significance of the M2 ORF-1 sequence on the initiation of M2 ORF-2 has been well documented, the effect that this sequence has on elongating ribosomes has not. Ribosomal profiling was used to investigate the role of the secondary structure in M2 ORF-1. This technique uses next generation sequencing to map the position of all elongating ribosomes in a cellular population. Using this technique, it was reported that throughout the entire M2 ORF-1 sequence there were regions of sequence where a high frequency of elongating ribosomes were stalling that could be attributed to structured regions slowing down the elongating ribosomes. A significant peak comprising of over 5% of all elongating ribosomes translating M2 ORF-1 were identified pausing on a single nucleotide on stemloop II in the highly structured region critical for initiation of M2 ORF-2 (P. Gould and A. Easton, personal communication).

Combined with these previous data (Ahmadian et al., 2000; Gould and Easton, 2005), it has been suggested that the secondary structure is critical for the mechanism of coupled translation by pausing elongating ribosomes translating M2 ORF-1 for a significant amount of time, which allows terminating ribosomes further downstream to translocate upstream to reinitiate at one of the M2 ORF-2 start codons in the M2 ORF-1/ORF-2 overlap region. Without this secondary structure, elongating ribosomes translating the M2 ORF-1 would prevent the initiating ribosome from translocating upstream. Whether this pause is caused by an interaction with DDX3X is not known. However, the protein may act in a complementary way by unwinding the structure thereby allowing the ribosome to continue translating the first ORF.

DDX3X is known to act as a initiation factor for translation in certain cellular and viral mRNAs and has been suggested that the protein aids eIF4A in unwinding certain secondary structures in the 5´UTR (Soto-Rifo et al., 2012). It is therefore conceivable that DDX3X also is able to unwind certain region in the M2-1 ORF. Further studies are needed to clarify the role of this protein (P. Gould and A. Easton, personal communication).

1.5.1. The mechanism of coupled translation termination/initiation used in other organisms

This translational mechanism of coupled translation has also been documented in other viral, eukaryotic and prokaryotic organisms (Gould et al., 2014; Gould and Easton, 2005; Oppenheim and Yanofsky, 1980). In the majority of cases except some prokaryotic examples (Thomas et al., 1987), the organization of the ORF's tend to follow a similar pattern (Fig. 1.7) (Gould et al., 2014; Gould and Easton, 2005; Oppenheim and Yanofsky, 1980). The ORF using the mechanism tends to be found downstream of the ORF that it is reliant on. In many viral cases the two ORFs will overlap each other from a few nucleotides to tens of nucleotides (nts) (Ahmadian et al., 2000). This is also the case in many prokaryotic examples (Aksoy et al., 1984; Oppenheim and Yanofsky, 1980). However in prokaryotes, the two ORFs may also be separated by a small distance ranging up to tens of nucleotides away from each other (Hellmuth et al., 1991). In all cases the mechanism is reliant on the termination of the ribosome translating the upstream ORF.

Fig. 1.7: Typical structure of gene transcript, where the mechanism of coupled translation is used for translation of ORF-2
Typical structure of gene transcript, where coupled translation is used for initiation of translation of ORF-2. The black line represents the mRNA strand. Boxes represent ORFs located on the mRNA strand. Translation of ORF-1 is required for successful initiation of translation of ORF-2. Both ORFs usually overlap by a number of nucleotides.

Initially the mechanism of coupled translation was first discovered in phage viruses in the 1960s (Gussin, 1966; Horiuchi et al., 1966). A well characterised example found in phages is the translation of the MS2 bacteriophage replicase ORF. The replicase ORF is reliant on the translation of the coat ORF. The translation of the coat ORF disrupts two long range interactions between the two ORFs, which allows reinitiating ribosomes access to the initiation site of the replicase gene for translation (Licis et al., 1998).

Coupled translation was first characterized in prokaryotes in the *E.coli* tryptophan operon. This polycistronic mRNA contains five ORFs that once translated to proteins are responsible for tryptophan synthesis. As seen in Fig. 1.8, two ORF pairs in the gene transcript are found to be translationally coupled (*trpE/trpD* and *trpB/trpA*, with *trpA* (Aksoy et al., 1984) and *trpD* (Oppenheim and Yanofsky, 1980) using the mechanism). The two ORF pairs both have overlapping regions of 1 nucleotide (UGAUG). Similar to other examples, coupled translation of both *trpD* and *trpA* are reliant on terminating stop codon of *trpE* and *trpB* being in the vicinity of initiation codons for *trpD* and *trpA* (Das and Yanofsky, 1989). There are three suggested mechanisms by which coupled translation is used both in the tryptophan operon and in other prokaryotic examples. The first mechanism involves ribosomes that are translating the upstream ORF of the ORF pair removing secondary structures. These secondary structures in the mRNA sequence sequester initiation of the second ORF by either blocking the Shine Dalgarno (SD)/ribosome binding site (RBS) or start codon. This removal also increases the concentration of ribosomes around the RBS or SD sequence, that allows efficient translation of the second ORF to commence (Oppenheim and Yanofsky, 1980). The second mechanism suggests that the removal of the secondary structure by the elongating ribosome allows de-novo ribosomes to bind to the RBS site of the second ORF and initiate translation (Oppenheim and Yanofsky, 1980). The third mechanism involves the translating ribosome not disassociating once it has reached the stop codon of the first ORF. It will instead translocate upstream in search of a start codon of the second ORF. This will generally be under the circumstances where both ORFs overlap each other (Oppenheim and Yanofsky, 1980). As well as the tryptophan operon, coupled translation has been characterised in other prokaryotic mRNAs. This includes the *E.coli galK* ORF located in the galactose operon (reliant on translation of the *galT*

ORF*)* (Schümperli et al., 1982) and the *E.coli atpA* ORF*,* which is reliant on translation of the upstream *atpH* ORF (Rex et al., 1994). Unlike other organisms that use coupled translation, prokaryotic examples tend to use coupled translation to translate an ORF in 1:1 ratios in relation to the upstream ORF which the mechanism is reliant on (Oppenheim and Yanofsky, 1980; Rex et al., 1994; Schümperli et al., 1982).

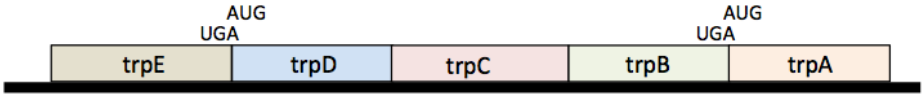


Fig. 1.8: Structure of *E.coli* tryptophan operon
Structure of the polycistronic *E.coliI* tryptophan operon. The black line represents the mRNA strand. Boxes represent ORFs located on the mRNA strand. Sequences for the stop and start codons of the ORF pairs *trpE/trpD* and *trpB/trpA*. The mechanism of coupled translation termination/initiation is used for initiation of translation of *trpD and trpA* ORFs. These two ORFs are reliant on the translation and termination of the *trpE* and *trpB* ORFs respectively.

Coupled translation has also been reported in polycistronic mRNAs in human cells. Although the majority of human mRNAs transcribed are monocistronic, it has been reported that there are examples of polycistronic mRNAs. The mechanism identified for initiation of translation of the second ORF in some of these polycistronic mRNAs is that of coupled translation termination/initiation. However in comparison to both prokaryotic and viral examples, the mechanism differs significantly (Gould et al., 2014).

The mechanism has been reported in five gene transcripts, where removal of the terminating stop codon for the first ORF inhibited translation of the second ORF. In all examples, the two ORFs overlap each other but vary in size. All five gene transcripts are in the second reading frame in relationship to the first ORF. Unlike examples in other organisms, the overlap regions are rich in multiple AUG codons that are in the frame of the second ORF (average of 7 AUG codons). It was reported however, that not all of the AUGs in the overlaps were used for initiation of the second ORF (Gould et al., 2014).

Out of the 5 examples identified that use the mechanism of coupled translation, the mechanism for the translation of *CASQ2* ORF-2 is the most characterised. It was reported that the overlap region was the only factor necessary for successful initiation of translation. In the *CASQ2* gene transcript, the overlap region (81 nts) is rich in aspartate residue (Asp) (GAU) repeats that are in the ORF-1 reading frame (Fig. 1.9). Initiation can occur at AUG codons at both ends of the overlap, with a slight preference for initiation at AUG codons located at the 3′ end of the overlap region. The GAU repeats are critical for the initiation of the second ORF in *CASQ2*. It has been suggested that instead of the repeated GAU sequence directing initiation of the second ORF, the Asp amino acids are responsible for successful initiation. It has been suggested that the Asp amino acids at the carboxyl terminus may cause interactions in the exit tunnel of the ribosome leading to stalling. This stalling may allow terminating ribosomes to reinitiate translation of the second ORF without encountering elongating ribosomes translating the first ORF. Replacing the Asp codons with other codons inhibited the mechanism. The other 4 examples that have been identified as using mechanism of coupled translation are also reliant on Asp residue repeats in the overlapping region. These Asp residue repeats are thought to have the same effect on the reinitiating ribosome as suggested during initiation of *CASQ2* ORF-2. Although only five polycistronic mRNAs were identified as using this mechanism, bioinformatic analysis has identified many more potential polycistronic mRNA that may use this mechanism (Gould et al., 2014).

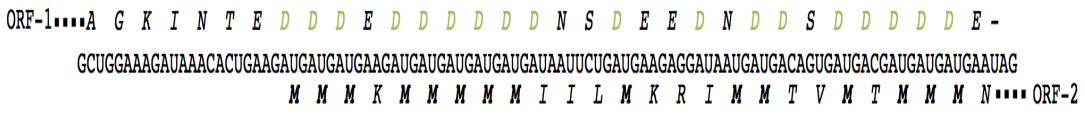


Fig. 1.9: Sequence of *CASQ2* bicistronic gene transcript
Sequence of the *CASQ2* gene transcript ORF-1/ORF-2 overlap region and the translated amino acid sequence for each ORF is shown above and below the mRNA sequence. Aspartic acid residues that have been identified as being critical to initiation mechanism of translation of ORF-2 are highlighted in green. AUG codons in the ORF-1/ORF-2 overlap region in ORF-2 are highlighted in bold.

The mechanism of coupled translation is also used in other viruses that infect eukaryotic organisms. There are six families of viruses which belong to groups III,

IV and V that are known to use this mechanism (Ahmadian et al., 2000; Fütterer and Hohn, 1991; Guo et al., 2009; Horvath et al., 1990; Li et al., 2011; Meyers, 2003). The majority of viruses that use this mechanism are reliant on the RNA sequences surrounding the initiation site of the ORF that uses this mechanism. Viruses that use this mechanism in the cauliflower mosaic virus (*Caulimoviridae* family) are not reliant on the sequence, but are reliant on the expression of a viral protein (Fütterer and Hohn, 1991).

A number of viruses in the *Caliciviridae* family (group IV) have been identified as using coupled translation to initiate translation of ORFs located on bicistronic subgenomic (sg) RNAs. As in all other examples, this mechanism is dependent on termination of elongating ribosomes translating the upstream ORF. Coupled translation has been reported in five members of this positive-sense single stranded RNA genome family (feline calicivirus (Luttermann and Meyers, 2007), murine norovirus (Napthine et al., 2009), human norovirus (Luttermann and Meyers, 2014), rabbit haemorrhagic disease virus (RHDV) (Meyers, 2003) and bovine norovirus (McCormick et al., 2008)). These are positive-sense single stranded RNA viruses that contain no viral envelope. The genome is approximately 7 to 8 kb long, which comprises of either two or three ORFs. As well as using the genome strand for translation, the virus will transcribe a subgenomic RNA strand that consists of ORF-2 and ORF-3 of the virus's three ORFs (Fig. 1.10). For the majority of viruses that use this mechanism (except bovine norovirus (McCormick et al., 2008)), coupled translation is used for the translation of ORF-3 on the subgenomic bicistronic RNA (Luttermann and Meyers, 2007) (Napthine et al., 2009) (Meyers, 2003). These ORFs are translated to produce the capsid proteins VP1 and VP2 (Luttermann and Meyers, 2007) (Napthine et al., 2009) (VP60 and VP10 in RHDV (Meyers, 2003)) for ORF-2 and ORF-3 respectively. The two ORFs found in the subgenomic RNA overlap each other by varying lengths; with human and murine norovirus both overlapping by one nucleotide (UAAUG) (Napthine et al., 2009; Luttermann and Meyers, 2014), feline calicivirus (FCV) overlapping by 4 nucleotides (AUGA) (Luttermann and Meyers, 2007) and RHDV by eight nucleotides (AUGUCUGA) (Meyers, 2003). For all these viruses, translation of the ORF-3 is reliant on the translation of ORF-2 (Luttermann and Meyers, 2007) (Napthine et al., 2009) (Meyers, 2003). The levels of protein translated from the third ORF are much lower than levels translated in the second

ORF, with the feline calicivirus producing VP1 and VP2 in a molar ratio of 18:1 (Luttermann and Meyers, 2007).

Fig. 1.10: Structure of the bicistronic calicivirus subgenomic RNA
Schematic of calicivirus subgenomic RNA. The black line represents the subgenomic RNA strand. ORF-2 is highlighted in yellow and ORF-3 is highlighted in blue.

However, unlike the other three examples that use coupled translation, bovine norovirus differs by using this mechanism to translate the capsid protein from ORF-2 via the genomic RNA. The translation of ORF-2 is reliant on the translation of ORF-1. Other viruses in this family usually translate ORF-2 from the subgenomic RNA. Although this still occurs in bovine norovirus as the capsid protein is translated from both genomic and subgenomic mRNA, it has been suggested that the capsid protein in bovine norovirus may also be involved in the earlier stages of the viral life cycle and therefore is required to be translated earlier from the genomic RNA (McCormick et al., 2008).

Although bovine norovirus uses the same mechanism for coupled translation for ORF-2 as other members use for ORF-3, the rest of this review will refer to the translation of ORF-3 in the subgenomic mRNA using the mechanism of coupled translation.

All caliciviruses that use this mechanism are reliant on a region of sequence that is located approximately 70 nucleotides upstream from the terminating stop codon of ORF-3 (Fig. 1.11A). This region is referred to as the termination upstream ribosomal binding site (TURBS). The region consists of three sequence motifs one of which (Motif 1) contains a five nucleotide sequence (UGGGA), which is conserved between all members of this family that use this mechanism (Fig. 1.11B)

(Luttermann and Meyers, 2007; Meyers, 2007). The size of the motif varies between different viruses, for example, feline calicivirus (6 nts) and RHDV (15 nts). Motif 2 in the TURBS sequence is non-conserved throughout this viral family. The sequence of Motif 2 differs significantly between different viruses. For example, FCV Motif 2 sequence is AGGAGU (Luttermann and Meyers, 2007). Whereas CUGGGG (Meyers, 2007) is RHDV's Motif 2 sequence. The third motif (2*) is found directly upstream of Motif 1. It is similar in composition to Motif 2 in that it is not conserved throughout this family of viruses. Between the motifs are spacer sequences and whilst are not essential, have been suggested to determine the difference in efficiency of this mechanism between different members of this family (Luttermann and Meyers, 2009; Meyers, 2007).

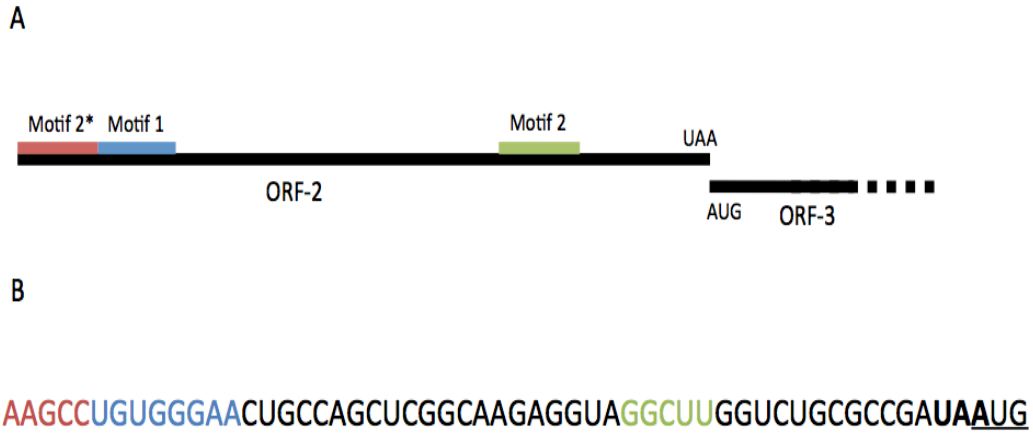


Fig. 1.11: TURBS used for initiation of ORF-3 in human norovirus subgenomic RNA
(A) Location of TURBS motifs located in ORF-2 of human norovirus subgenomic RNA. The black line represents the subgenomic RNA strand. Motif 1 is highlighted in blue, Motif 2 is highlighted in green and Motif 2* is highlighted in red. (B) Sequence of TURBS and ORF-2/ORF-3 overlap region of the subgenomic human norovirus. The initiation codon used for ORF-3 is underlined. The terminating stop codon for ORF-2 is in bold. For the TURBS sequence, the sequence for Motif 1 is highlighted in blue. The sequence for Motif 2 is highlighted in green. The sequence for Motif 2* is highlighted in red.

It has been shown that Motif 1 binds to helix 26 (h26) in the 18S rRNA sequence found in the 40S ribosomal subunit (Luttermann and Meyers, 2009; Meyers, 2007; Pöyry et al., 2007). Motif 2* is thought to form a secondary structure with Motif 2 and is thought to place Motif 1 into an optimal position which efficiently binds to h26. Mutations altering the structure cause a dramatic decrease in translation efficiency of ORF-3. It is thought, that by binding to h26, the interaction brings the 18S rRNA into an optimal position to the viral RNA. This allows certain residues

located on the 18S rRNA that are associated with canonical initiation of translation, to be brought into their optimal positions on the viral RNA strand (Martin et al., 2016). This in turn is thought to position the ribosome so that the start site for ORF-3 is positioned in the P site of the ribosome (Luttermann and Meyers, 2009) .

There have been three proposed models that may be used by ribosome to reinitiate translation at the ORF-3 initiation site. Two of these use the 40S subunit and one the 80S subunit. It is thought that all three models can be used by the ribosome for initiation of translation of ORF-3.

For the first model, the mechanism uses eIF3 for successful initiation through interactions between the TURBS sequence and eIF3. It was initially suggested that by binding the 40S subunit and eIF3, the TURBS sequence binds to the 40S subunit long enough to encourage initiation factors to bind and aid in the recruitment of the 60S subunit (Pöyry et al., 2007). Although little is known about the interaction between the mRNA and eIF3, it has been shown that mutagenesis of Motif 1 has little effect on the binding of eIF3 (Luttermann and Meyers, 2009).

A second and more recent model, has suggested that initiation of ORF-3 is eIF3 independent. For this model, the 40S subunit following termination is capable of reinitiating translation using either eIF2/1/1A or Ligatin as cofactors with the TURBS sequence to stabilise the subunit allowing the recruitment of the 60S subunit. However, in this model eIF3 was found to still hold importance in the mechanism under conditions where there was a lack of alternate initiation factors (Zinoviev et al., 2015).

For the third model, reinitiation is achieved without the separation of the ribosomal complex. 80S ribosomal subunits have been found to be also capable of reinitiating translation due to the stabilizing effect that the TURBS sequence provides (Zinoviev et al., 2015).

All three ribosomal mechanisms (80S and both 40S methods) that are used for reinitiation use either the native or near cognate start codons for initiation, with the mechanism using eIF2/1/1A being the most tolerant to deviation from the start codon

(Zinoviev et al., 2015). The start codons used for initiation of translation of ORF-3 are not essential for reinitiation. In some members of the calicivirus family, replacing one, two or even three nucleotides in the AUG codon used for initiation of ORF-3 still leads to successful initiation of translation of ORF-3. In FCV, these mutations reduced the levels of ORF-3 expression by 40%, 70% and 88% respectively. This is slightly lower in RHDV with the replacement of one nucleotide reducing expression by 30% and replacement of the entire start codon completely inhibiting translation of ORF-3. The context of the Kozak sequence surrounding the AUG codon also has no bearing on the efficiency of reinitiation of ORF-3 in all viruses in this family (Luttermann and Meyers, 2014).

It has been discovered that in human norovirus and FCV, initiation of ORF-3 can also occur downstream of the ORF-2/ORF-3 overlap region when the native initiation codon for ORF-3 is mutated. Initiation using the start codons located downstream is also dependent on the Kozak sequence surrounding the start codon (Luttermann and Meyers, 2014). It has been suggested that a fraction of the ribosomes used for initiation of ORF-3 are capable of scanning downstream if the ribosome is unable to use the start codons associated with coupled translation. This fraction of terminating ribosomes is thought to use a classical scanning-like mechanism post termination to reinitiate at this downstream start codon. 40S subunits are likely to be involved in this mechanism in conjunction with eIF4F to scan for the downstream start codons. This alternative model is not used in RHDV where as mentioned, removal of the main start codons leads to complete inhibition of translation for the second ORF (Luttermann and Meyers, 2014; Zinoviev et al., 2015). Under standard conditions the ribosome will almost always prefer to use the native start codon. This start site is preferred by all three separate mechanisms (Zinoviev et al., 2015). It has been reported that sequence directly downstream of the ORF-2 terminating stop codon is of importance for the mechanism. Although the specific purpose of this sequence is unknown, it is thought to modulate the mechanism. This regulation is believed to be both reliant on the location of sequence in relation to the initiation site and also the specific sequence. It has been suggested that this acts through some form of secondary structure although mutational studies could not confirm a predicted structure in this region (Haβ et al., 2014).

Depending on whether the codon is non-cognate or cognate, elongator tRNAs have been found to be used in reinitiation instead of Met-tRNA$^{Met}_{i}$. However it is suggested that fewer elongator tRNAs will be used in the presence of a native AUG as Met-tRNA$^{Met}_{i}$ and so will always have a higher affinity. The tRNA found in the P site during termination has also been found to have an impact on the levels of reinitiation which may be due to the affinity in the P/E hybrid state. Examples such tRNA^{Val} and tRNA^{Arg} reduced the efficiency for translation. This therefore may have been due to sensitivity of the tRNA to Mg^{2+} (Zinoviev et al., 2015).

The mechanism of coupled translation using a TURBS sequence to reinitiate translation in the second ORF of a bicistronic mRNA has also been reported in the influenza B virus for translation of the protein BM2. The negative-sense single stranded genome virus influenza B is a member of the *Orthomyxoviridae* family (Powell et al., 2008). The gene for the BM2 protein is located on the seventh genomic segment, which is transcribed into a bicistronic mRNA strand with the first ORF being translated to the M1 protein and the second ORF translated to the BM2 protein. The two ORFs overlap by 1 nucleotide UAAUG. It has been reported that translation of BM2 is reliant on the termination of translation of M1. Removal or changing the location of the M1 terminating stop codon inhibits the initiation of translation of BM2 (Horvath et al., 1990). It also been reported that similar to members that use this mechanism in the *Caliciviridae*, translation of the BM2 ORF is reliant on a 45nts sequence directly upstream of the AUG codon. This sequence is similar to the TURBS Motif 1 sequence observed in caliciviruses (Fig. 1.12). Unlike TURBS sequences in caliciviruses, the TURBS sequence is unstructured, this is likely due to the lack of Motif 2 and Motif2*. Motif 1 of influenza B similar to caliciviruses, has been shown to interact with the h26 helix on the 18S rRNA (Powell et al., 2011; Powell et al., 2008). The 40S ribosomal subunit is thought to be responsible for reinitiation of translation. eIF3 has also been suggested to aid in the process of reintiation. It also has been suggested that the distance between the stop codon and the TURBS sequence is more important to the mechanism than the distance between the AUG and the stop codon. Initiation can still occur albeit at a significantly lower level if the AUG located in the overlap is removed and the next viable start codon is located up to 63 nucleotides downstream. However the ribosome will reinitiate at the start codon in the overlap region if given a choice

between the AUG codon in the overlap and an alternate AUG codon in close proximity (Powell et al., 2011).

Fig. 1.12: Sequence of the influenza B segment 7 M1 ORF/ BM2 ORF overlap region and TURBS like Motif 1.
Sequence of the influenza B segment 7 M1 ORF/ BM2 ORF overlap region. The AUG codon used for initiation of translation of BM2 ORF is underlined. The terminating stop codon for the M1 ORF is in bold. The sequence for Motif 1 used for initiation of translation of BM2 ORF through the mechanism of coupled translation termination/initiation is highlighted in red.

Whilst the majority of viruses that use coupled translation tend to infect animal hosts, coupled translation has been identified in the fungal-infecting *Helminthosporium victoriae* virus 190S from the *Totiviridae* subfamily. This double stranded RNA virus is a non-segmented virus comprising of two ORFs that code the capsid protein and the RNA polymerase protein (RdRP). These two ORFs overlap each other with the following sequence AUGA. A second AUG codon can be used for reinitiation of the second ORF, which is located 90 nucleotides downstream of the overlap region. However, initiation is preferred at the start codon located in the overlap region (Li et al., 2011).

Translation of the RdRP ORF is reliant on the stop codon. Studies have also identified a predicted pseudoknot like structure located within 13 nucleotides upstream of the overlapping region. This predicted secondary structure has been found to be of importance in the translation of the second ORF, however the precise role of this structure is still unknown. There is evidence to suggest that the structure may be similar to a TURBS sequence found in caliciviruses and influenza B. This is due to the predicted structure having a complementary sequence (CUGAUCG) to the 18S rRNA of the infected host cell (helminthosporium savius) ribosome. However, this potential function has been deemed improbable, as this complementary sequence is not in the same region of the 18S rRNA that TURBS motifs bind to. Structural predictions also suggest that this sequence is also located in the stem of the structure,

which might mean that it may be impossible for this sequence to bind to the rRNA. It has also been suggested this predicted structure may act as barrier to prevent the ribosome from translocating too far upstream from the overlap region which allows the ribosome to reinitiate in the vicinity of the start codon for the RdRP (Li et al., 2011, 2015a).

Another possibility that has not been discussed is that it still is feasible that the unwinding of this stem loop structure by elongating ribosomes or other cellular factors, allows an interaction between the mRNA strand and the 18S RNA to occur similar to examples of coupled translation observed in prokaryotes (Oppenheim and Yanofsky, 1980). Although it is noted that this sequence of complementarity is not located in the same vicinity in the 18S rRNA sequence used by influenza B and caliciviruses for its TURBS Motif 1 sequence. The interaction may still allow the ribosome to be tethered in the correct position allowing reinitiation to occur.

Another fungus infecting virus associated with the picorna-like super family has been documented as using coupled translation. *Cryphonectria hypovirus 1* (CHV1), a double stranded RNA genome virus that infects chestnut blight fungus, has been found to use the coupled translation mechanism to translate its second ORF. The viral genome contains two ORFs. The first ORF contains the p69 polyprotein, which is further cleaved into the p29 and p40 proteins. The second ORF contains another large polyprotein, which contains the multifunctional protein p48 (an RdRP and a RNA helicase) (Guo et al., 2009).

The two ORFs overlap by one nucleotide UAAUG, which is also an identical overlap to the coupled gene of BM2 in influenza B. Studies have shown that the translation of the first ORF is essential for the translation of the second ORF and the terminating stop codon of the first ORF is also required. Removing or moving the terminating stop codon further upstream as to prevent or cause premature termination of translation of the first ORF prevents translation of the second ORF. Similar to other viruses that use this mechanism the levels of the second ORF translated are about 4-5% compared to the levels of the first ORF. Upstream sequences found in the p40 coding sequence in the first ORF have been found to also play a role in translation, although the precise role is not known. It is possible that there may be structural

elements in this region as closely related viruses have structural regions in the p40 ORF which have sequences complementary to the 18S rRNA. These regions are however a significant distance from the h26 sequence that binds to TURBS sequences at the 3′ end of the 18S rRNA. This sequence at the 3′ end of the 18S rRNA is not conserved in this host organism, which may explain why no complementarity to the viral mRNA is found here. It is still not known how this complementary sequence to the 18S rRNA aids in the mechanism. Finally, it has also been suggested that the actual viral protein of p40 may also be involved. Further mutational studies in regions that are of great distance from the overlap caused a reduction in translation efficiency suggesting that the protein p40 itself may be involved in translation. This distance is considered too far from the ORF-2 initiation site to influence the mechanism of coupled translation. Therefore it is expected that these mutations disrupt the function of the p40 protein which may aid in translation of ORF-2 (Guo et al., 2009).

1.6.1. Other non-canonical mechanisms for initiation of translation

The mechanism of coupled translation for the initiation of M2 ORF-2 has been well documented in members of the *Pneumoviridae* family, which includes both the *Orthopneumovirus* (hRSV (Ahmadian et al., 2000), PVM (Gould and Easton, 2007)) and the *Metapneumovirus* (APV (Gould and Easton, 2007)) genera. These data would suggest that as the sequence of M2 ORF-1 differs in members that use the mechanism coupled translation; it is feasible that the mechanism of coupled translation is used for the initiation of M2 ORF-2 by all members of this family.

There is however evidence from a single report, that another member of the *Metapneumovirus* genus (human metapneumovirus) does not use this mechanism. The human metapneumovirus (HMPV) M2 ORF-2 is located on the bicistronic M2 gene transcript and similar to APV, contains the two ORFs of M2 ORF-1 and M2 ORF-2. These two ORFs overlap each other by 53nts. The mechanism of initiation of translation of M2 ORF-2 in the second ORF is yet to be identified, however it has been reported that translation of M2 ORF-2 is not reliant on the translation of the first ORF M2 ORF-1. In this study, plasmids for reverse genetics for the M2 gene were created, where full M2 ORF-1 translation was inhibited through the introduction of premature terminating stop codons for M2 ORF-1 and the removal of the M2 ORF-1 start codon. The inhibition of translation of M2 ORF-1 did not affect translation of M2 ORF-2 (Buchholz et al., 2005). In members of this family that use the mechanism of coupled translation, inhibition of translation of M2 ORF-1 would lead to inhibited translation of M2 ORF-2 (Ahmadian et al., 2000; Gould and Easton, 2007). This is a prerequisite of coupled translation, where successful translation of the first ORF is required (Buchholz et al., 2005). As it is unlikely that a canonical initiation mechanism is not used due to the presence of the first ORF, it is possible that mechanism used for initiation of M2 ORF-2 is another non-canonical mechanism. The data also suggest the possibility that not all viruses in this family use the mechanism of coupled translation for the initiation of translation of M2 ORF-2 (Buchholz et al., 2005).

1.6.2. Leaky scanning

Of all alternate non-canonical mechanisms that could be used for initiation of M2 ORF-2 in pneumoviruses with the mechanism uncharacterised, leaky scanning is the closest related mechanism to canonical cap-dependent translation. Leaky scanning, is a cap-dependent mechanism where the scanning 48S ribosome bypasses the first AUG and continues scanning further downstream and initiates translation at an alternate start codon. In some cases, the scanning 48S ribosome may initiate at multiple AUG codons and may also bypass multiple AUG codons to initiate at a downstream site.

The underlying mechanism is connected to the strength of the context of the Kozak sequence surrounding the AUG. As mentioned in Section 1.4.2, initiation is reliant on the scanning ribosome recognising an AUG codon with an optimal context of GCCACCAUGG (Kozak, 1986). With the most vital nucleotide at position -3 from the A in the AUG codon required to be a purine with a slight preference for an A nucleotide. The second most vital nucleotide is at position +4, which is required to be a purine nucleotide with a slight preference for a G nucleotide. The Kozak sequence surrounding these two vital nucleotides as well the AUG codon can vary between different species (Kozak, 1986, 1997). However, the requirement for the nucleotides at position -3 and +4 is consistent for all species (Grzegorski et al., 2014; Nakagawa et al., 2008). This Kozak sequence is required to ensure binding fidelity by the Met-tRNA$^{Met}_{i}$ anticodon to prevent the anticodon binding to other codons. Under some circumstances where this context is weak or non-existent a small proportion of the scanning ribosomes will initiate at this AUG codon. The remainder of the scanning ribosomes are capable of scanning further downstream and will initiate at an AUG with a better context. It is also important to note that there are examples where an AUG with a poor context at the 5′ terminal of the mRNA will be capable of binding to the majority of the scanning 43S ribosomes and initiate translation at that site. This is thought to be due to the presence of highly structured AUGs which may slow the scanning ribosome down sufficiently to bind effectively at AUGs with a poor surrounding Kozak sequence (Kozak, 2002; Lin and Lo, 1992).

Leaky scanning is a relatively common occurrence in viruses that transcribe polycistronic mRNAs and has been reported in examples involving a wide range of

infected host organisms including plants. In cases of leaky scanning used to bypass the first ORF, the context of the Kozak sequence of the first AUG codon may be weak allowing scanning ribosomes to scan past the first AUG. This is used in examples such as initiation for the p17 cistron in segment S1 in Avian reovirus where the consensus sequence for the bypassed AUG codon is UCGAUGC (Racine et al., 2007) or segment 2 of influenza A virus for the initiation of PB1-F2 and N40 cistrons where the consensus sequence for the bypassed AUG codon is UGAAUGG (Wise et al., 2009). Another possibility is that the start codon for the first ORF is a non AUG codon, this is the case in such examples as the pre-genomic RNA of rice tungro bacilliform virus (AUU) (Fütterer et al., 1997) and in the eukaryotic example of *c-myc* (CUG) (Spotts et al., 1997). However in all these cases the first ORF is still translated at significantly lower rates compared to downstream ORFs in polycistronic mRNAs.

There is no defined pattern between different examples of viruses in relationship to the outcome of the scanning ribosome once it passes the first AUG codon. The ribosome is capable of still bypassing multiple AUGs in order reach one with an optimal Kozak sequence such as in the expression of PB1-F2 in segment 2 of the influenza virus, which scans past 3 AUG codons to reach the initiation codon for PB1-F2 (Wise et al., 2009).

The ribosome is also capable of initiating translation at the next available codon such as for the NSs protein that is expressed on viral small mRNA from the Andes hantavirus. The ribosome is capable of scanning past the N initiation codon and initiating translation at the initiation codon for the NSs ORF even if the sequence of the N initiation codon is mutated to have an optimal Kozak sequence (Vera-Otarola et al., 2012).

In polycistronic mRNAs where multiple ORFs are translated through leaky scanning, a proportion of the ribosomes that have bypassed the first AUG codon can initiate at the second ORF. However, the majority of ribosomes will bypass the AUG codon used for initiation of the second ORF and initiate at an AUG codon further along. This again is due to the context of the Kozak sequence. Rice tungro bacilliform virus pregenomic RNA is a polycistronic RNA containing three ORFs (Fig. 1.13). The

first ORF initiates translation through ribosomal shunting, and the second and third ORFs initiate translation through leaky scanning. The first two ORFs have poor start codons (first ORF with a non AUG codon and the second containing neither optimal nucleotide at -3 or +4 UACAUGAGC), however the initiation codon for the third ORF has a much stronger Kozak sequence (AGCAUGA). As result a small proportion of ribosomes initiate translation through leaky scanning at the AUG codon of ORF-2 and the remaining scanning ribosomes will continue scanning and initiate at the AUG codon for ORF-3. There is also one AUG codon of poor context (UUUAUGUGC) that is bypassed by the scanning ribosome between ORF-2 and ORF-3. This is also the longest distance reported where initiation has occurred through leaky scanning at a distance of 894nts (Fütterer et al., 1997). There are also examples where the both the second and third ORFs have equally optimal Kozak sequences. This is observed in segment 2 of the influenza A virus, where the AUG codons PB1-F2 and N40 have equally optimal Kozak sequences (Wise et al., 2009). The reason why a proportion of scanning ribosomes pass by the second ORF is still unknown.

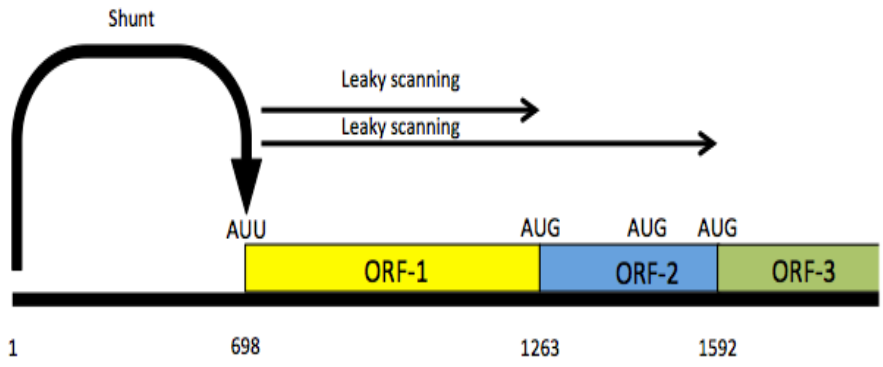


Fig. 1.13: Schematic of the rice tungro bacilliform virus polycistronic pregenomic RNA and mechanisms used for initiation of translation of the three ORFs
Schematic of the rice tungro bacilliform virus polycistronic pregenomic RNA. The black line represents the mRNA strand. Boxes represent ORFs located on the mRNA strand. Ribosomal shunting is used for translation of ORF-1 (yellow). Ribosomal shunting is further described in Section 1.6.4. Leaky scanning is used for translation of ORF-2 (blue) and ORF-3 (green).

The frame of the AUG that the bypassing 43S ribosome will initiate at varies from example to example. For instance in the rice tungro bacilliform virus, all three ORFs are in the same reading frame (Fütterer et al., 1997). However in other cases such as

the pelargonium line pattern virus, the polycistronic sgRNA contains three ORFs, all of which are reliant on initiation of translation through leaky scanning and are all in different reading frames (Castaño et al., 2009). There are also examples of viruses that use a mixture of frames such as segment 2 of the influenza A virus which is mentioned earlier (Wise et al., 2009). This however, is unsurprising as scanning ribosomes are not influenced by frame. There are also cases on bicistronic mRNAs where the two reading frames overlap such as in Andes hantavirus where NSs ORF overlaps the N ORF by 18nts (Fuller et al., 1983; Vera-Otarola et al., 2012). It is also possible for leaky scanning to be used in conjunction with other non-canonical mechanisms. For example in the tricistronic S1 mRNA of avian reovirus, leaky scanning is used for the initiation of translation of the p17 ORF which is the second ORF on the mRNA strand. However the third ORF which is translated to the σC protein, uses the mechanism of leaky scanning for initiation of translation but uses it in conjunction with the mechanism of ribosomal shunting (Racine et al., 2007; Racine and Duncan, 2010).

In the majority of cases mentioned previously, a lower level of expression of the first ORF is observed in relationship to a downstream ORF where the preferential ribosome initiation site is. This however is not the case in all examples of leaky scanning. For instance in VSV, two truncated versions of the Matrix or M protein (33 and 51 amino acids away from the methionine) are translated by leaky scanning. These two truncated versions of the protein are produced in lower quantities compared to the first ORF. The context of the Kozak sequence surrounding the matrix protein AUG codon is suitably strong (AUCAUGA). It is unknown what causes a small proportion of the ribosomes to bypass the initial AUG even though it has a strong Kozak sequence (Jayakar and Whitt, 2002).

1.6.3 Programmed ribosomal frameshifting

Programmed ribosomal frameshifting (PRF) is a mechanism used by the same elongating ribosome to translate two separate ORFs in different reading frames to generate a fusion protein. In this process the elongating ribosome translating the first ORF will shift the frame in which the mRNA is being read without terminating or reinitiating translation. This shift will cause the elongating ribosome to translate the second ORF without disrupting the process of elongation and as a result two separate

proteins are translated in the form of a fusion protein which may later be cleaved to two separate proteins.

In the vast majority of examples of PRF, there is a shift by the elongating ribosome to the -1 frame in relationship to the frame of the first ORF. There are however examples where PRF occurs in +1/-2 direction, these examples are less common and the mechanism is less well defined. PRF has been most commonly documented as a mechanism of translation of viral mRNAs. There also been documented cases in both eukaryotic (Dinman, 2012) and prokaryotic organisms (Flower and McHenry, 1990). This also includes the mechanism being used in mammalian mRNAs (Dinman, 2012).

For programmed -1 ribosomal frameshifting, all examples documented are reliant on two factors for successful frameshifting of the ribosome. Firstly the site of the frameshift must be in a specific location and within a certain sequence called a slippery sequence. Secondly, this slippery sequence must be in close vicinity to a structured region downstream of the slippery sequence (Caliskan et al., 2015). The slippery sequence can vary in sequence, but must have the same format of XXXYYYZ where X can be any nucleotide, Y can be either A or U and Z can be any nucleotide except for G. As seen in Fig. 1.14, for the elongating ribosome translating the first ORF, the slippery sequence must be in the format X_XXY_YYZ (where _signifies the break between each codon). Therefore a shift in the -1 frame would lead to this slippery sequence being read as XXX_YYY_Z (Howard et al., 2004; Plant and Dinman, 2006; Yu et al., 2014).

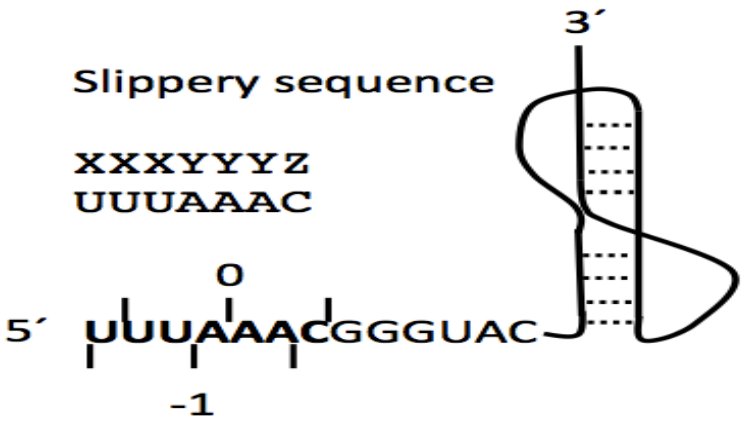


Fig. 1.14: Sequence of slippery sequence used for the mechanism of programmed ribosomal frameshifting used by infectious bronchitis virus (IBV) for the initiation of translation of the 1a/1b polyprotein ORF.
Figure displays the slippery sequence used by IBV for programmed -1 ribosomal frameshifting used for initiation of translation of the 1a/1b polyprotein. Figure also displays the critical secondary structure for PRF located 6 nts downstream of the slippery sequence (Brierley et al., 1989).

The secondary structure varies between examples, but is always within several nucleotides of the slippery sequence. These structures tend to form pseudoknots where the stem loop structures bind with distant sequences downstream leading to the formation of large loops between stems (Namy et al., 2006). The exact process of how the slippage occurs is still debated.

There are currently three suggested models of how the slippage occurs based on the point in time during elongation when the slippage occurs. These three models are all reliant on GTP hydrolysis. The first model suggests that the slippage occurs when the XXY codon is in the process of being translocated to the P site and the energy required for frameshifting is derived from the eEF2 GTP hydrolysis. Upon the ribosome reaching the structured region, a pause in translocation is caused. This pause causes the ribosome to slip by one nucleotide in the -1 direction so the P site instead of having XXY in the codon, now has XXX. The anticodon of the sequence can still bind to positions 1 and 2 in the codon and the third nucleotide in the codon may bind putatively due to wobble base pairing (Namy et al., 2006).

Another model of slippage has been suggested to occur during the delivery and binding of the aminoacyl-tRNA to the A site but before the peptidyl transfer. The energy for this reaction comes from the hydrolysis of GTP from eEF1A. The model suggests that before the accommodation of aminoacyl-tRNA and during

translocation, the structural element causes a significant level of pressure for the ribosome during translocation. Once the aminoacyl-tRNA binds in the A site, the pressure is relieved by the release of tRNA from the mRNA and the movement of the ribosome in the -1 direction. This again leads to wobble base pairing for position 3 in both codons due to the nature of the slippery sequence (Plant et al., 2003).

The third model is similar to the second where the pressure is built before translocation but before peptidyl transfer to the A site. However this model is thought to occur when the ribosome contains tRNAs in all three sites. When the slippage occurs, the XXY in the P site changes to XXX and the A codon changes to YYY (where N is any nucleotide). The energy is also derived from the hydrolysis of GTP from eEF2 (Léger et al., 2007). The debate continues which model is used, however there is growing evidence that the second model is used for (-1) frameshifting.

The first documented case of -1 frameshifting was reported in rous sarcoma virus for the translation of the Gag-Pol polyprotein located on the unspliced viral mRNA (Jacks et al., 1988a; Jacks and Varmus, 1985). Frameshifting is also used in a similar mechanism for the translation of the human immunodeficiency virus (HIV) Gag-Pol polyprotein (Jacks et al., 1988b). For both viruses, the unspliced viral mRNA contains two ORFs, with the Gag polyprotein translated from the first ORF and the Pol polyprotein translated from the second ORF (Jacks et al., 1988b). The Gag-Pol fusion polyprotein once translated is cleaved into the respective proteins found in each ORF. For HIV, the levels of translation of Gag-Pol polyprotein are not produced in equal amounts compared to the levels of Gag polyprotein translated. Frameshifting only occurs in a small proportion of cases (5% (Jacks et al., 1988b)). The majority of ribosomes will continue past the slippery sequence and terminate translation at the Gag ORF terminating stop codon (Biswas et al., 2004; Jacks et al., 1988b; Nikolaitchik and Hu, 2014). Many retro viruses and positive-sense RNA genome viruses use frameshifting for the expression of the viruses' RdRP. In positive-sense single stranded RNA viruses such as coronaviruses (Fig. 1.14), programmed -1 ribosomal frameshifting is used for the translation of the ORF1b polyprotein. This mechanism allows the virus to tightly control the levels of the replicase protein produced (Brierley et al., 1987).

It has been suggested that the downstream structured regions used for PRF in viruses, play a key role in controlling the levels of frameshifting. It is also of note that these structures differ in different members of the same family that use PRF. They do however lead to roughly similar levels of frameshifting (Plant et al., 2010; Thiel et al., 2003).

There have also been examples of frameshifting in negative-sense single stranded RNA genome viruses such as the measles virus (a member of *Paramyxoviridae* family). In measles virus, frameshifting occurs in the P gene transcript for the synthesis of the R protein. As seen in Fig. 1.15A, the P polycistronic gene transcript contains 4 ORFs (P, C, V and R). The R ORF is translated to the R protein which is a fusion product consisting of 294 amino acids (aa) translated from the P ORF and 5 aa translated from the V ORF. The site of the frameshift occurs in between the overlapping region between the P and V ORFs. The -1 shift during translation of the P ORF results in the elongating ribosome translating the final 18nts of the V ORF and terminating translation at the V terminating stop codon. Evidence suggests that the slippery sequence may slightly differ to the consensus sequence (Fig. 1.15B). It is also ambiguous whether a structural region is involved. The study's conclusions were unclear whether a structural region downstream of this alleged slippery sequence was present due to the limitations in modelling software (Liston and Briedis, 1995).

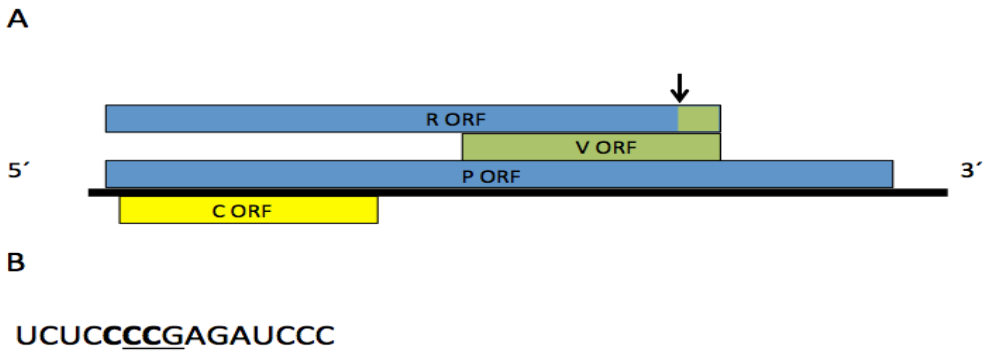


Fig. 1.15: Structure of the polycistronic P gene transcript in measles virus and site of frameshift for the R protein
(A) Structure of the polycistronic P gene transcript in measles virus containing the P (blue), C (yellow), V (green) and R (blue/green). The black line represents the mRNA strand. Boxes represent ORFs located on the mRNA strand. All three ORFs are in separate reading frames. The R ORF (blue and green) is displayed in two separate colours to represent the combined two frames which originate from the P and V ORFs. The site of the frameshift is shown through an arrow. (B) Site of -1 slippage for the R protein. The codon in the first ORF is underlined and codon after the slippage is highlighted in bold

The less well defined example of programmed +1 ribosomal frameshifting has been observed in both eukaryotes and prokaryotes. In eukaryotes, the yeast gene *OAZ1* and its mammalian homologue have been identified as using +1 frameshifting. In both eukaryotes and prokaryotes the slippery sequence used in +1 frameshifting appears to be case specific. In eukaryotes the mechanism is still reliant on a highly structured region downstream of the slippery sequence. The slippery sequence, for example in *OAZ1*, is UCCUGAU with the UGA codon located in the A site during frameshifting. A feedback loop is thought to also control whether frameshifting occurs (Ivanov and Atkins, 2007; Matsufuji et al., 1996; Palanimurugan et al., 2004). There have also been examples of +1 frameshifting in viruses. Lake Sinai viruses 1 and 2 use +1 frameshifting for the translation of the RdRP (Olivier et al., 2008). +1 frameshifting has also been identified as the mechanism to translate the fusion protein PA-X in influenza A. The fusion protein is thought to be used for host gene shut off. The slippery sequence again differs with the sequence for influenza A being UCCUUUCGU (Firth et al., 2012). There have also been examples of -2 frameshifting in viruses, such as the trichomonas vaginalis virus 1 using frameshifting for the production Pol protein (Ai, 2016; Goodman et al., 2011).

1.6.4. Ribosomal Shunting

Ribosomal shunting is a cap-dependent mechanism that is partly reliant on scanning. The premise of ribosomal shunting is the 40S ribosomal subunit bypassing (through a jump) large structured elements typically in the 5′ UTR and resuming scanning post jump (Fig. 1.16). The ribosome is thought to achieve this by retaining certain initiation factors pre-jump through the translation of a short ORF (sORF) before reaching the structure, which enables the ribosome to jump/bypass the secondary structure and bind to the mRNA on the other end of the structure and continue scanning. In general it is thought that shunting is used to bypass long 5′ UTRs that contain multiple AUGs, which in some cases have as good a context Kozak sequence as the AUG used for translation of the desired ORF. The majority of ribosomal shunting examples have been characterised in viruses and only a small number of examples have been reported in eukaryotic organisms (Ogawa, 2013).

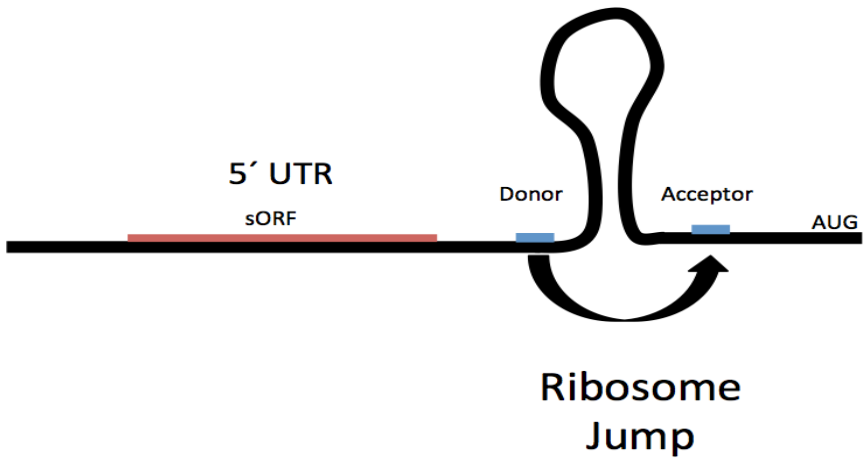


Fig. 1.16: Schematic of the site and mechanism of ribosomal shunting
Schematic of the site and mechanism of ribosomal shunting. Ribosomal shunting will initially translate a sORF (red) in the 5′ UTR. Upon termination, the ribosome will retain some of it's initiation factors and will bypass a large structured region of sequence by jumping from the donor site to the acceptor site (blue). The ribosome will then continue scanning until it reaches the AUG codon.

It appears that throughout all examples of shunting, the exact mechanics behind the mechanism differ significantly from one example to the other. One of the first and most well characterized examples of the mechanism of ribosomal shunting is from the cauliflower mosaic virus, which is a member of the *Caulimoviridae* family. The mechanism is used to bypass a large 5′UTR on the 35S mRNA and translate ORF VII. In this example, the scanning 43S ribosome will initiate translation at a short

sORF located in the 5′ UTR, which is in the vicinity of a large stem loop structure. Upon termination the vast majority of ribosomes will dissociate from the mRNA. However in a small number of cases, some of the 40S ribosomal subunits retain some key initiation factors that do not include any factors that can act as a RNA helicase. The ribosome will move several nucleotides upstream to the beginning of the stem loop structure called the donor site. The ribosome however, due to not having the ability to unwind the stem loop, will instead jump/bypass the structured region (around 480nts in length), where it will bind to the mRNA acceptor site and continue scanning downstream and initiate translation of ORF VII (Ryabova and Hohn, 2000). This has also been documented in the rice tungro bacilliform virus (RTBV), another member of the *Caulimoviridae* family. This virus also uses a similar mechanism to bypass a large stem loop structure. Although both the stem loop and translated sORF differ from the cauliflower mosaic virus example. In both of these cases the acceptor and donor sites where the ribosome jumps and lands are found within the vicinity of the stem loop structure (Pooggin et al., 2012; Pooggin et al., 2006).

Other viruses use shunting to access ORFs using complementary sequences to the 18S rRNA. Adenovirus uses ribosomal shunting to translate ORFs in late mRNAs. During this point in the infection eIF4E is dephosphorylated which prevents the initiation eIF4F complex from efficiently binding to the m^7G cap. As a result the majority of capped mRNAs are not translated as ribosomes are unable to scan through the 5′ UTR and initiate translation due to lack of the eIF4F complex containing an RNA helicase (eIF4A) in the scanning ribosomal complex. Ribosomal shunting is therefore employed by the virus to translate ORFs during this point of infection. This shunt allows the ribosome to bypass the 5′ UTR which is highly structured and would otherwise be unable to be bypassed. Analysis of the 5′ UTR has revealed that shunting is directed by three conserved regions that are complementary to the 3′ region of the 18S rRNA. These complementary sequences are separated from each other by ~30nts and are labelled C1 to C3. The three regions are thought to bind to a stem loop structure found at the 3′ end of the 18S rRNA in 40S subunits, which in turn allows the recruitment of initiation factors. The precise mechanism by which the C1 to C3 regions direct shunting is unknown (Yueh and Schneider, 2000). However, unlike members who use shunting in the *Caulimoviridae*

family, the adenovirus does not rely on the translation of a sORF. This shunting mechanism for translation has also been reported in the *hsp70-1* gene transcript in human cells (Chappell et al., 2006).

Another mechanism of shunting has been reported in the translation of the Sendai virus Y1 and Y2 proteins. The two ORF belong to a polycistronic mRNA, in which multiple mechanisms are used to translate each of the residing ORFs. Shunting has been identified as the mechanism of translation for two of these ORFs (Fig. 1.17A). This shunting mechanism differs from others as no donor site has been identified. As seen in Fig. 1.17B, it was reported that critical sequences directly upstream of the Y2 initiation site may act as an acceptor site for both Y1 and Y2 ORFs. It is thought that this acceptor site is presented to the ribosome inside a secondary structure (de Breyne et al., 2003; Latorre et al., 1998).

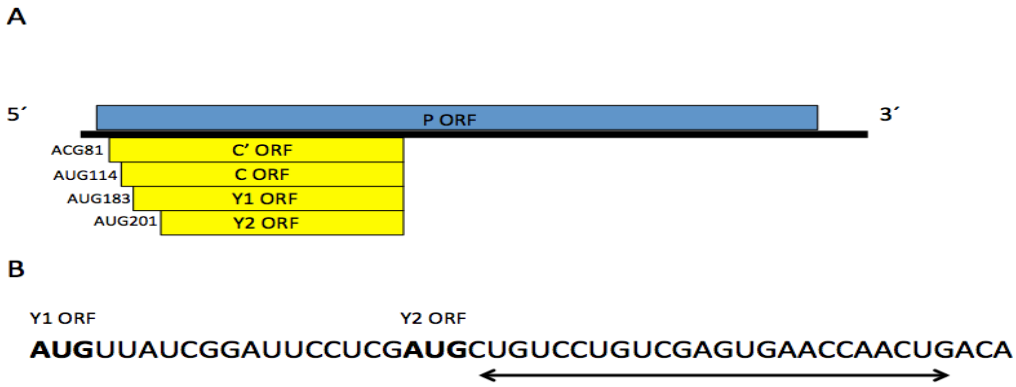


Fig. 1.17: Structure of the polycistronic P gene transcript of Sendai virus and site of initiation for the Y1 and Y2 proteins
(A) Structure of the polycistronic P gene transcript in Sendai virus containing the P, C', C, Y1 and Y2 ORFs. The black line represents the mRNA strand. Boxes represent ORFs located on the mRNA strand. The colour of the ORF represents the reading frame of the ORF (blue +1 and yellow +2). (B) Initiation sites (in bold) for translation of Y1 and Y2 ORFs. The arrow below identifies sequence critical for the mechanism of ribosomal shunting that is thought to act as an acceptor site.

Shunting can be used in conjunction with other mechanisms to initiate translation of ORFs on a polycistronic mRNA. As mentioned in Section 1.6.2, avian reovirus uses leaky scanning for the translation of all three ORFs in the S1 mRNA. However, the σC ORF also uses a shunting mechanism as well as leaky scanning (70/30 leaky scanning/shunting). The S1 mRNA has a short 5´ UTR, which is highly unusual for a virus using the shunting mechanism. There is no room for sORF to be translated as

described in the cauliflower mosaic virus. It has been suggested that the shunting occurs after the termination of one of the upstream ORFs to that of σC. As only a small number of proteins are present in the virus due to leaky scanning being used, the efficiency of this shunting mechanism can afford not to be highly efficient (Racine and Duncan, 2010).

1.6.5. Internal Ribosome entry site

Internal ribosome entry sites (IRES) are capable of initiating translation in a cap-independent manner in almost any location along an mRNA strand. These IRESs are either capable of assembling the 80S ribosomal complex and initiating translation with few or no initiation factors by assembling the 80S complex entirely through interactions with the mRNA. They tend to be used in conditions where the m^7G cap is not accessible/doesn't exist or in conditions where initiation factors are not available. IRESs are highly structured regions of mRNA either located at the 5′ UTR or found within cistrons. These highly structured regions tend to form high affinity complexes with the ribosomal subunits. In turn manipulating the ribosomal subunits in order to assemble the 80S subunit and initiate translation in the vicinity of the IRES, usually at an AUG codon. IRESs tend to have an incredibly high degree of variability with very little sequence or structural homology between different examples (Balvay et al., 2009). First discovered in picornaviruses (Jang et al., 1988; Pelletier and Sonenberg, 1988), IRESs are primarily found as examples of internal cap-independent initiation in viral mRNAs. However, there are an increasing numbers of examples being discovered in eukaryotic mRNAs (Spriggs et al., 2009). IRESs as initiation systems can be easily adopted to be used by other mRNAs and placed in alternate examples of cellular environments and still successfully initiate translation.

Attempts to classify viral IRES have resulted in a large quantity of well defined IRESs being divided into 4 distinctive groups. However there are many other examples that do not fit into these four groups, which is testament to the high diversity of IRESs found on viral mRNAs alone (Kieft, 2008).

Group 1 are the simplest group both in size and mechanism. These IRESs do not require any eIFs to assemble the 80S ribosomal complex and instead assemble the

ribosomal subunit through mRNA interaction alone. They also tend to be smallest in size of the four major groups, with one of the two cricket paralysis virus (CrPV) IRESs only being 190nts in size. The majority of the viruses that use group 1 IRESs belong to the *Dicistroviridae* family. This includes the mentioned CrPV (Wilson et al., 2000b), plautia stali intestine virus (Sasaki and Nakashima, 1999), Israeli acute paralysis virus (Ren et al., 2012) and the shrimp tuara syndrome virus (Mari et al., 2002). With the most well characterised IRES in this family being the CrPV IRES used for initiation of ORF-2. This IRES is located in an intergenic region between two cistrons on a bicistronic mRNA. The IRES directs initiation of translation at a non-AUG codon. For CrPV, the IRES consists of three pseudoknots. Initially the sole mechanism thought to be used by CrPV IRES involved the pseudoknots binding to the 40S ribosomal subunit, resulting in the recruitment of the 60S ribosome for the formation of the 80S ribosomal subunit (Wilson et al., 2000a). There is however evidence that the IRES can also directly recruit the 80S subunit and multiple mechanisms are thought to be used by the IRES in order to maximise translational efficiency (Petrov et al., 2016). As the initiation site is not an AUG, the first tRNA that is delivered to the ribosome is by eEF1A. It is believed that one of the pseudoknots, by forming a similar confirmation to tRNA, binds in the A site and then is moved to the P site by eEF2. This thereby allows the ribosome to form a correct conformational shape reminiscent of the ribosome in elongation phase more similar to a pretranslocation state (Fernández et al., 2014).

Group 2 IRESs are reliant on eIFs to successfully initiate translation. These IRESs are also reliant on an AUG site for initiation. Unlike group 1 IRESs, these viral IRESs originate from more than one family, examples include: classical swine fever virus (Rijnbrand et al., 1997), porcine teschovirus 1 (Pisarev et al., 2004) and the most well characterised member of this group, hepatitis C virus (HCV) (Tsukiyama-Kohara et al., 1992). For the HCV, the IRES is located in the 5′ UTR of the positive-sense RNA genome. The IRES is responsible for initiating translation of the entire polycistronic RNA which is translated into a polyprotein that is later cleaved into the individual HCV proteins. Compared to the structure of the CrPV, the structure is more complex. The structure is 340 nts in size and comprises of 4 domains of which II,III and IV independently fold. The mechanism used by the HCV IRES for initiation of translation occurs over several steps. The first step is the recruitment of

the 40S complex to the IRES. The 40S complex is bound to domain III of the IRES which mediates further interactions with other domains. The second step is recruitment of eIF3 and the ternary complex, which also binds to domain III. Certain eIF3 subunits are involved in aiding the process of 80S ribosomal assembly. The AUG codon used for initiation is located on domain IV is thought to fold into the 40S decoding groove, thereby allowing binding with Met-tRNA$^{Met}_{i}$ in the P site. The final step is the association of the 60S subunit to form the 80S subunit, this is mediated by eIF5 and eIF5B to hydrolyse the GTP on eIF2, which leads to the release of eIF2-GDP and the joining of the 60S subunit to the 40S complex. This is orchestrated by the domain II in the IRES. eEF1A will then bring the aminoacyl-tRNA into the A site to initiate elongation. Under certain conditions this initiation process may vary in examples such as ion concentration variation or inactivation of eIFs. The IRES is still capable of initiating translation even when certain initiation factors are not available or if the Mg^{2+} ion concentration changes from optimal levels leading to change in structure (Jaafar et al., 2016; Pérard et al., 2013).

Groups 3 and 4 are derived from the *Picornaviridae* family. This family has multiple members that use IRESs for cap-independent translation. Although this family has members belonging to both groups 3 and 4, all IRESs belonging to this family have also been separately classified into five separate types. Types 1 and 2 in this family can also be classified into Groups 4 and 3 respectively. Types 3 to 5 are not classified into the 4 IRES groups but do share similar characteristics to both group 3 and 4 IRESs (Martínez-Salas et al., 2015).

Group 3 and 4 IRESs are also reliant on eIFs, the ternary complex and other cellular cofactors called IRES-transacting factors (ITAFs) for successful initiation of translation (Pestova et al., 1996). For group 3 IRESs, the AUG codon is located in the direct vicinity of the IRES and the 40S ribosomal complex is loaded directly onto the AUG in a Yn-Xm-AUG motif (Kolupaeva et al., 2003). This group comprises of such example as the encephalomyocarditis virus (EMCV) IRES (Ghattas et al., 1991) and the foot and mouth disease virus (FMDV) IRES from the *Picornaviridae* family (Martínez-Azorín et al., 2008). In group 4, the 40S complex is loaded onto the same motif, however the ribosome will not use the AUG as the initiation site and instead will use an AUG 150 nts downstream of the IRES. The ribosome may scan or shunt

itself over 100-15 0nts to the initiation codon. Group 4 IRESs comprise of such examples as the polio virus IRES (Pelletier and Sonenberg, 1988)

Almost the entire group of eIFs is required for successful initiation of both groups 3 and 4 IRESs with the exception of eIF4E. The IRESs of both groups are reliant on other cellular cofactor named ITAFs. These tend to be nuclear based proteins that migrate to cytoplasm, which include polypyrimidine-tract binding protein PTB. There is also evidence of interactions between the poly-A tail and the IRES during initiation (Sweeney et al., 2014).

Although this review covers the 4 major group of IRESs, it is important to note that there are constantly new IRESs being discovered that do not fit into the 4 groups of IRESs. There are also multiple examples in eukaryotic mRNAs that are not mentioned in this review.

1.7. Aims

The aim of this project was to identify and investigate the use of regulatory mechanisms used by the *Pneumoviridae* family to control viral gene expression. The technique of ribosomal profiling was used to create transcriptional and translational profiles of both the hRSV and bRSV viral genomes. Transcriptional profiles were used to identify if the levels of viral mRNA abundance followed the levels predicted by the polar gradient model and therefore identify if this model primarily controlled viral gene expression. Translational profiles were used to investigate whether any viral mRNAs other than the M2 mRNA used translational regulation.

The only translational regulation mechanism identified in this family is for initiation of translation of M2 ORF-2, which has been identified in several members of this family. This mechanism is currently uncharacterised in bRSV. Therefore reporter plasmid assays were used to investigate the mechanism used for translation of bRSV M2 ORF-2. Reporter plasmid assays were also used to further characterise the regulatory mechanism of coupled translation termination/initiation used for initiation of translation of PVM M2 ORF-2.

Chapter 2
Materials and Methods

2.1. Tissue culture and virus propagation

2.1.1. Tissue culture media

All cells lines were cultured in Glasgow Modification of Eagle's Medium (GMEM). All media contained 0.2 mM of L-glutamine, 100 μg/ml of penicillin and streptomycin and 10% (v/v) foetal bovine serum (FBS). Under conditions where virus was used which included plaque assays, transfections and growth of viral stocks, GMEM media would contain 2% FBS instead of 10%. For the remaining chapter the percentage at the beginning of GMEM refers to the concentration (v/v) of FBS supplemented in the media.

2.1.2. Phosphate buffered saline (PBS)

2.5mM KCl, 0.15 M NaCl 1.5 mM KH_2PO_4, 6.5 mM Na_2HPO_4, pH7.4

2.1.3. Cell lines

Human epithelial type 2 (HEp-2) cells, epithelial cells of African green monkey origin (BSC-1) cells and Madin-Darby bovine kidney (MDBK) cells were all acquired from Dr Phillip Gould and Prof. Andrew Easton.

2.1.4. Maintenance of cell lines

All cell lines were grown at 37ºC at 5% CO_2 in Falcon^{TM} T-175 175 cm^2 flasks. Cells were grown in 50 ml of 10% GMEM in T-175 flasks. Cells typically doubled in number between every 12 and 24 hours. Once cells had reached 100% confluency, cells would be passaged to reduce the confluency of the cells. The 10% GMEM media would be removed from the flask and versene (0.02% (w/v) EDTA and 0.002% (w/v) phenol red in PBS, pH 7.4) was used to wash the cells. The cells would be disassociated by adding 5 ml of trypsin EDTA (0.02% (w/v) EDTA and 0.25% (w/v) trypsin in PBS, pH 7.4) and incubating at room temperature until all cells had disassociated from the flask's surface. 5 ml of 10% GMEM would be added to the disassociated cells to neutralize the trypsin EDTA. A proportion of cells would then be removed for the requirements of the experiments and the remaining media containing cells were made up to 50 ml using 10% GMEM. Cells would then be incubated back at 37ºC. The concentration of cells after a passage were never reduced to lower than 12.5% and were never passaged more than 20 times. All media used for cell culture unless otherwise stated was heated to 37ºC before use.

2.1.5. Storage of cell lines in Liquid N_2

All cells were stored in liquid N_2 for long-term storage. Cells destined for storage were passaged so that all cells were disassociated from the flask's surface and pelleted at 2000 rpm for 4 min. The supernatant was removed and cells would be suspended in 40% (v/v) of 10% GMEM, 50% (v/v) FBS and 10% (v/v) Dimethyl sulfoxide (DMSO). The newly suspended cells were placed in cryotubes, the cells were then frozen by storing the cryotubes in isopropanol at -70^{o}C over a period of 24 hours. The frozen cells were then transferred to liquid N_2 for long-term storage.

Cells that were removed from liquid N_2 were thawed at 37^{o}C and added to 10 ml of 10% GMEM. The cells were pelleted at 2000 rpm for 4 minutes and resuspended in 10 ml 10% GMEM. The cells were then seeded into Falcon^{TM} T-75 flasks and left to grow until confluent and then transferred to a T-175 flask as described in Section 2.1.4.

2.1.6. bRSV

Stocks of bRSV (Snook strain) were acquired from Prof. Andrew Easton. This strain was used for isolation of the M2 gene sequence and for ribosomal profiling. Stocks of bRSV were grown in MDBK cells. MDBK cells were grown in NUNC^{TM} Cell Culture Treated TripleFlasks with standard cell culture techniques described in Section 2.1.4. Cells were supplemented with 100 ml 10% GMEM media whilst reaching 100% confluency. Upon cells reaching 100% confluency, the 10% GMEM media was removed and 90 ml of 2% GMEM was added to the flask. Virus was then added at a multiplicity of infection (MOI) of 0.01. Cells were incubated at 31^{o}C and incubated until all cells were showing signs of infection, the time frame of this ranged from 5 to 8 days. Cells were harvested by adding glass beads to the flask and gently rocking the flask until all cells attached to the flask surface were removed. All media was then removed from the flask and was centrifuged at 3200g for 20 minutes. The supernatant was either frozen at -70^{o}C or precipitated using the protocol described in Section 2.1.8. Plaque assays as described in Section 2.1.9 were performed using MDBK cells.

2.1.7. Vaccinia virus

Stocks of the vaccinia virus (VTF7-3) strain were acquired from Prof. Andrew Easton. Stocks were replaced by infecting BSC-1 cells with vaccinia virus at a MOI of 0.01 when cells were 90% confluent. The cells were incubated at 37°C for 4-5 days and harvested once all cells showed extensive cytopathic effect. Cells were harvested by adding glass beads to the flask and gently rocking the flask until all cells attached to the flask surface were removed. All media was then removed from the flask and was centrifuged at 3200g for 20 minutes. The supernatant containing the virus was stored at -70°C. Plaque assays as described in Section 2.1.9 were performed using BSC-1 cells

2.1.8. Concentrating Virus

In order to increase the titre of viral stocks, media containing harvested virus was mixed with 50% PEG 6000 in NTE buffer (w/v) ratio (0.15 M NaCl, 0.05 M Tris-HCl, 0.001 M EDTA pH7.5) to give a final concentration of 10% (v/v) PEG. The mixture was mixed for 90 minutes at 4°C with low stirring. The sample was pelleted by centrifuging at 3200g for 5 minutes at 4°C and then resuspended in the desired volume to acquire the desired titre.

2.1.9. Quantifying virus

Virus titres were quantified using plaque assays. Serial dilutions were made by diluting the virus in PBS in 10 fold increments ranging up 10^{-7}. Cells were grown in 12 well plates in 10% media until they reached 90% confluency. The cell media was then removed and 250µl of a diluted virus was added to each well and incubated with intermediate rocking at 37°C. Two wells were used as repeats for each dilution. The PBS containing virus was removed from the wells after the hour incubation and 1ml of CMC overlay (1% (w/v) carboxymethyl cellulose in 2% GMEM) was added to each well. The cells were be incubated at 37°C until visible plaques formed. The infected cells were fixed using 4% glutaraldehyde solution (4% (v/v) glutaraldehyde in PBS) and left in solution for 2 hours. The plates were then washed and crystal violet solution (0.5% (w/v) crystal violet, 20% (v/v) methanol in H_2O) was added and incubated at room temperature for 40 minutes. The crystal violet solution was washed off the plates and the plates dried. Plaques were then counted and the titre

(Plaque forming units (Pfu/ml of stock virus) calculated from the number of plaques present.

2.2. Protein techniques

SDS-PAGE and western blot analysis of proteins to determine the presence of fusion proteins.

2.2.1. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Prior to western blot analysis, proteins were separated using SDS-PAGE. Each gel comprised of both a running and stacking gel. All running gels were 10% gels which comprised of 2.8 ml H_2O, 3ml 30% acrylamide mix, 4 ml 0.5 M Tris-HCl (pH8.8), 0.15 ml 10% SDS 0.1 ml, 10% ammonium persulfate and 4µl TEMED. Stacking gels comprised of 8.45 ml H_20, 30% acrylamide mix, 3.75 ml 0.5 M Tris-HCl (pH6.8), 0.15 ml 10% SDS, 0.15 ml 10% ammonium persulfate, 4 µl TEMED.

Prior to loading samples on the gel, protein samples were mixed with 4x loading dye, which contained beta-mercaptoethanol. The quantity of lysate required was either dependent on well volume or concentration of protein required for western blot analysis. A PageRuler Prestained protein ladder 10 to 180 kDa (ThermoFisher Scientific) was used to determine protein size. 5 µl of the ladder was added to one well in the gel. Samples were then incubated for 5 minutes at 95ºC to denature the protein. Samples once loaded were run at 125V for 2 hours in 1x running buffer (15 g/l Tris, 72 g/l glycine, 5 g/l SDS pH 8.2 for a 5X concentration).

After the gel had sufficiently separated, the gel was then washed in double-distilled water for 5 minutes. For western blot analysis, preparation of the gel continued as described in Section 2.2.2. For non-specific protein analysis, the gel was stained using Coomassie stain (Bio-Rad). 50 ml of stain was added to the gel and left rocking for 30 minutes. The gel was then washed in destain (20% methanol, 10% acetic acid, 70% H_2O) until bands became visible. A protein ladder was used to identify the size of the protein bands in the gel. Bands were visualized on an illuminator box.

2.2.2. Protein detection using western blot analysis

The gel from Section 2.2.1 once washed was rewashed in transfer buffer (10% 10x tris glycine, 20% methanol, 70% water) for 5 minutes. The gel was then placed on a nitrocellulose membrane and run in a gel electrophoresis tank using transfer buffer for 1 hour at 200 mA. Upon completion of transfer of proteins onto a nitrocellulose membrane, the membrane was then washed in TBS tween (50mM Tris, 150 mM NaCl and 0.05% tween 20) for 5 minutes. The membrane was then immersed in 2% bovine serum albumin (BSA) in TBS tween for 1 hour to act as a blocking agent. The membrane was then incubated with 25ml of primary antibody overnight at 4 ^{o}C. The primary antibody (Goat Anti-CAT IgG Fc Abcam) was diluted 1 in 3000 in PBS solution (1% w/v BSA and 0.08% v/v sodium azide). The membrane was then washed for 1 hour in TBS tween. 10ml of secondary antibody for the specific primary antibody (Rabbit Anti-Goat IgG H&L (HRP) Abcam) was incubated for 2 hours with the membrane. The antibody was dissolved in powdered milk in 5% w/v ratio in TBS tween at a dilution of 1 in 3000. The membrane was then washed for 1 hour in TBS tween. ECL western blotting substrate (Pierce) was added to develop the membrane as described by manufacturers protocol. The membrane was then placed onto X-ray film and kept in a developer cassette for a set period of time. Films were developed in a X-ray film developer machine.

2.2.3. Bicinchoninic acid protein assay

A bicinchoninic acid protein assay (BCA) was performed using a Pierce BCA protein Assay Kit from ThermoFisher Scientific. Assay was performed according to manufactuers instructions.

2.3. Cloning

All constructs were created using the following methods below. For an in depth cloning strategy see relevant results section. Primer sequences can be found in Appendix C.

2.3.1. PCR

For all PCR reactions, the proof reading polymerase *Pfu* DNA polymerase (Promega) was used. A total reaction mixture of 50 µl in 500 µl PCR tubes was used for each reaction. Primers were added to make a final concentration of 0.2pmol/µl.

Primers were acquired from IDT and were resuspended at concentration of 10 pmol/ul. dNTPs were added to make a final concentration of 200 μM for each dNTP. A maximum of 50 ng of template was added. *Pfu* DNA polymerase reaction buffer (Promega) was added to the reaction mixture from a from 10x solution to create a final 1x concentration (200mM Tris-HCl (pH 8.8 at 25°C), 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, 1.0% Triton X-100 and 1mg/ml nuclease-free BSA). 1.25 units of *Pfu* DNA polymerase were used for each reaction. A typical reaction mixture is shown in Table 2.1. Reactions were performed in an Eppendorf Mastercycler Gradient PCR machine. The reaction cycle is shown in Table 2.2. PCR products were purified using a QIAGEN Gel extraction kit (Section 2.3.14) once correct band was identified on a 1% agarose gel (Section 2.3.2).

Reagent	Quantity	Final Concentration
Pfu DNA Polymerase buffer 10x	5 μl	1x
dNTP Solution 10mM of each dNTP	1 μl	200 μM each dNTP
Forward Primer 10pmol/μl	1 μl	0.2 pmol/μl
Reverse Primer 10pmol/μl	1 μl	0.2 pmol/μl
DNA template	≈1 μl	50 ng
Pfu DNA polymerase	1 μl	1.25 units
Water	To total volume of 50μl	NA

Table 2.1: PCR reaction mix used for all PCR reactions
PCR reagent mixture for *Pfu* DNA polymerase.

Temperature	Time	Cycles
95°C	5 minutes	1
95°C	30 seconds	30
58 °C dependent on primer TM	30 seconds	
72 °C	1 minute per kbp	
72 °C	10 minutes	1
4 °C	Indefinite	NA

Table 2.2: PCR reaction conditions used for all PCR reactions
PCR cycle program used for PCR reaction using the *Pfu* DNA polymerase

2.3.2. Agarose gel electrophoresis

Agarose gels were used to quantify and visualize DNA. Gels were made to 1% w/v agarose in TBE buffer (45 mM Tris-Borate, 1 mM EDTA) and 1 µl of Midori Green staining solution was added. Gels were set and run in TBE buffer. DNA samples were mixed in 6x xylene cyanol or 6x bromophenol blue. A HighRanger plus 100 bp DNA ladder was used in all experiments. Gels were run at 120V for 40 minutes and visualized using a UV box.

2.3.3. Quantification of DNA and RNA

DNA and RNA was quantified using a Implen NP80 Nanophotometer. Both DNA and RNA quantification was performed at a wavelength of 260nm. The Nanophotometer was initially blanked with the 1 µl of solution that the nucleic acid was dissolved in. 1 µl of sample was then used to quantify the levels of nucleic acid.

2.3.4. Restriction digest

Restriction digests were performed using FAST DIGEST restriction enzymes from Thermoscientific. 5 units (1 µl) of enzyme was used to digest 1 µg of DNA (quantified as described in Section 2.3.3) in a total reaction mixture of 20 µl. A Fast Digest buffer was added at $1/10^{th}$ the volume of the reaction mixture (10 mM Tris-HCl (pH 7.5 at 37°C), 10 mM $MgCl_2$ 0.1 mg/mL BSA). The remainder of the reaction mixture was made up to 20 µl with nuclease free water. Reaction samples were incubated at 37 °C for 2 hours. For each reaction, a control was used for each restriction endonuclease. For a control reaction, each restriction endonuclease was

used to digest the plasmid pBluescribe. The successful digestion of the plasmid was confirmed on a agarose gel with the plasmid being visually linearized compared to uncut plasmid run on the same gel.

2.3.5. Digestion of insert and vector samples for cloning

Inserts and vectors used in cloning underwent the same restriction digest as described in Section 2.3.4. Digested insert samples were purified using a QIAGEN PCR Purification kit (Section 2.3.15). Digested vectors were purified based on the size of the fragments generated. If the unwanted fragment was under 100nts, then the sample was purified with a QIAGEN PCR Purification kit (Section 2.3.15). However, if the unwanted fragment was over 100 nts in size, then the entire restriction digest mixture was run on a 1% agarose gel (Section 2.3.3). The desired fragment used for the vector was excised from the gel and purified using QIAGEN Gel Extraction Kit (Section 2.3.15). Upon purification, all samples had the levels of DNA quantified using an Implen Nanophotometer.

2.3.6. Dephosphorylation of vector

The digested vector was dephosphorylated in order to prevent self-ligation. 1 unit of Antarctic phosphatase was added to the previous digested reaction mixture. Antarctic phosphatase buffer was added at $1/10^{th}$ the volume of the total reaction mixture (50 mM Bis-Tris-Propane HCl 1 mM $MgCl_2$ 0.1 mM $ZnCl_2$ pH 6.0 at 25°C). The reaction mixture was then incubated at 1 hour at 37 ^{o}C. Samples were then purified using a QIAGEN PCR Purification kit (Section 2.3.15).

2.3.7. Ligation

Digested inserts were mixed at molar ratio of 3:1 with a digested and dephosphorylated vector. Quantities required were calculated using NEB clone ligation calculator. T4 DNA ligase from Thermoscientific was used for the reaction. 5 units (1 µl) of enzyme was added to a total reaction volume of 20 µl. T4 DNA ligase buffer was added at $1/10^{th}$ the volume of the total reaction mixture (400 mM Tris-HCl, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP (pH 7.8 at 25°C)). The remaining volume comprised of digested vector and insert DNA. Reaction mixtures were incubated for 1 hour at 37^{o}C. Samples were then directly transformed into DH5α competent bacteria.

2.3.8. Transformation

Transformations were performed using DH5α competent bacteria. For transformation of ligations, ligation mixture (never exceeding 100 ng of DNA) equalling $1/10^{th}$ the volume of competent bacterial cells were mixed with the bacteria on ice. The mixture was then incubated on ice for 30 minutes. The transformation mixture was then heat shocked at 42^{o}C for 45 seconds. The mixture was then placed on ice and incubated for another 2 minutes. All plasmids used contained an ampicillin (AMP) resistance gene. The competent bacteria were therefore incubated at 37^{o}C in 10x volume of L-broth (LB) broth for 45 minutes in order for the bacteria to express the ampicillin resistant genes. After the 45 minutes incubation, bacterial cells was plated on ampicillin LB agar plates (ampicillin at a concentration of 100 μg/ml). For certain ligation mixtures requiring blue white selection, ampicillin LB agar plates was used that also contained Isopropyl β-D-1-thiogalactopyranoside (IPTG) and X-Gal (X-Gal 40 μg/ml, IPTG 100 μg/ml and ampicillin 100 μg/ml). Plated colonies were incubated for 48 hours at 30^{o}C before being picked.

2.3.9. Colony selection

White colonies were selected for bacteria that had been plated on LB agar plates with IPTG and X-Gal. This indicated that the *lacz* gene in the transformed plasmid had been disrupted indicating successful ligation. No disruption of the *lacz* gene would lead to a distinctive blue colour in colonies. Multiple colonies were picked if ampicillin LB agar plates were used. Selected individual colonies were picked into 5 ml of LB broth containing ampicillin (concentration 100 μg/ml) and incubated at 37^{o}C overnight. Plasmids from bacterial colonies were purified using a QIAGEN Plasmid Miniprep Kit (Section 2.3.12).

2.3.10. Sequencing of DNA

For sequencing of plasmids, up to 500 ng of plasmid was added to a total reaction mixture of 10 μl in a 1.5 ml microcentrifuge tube. The correct sequencing primer was used at a final concentration of 2.5 pmol/μl. Samples were then sent to GATC sequencing for sequence analysis.

2.3.11. Glycerol stocks

Successfully cloned plasmids were stored at -70^{o}C in glycerol stocks. Plasmids were transformed into DH5α competent bacteria, plated and grown overnight at 37^{o}C on ampicillin LB agar plates (Section 2.3.8). Colonies were picked into 5 ml LB with ampicillin at a concentration of 100 μg/ml and grown overnight at 37^{o}C. 0.5 ml of the bacterial culture was taken and mixed with 50% glycerol solution (50% (v/v) glycerol in H_2O) at a (v/v) ratio of 50%. The solution was stored in cryotubes and stored at -70^{o}C. Stored plasmids in glycerol stocks could then be plated on ampicillin LB agar plates when required.

2.3.12. Increasing stocks of plasmid

To increase stocks of plasmid, colonies were picked from plates inoculated with glycerol stocks of the respective plasmid required. A colony was picked into 5 ml LB with ampicillin at a concentration of 100 μg/ml and grown overnight at 37^{o}C. Bacterial cultures were then purified using QIAGEN Plasmid Miniprep Kits (Section 2.3.13). Plasmids preps were quantified using an Implen Nanophotometer.

2.3.13. Plasmid purification

Plasmids were purified using a QIAGEN Plasmid Miniprep kit. Purifications were performed to manufacturers instructions.

2.3.14. Gel purification

DNA fragments separated using agarose gel electrophoresis were purified using a QIAGEN Gel Purification Kit. Purifications were performed to manufacturers instructions.

2.3.15. PCR purification

PCR samples were purified using a QIAGEN PCR Purification Kit. Purifications were performed to manufacturers instructions.

2.4. Quantification of Coupled translation

Constructs created to characterise the mechanisms of M2 ORF-2 translation were transfected into cells and the levels of reporter gene expressed was measured, enabling the efficiency of coupled translation to be measured.

2.4.1. Transfections

Lipofection transfections were used to introduce plasmids into cells. HEp-2 cells were used for all transfections. Cells were plated into 12 well plates so at the time of transfection each well would be at 50% confluency. TransIT-LT1 transfection reagent (Mirus) was used as the vector for transfection. For each well, 500 ng of plasmid was added to 1.5 µl of TransIT-LT1 transfection reagent (Mirus) in 100 µl of Opti-MEM (Thermoscientific) and incubated for 30 minutes at room temperature. Initially cell media was aspirated from each well and vaccinia virus (VTF7-3 strain) in 1 ml of 2% GMEM at a MOI of 1 was added. This was to provide a T7 RNA polymerase to the cell to aid in successful transcription of the plasmid. The cells were then incubated with the virus for 1 hour at 37^{o}C. After the hour the viral infected media was aspirated off and the Opti-MEM mixture with the transfection reagent and plasmid was added to each well. 1.5 ml of 10% GMEM was then added to the cells. The cells were then incubated for 48 hours at 37^{o}C and were then harvested. The media was initially aspirated and each well was washed in sterile PBS. The PBS was then aspirated and 300 µl of lysis buffer (Roche CAT ELISA kit) was added to each well. The lysed cells were left for 30 minutes to ensure all cells were completely lysed. Lysed cells were then be removed and centrifuged at 13000rpm for 7 minutes to separate cell debris. The supernatant was frozen at -20 ^{o}C. For each experiment, three separate transfections were performed for each plasmid. The experiment was repeated a total of three times. This meant that a total of three triplicates of transfections were performed for each plasmid tested.

2.4.2. CAT Enzyme-linked immunosorbent assay (ELISA)

A Roche CAT ELISA kit was used to quantify the levels of CAT protein expressed from transfected plasmids in HEp-2 cells. 30 µl of harvested sample was initially mixed with sample buffer from ELISA KITS to make up to a total volume of 200 µl. The mixture was added to ELISA plates and incubated for 1 hour at 37^{o}C. After the incubation the plates were washed in TBS TWEEN four times and the secondary antibody Anti-CAT-DIG was added. The antibody from stock solution was diluted to 1/200 in sample buffer and 200 µl was added to each well. The plate was incubated for another hour at 37^{o}C. The plates were then washed with TBS Tween 4 times and the tertiary antibody was added (Anti-DIG-POD). This antibody was diluted 1/133 in sample buffer from stock solutions and 200 µl was added to each well and incubated

for one hour at 37^{o}C. The plates were then washed 4 times after incubation and 200 µl of peroxidase (POD) substrate would be added and incubated at room temperature for 5 minutes. The ELISA plate was placed on plate reader and read at 405nm wavelength.

2.4.3. GFP ELISA

An Abcam GFP ELISA was used to quantify the levels of eGFP protein expressed from plasmids transfected in HEp-2 cells. Initially cellular lysate harvested from transfected cells was diluted 1 in 100 with PTR buffer and 50 µl of the diluted sample was added to each well. 50 µl of antibody mixture (containing detector and capture antibodies) was added. The ELISA plate was incubated at room temperature for 1 hour on plate shaker set at 450 rpm. After the hour incubation, the wells were washed with 350 µl wash buffer 3 times. 100 µl of TMB substrate was added to each well and the ELISA plate was incubated whilst covered for 3 minutes on a plate shaker set at 400 rpm. 100 µl of Stop solution was then added and the ELISA plate was incubated for 1 minute on a plate shaker set at 400 rpm. The ELISA plate was then placed in a plate reader and was read at 450 nm wavelength. A standard curve was created using GFP recombinant protein using a serial dilution in 3 fold increments from a concentration of 4000 pg/ml to 2.7 pg/ml. Creation of the standard curve followed the same protocol for the GFP ELISA as descirbed above.

2.5. Competent bacteria production

DH5α cells were plated on LB agar plates. Colonies were picked in 5 ml LB broth and incubated at 37^{o}C overnight. Bacteria grown overnight were added to 100 ml of LB broth in a 250 ml conical flask and were grown to an OD_{600} of 0.39 (measured using photometer). Once the correct OD_{600} was reached the flask was placed on ice for 5 minutes. Bacterial colonies were centrifuged at 2000 rpm for 15 minutes at 4^{o}C. The pellet would then be resuspended in 40 ml of Tfb I buffer (30 mM KOAc acetate, 100 mM RbCl, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15% v/v glycerol to a pH to 5.8) and incubated for 1 hour at 4^{o}C for 1 hour. Cells were then centrifuged for 15 minutes at 3000 rpm at 4^{o}C. The pellet was then resuspended in 4 ml of Tfb II (10mM MOPS, 75 mM $CaCl_2$, 10 mM $RbCl_2$, 15% v/v glycerol to a pH to 6.5) and incubated for 5 hours at 4^{o}C. Bacteria in Tfb II were aliquoted in 100 µl quantities and snap frozen in liquid nitrogen and stored at -80^{o}C.

2.6. RNA extraction and reverse transcription (RT)

The M2 RNA of bRSV was isolated from virus using a QIAGEN Viral RNA extraction kit. Reverse transcription was used to create DNA copies of the viral RNA that could be used as a template to amplify the M2 Gene through PCR.

2.6.1. RNA extraction

RNA was extracted from bRSV (Snook strain) using a QIAGEN Viral RNA Extraction Kit. The method was completed according to the manufacturers instructions. $3x10^4$ Plaque forming units (Pfu) of bRSV were used to extract the viral genome.

2.6.2. Reverse transcription

M-MLV reverse transcriptase (Promega) was used for RT amplification of the viral genome using random oligo-nucleotides as primers (500 ng). A maximum reaction volume of 50 µl was used. The maximum quantity of RNA that was capable of being added to the reaction was added. 5 µl of reaction buffer was added to the reaction (250mM Tris-HCl, 15mM $MgCl_2$, 50 mM DTT, 375 mM KCl). dNTPs at a final concentration of 200µM for each dNTP were added. 200 units of M-MLV reverse transcriptase enzyme and 20 units of ribonuclease inhibitor (Promega) were finally added. The reaction mixture was incubated for 60 minutes at 37 ºC. cDNA was amplied as described in Section 2.3.1.

2.7. *In vitro* transcription reaction

A MEGAscript kit from Ambion was used to perform *in vitro* transcription reactions using a T7 RNA polymerase. Amplicons generated by PCR from plasmids used the same method and reaction conditions as described in Section 2.3.1. Specific primers used to amplify each plasmid are mentioned in respective results chapters. Prior to use in the *in vitro* reaction, PCR products were purified using a QIAGEN PCR purification kit as described in Section 2.3.15. For each reaction, 1 µg of purified PCR product was added to each reaction. This was added to a reaction master mix (2 µl ATP solution (75 mM), 2 µl CTP solution (75mM), 2 µl GTP solution (75mM), 2 µl UTP solution (75mM), 2 µl 10X reaction Buffer and 2 µl of enzyme mix) and nuclease free water was used to bring the total volume to 20 µl. The mixture was

incubated at 37°C for 2 hours. Samples were run directly on Novex precast 6% TBE urea-polyacrylamide gels to perform denaturing polyacrylamide gel electrophoresis (PAGE). Samples were mixed with an equal volume Gel loading buffer II (Ambion) and incubated at 90°C for 3 minutes. Samples were run at 100V for 1.5 hours in TBE buffer. Gels were stained with 50ml SYBR Gold and viewed using a UV box. A Geneflow 100bp ladder was used as a size marker and was prepared using the same method as described above.

2.8. Statistics

Standard deviation was calculated for all transfection results using Microsoft excel. Analysing whether there was statistical significant difference in CAT protein expression between two transfected plasmids was confirmed using the Welch's T-test using Microsoft excel.

2.9. Mfold analysis

Sequences that had suspected mRNA secondary structure were inputted into the mRNA secondary structure modelling software Mfold. The constraint information was kept at standard settings: RNA sequence, linear; folding temperature, 37°C; percent suboptimality number, 5; upper bound on the number of computed foldings, 50; window parameter, default; maximum interior/ bulge loop size, 30; maximum asymmetry of an interior/bulge loop, 30 and maximum distance between paired bases, no limit.

2.10. Ribosomal profiling

An Illumina ARTseq (now TruSeq) ribosomal profiling kit (mammalian) was used to prepare and produce sequencing libraries from RNA harvested from MDBK cells infected with bRSV for next generation sequencing using a MiSeq Sequencing System.

2.10.1. Infection of cells and harvesting of cellular lysate containing RNA

Both protected and total libraries were created from cellular lysate harvested from the same cellular population. MDBK cells were grown in 100 mm cell culture dishes as described in Section 2.1.4. Cells were infected with bRSV (Snook strain) at a MOI

of 1 when cells were 90% confluent. Initially media was aspirated off the dish and 5ml of 0% GMEM media containing bRSV was added to the cells. The cells were incubated at 37°C for 1 hour. After the incubation, the media was aspirated and 20ml of 2% GMEM media was added. Cells were either incubated for a further 3 or 7 hours at 37°C. The control cell culture dish (MockP/MockT libraries) followed the same protocol above except without the 0% GMEM media containing virus. The control cell culture dish used to create the control sequencing libraries was also harvested after 3 hours of further incubation. After the incubation period, 5 ml of 2% GMEM media containing cycloheximide was added that brought the concentration of cycloheximide in the cell culture dish to 0.1 mg/ml and was then incubated for 5 minutes at 37°C. After the incubation, the media was aspirated and cells were washed in 10 ml of ice cold PBS that contained cycloheximide at a concentration of 0.1 mg/ml. Cells were then scraped off with 800µl of lysis buffer (20% 5x Mammalian Polysome buffer, 10% Triton X-100, 1% 100mM DTT, 1% DNAase I, 0.4% 25 mg/ml Cycloheximide, 66.6% Nuclease free water). The lysis buffer containing the cellular lysate was passed through 21, 22, 25 gauge needles and was then transferred to a microcentrifuge tube. The cellular lysate was placed on ice and incubated for 10 minutes. The lysate was centrifuged at 20000xg at 4°C for 10 minutes. The supernatant was then transferred to a new fresh tube. Each sample was aliquoted into 200 µl aliquots. For each time point, 200 µl of the cellular lysate was used to create the protected library and the remaining 600 µl was snap frozen in liquid N_2 and stored at -80°C, where 200 µl would be later used to create the total library as described from Section 2.10.2. For the 200 µl of cellular lysate used to create the protected libraries at each time point, 1.5 µl of ARTseq nuclease was added to the mixture. The cellular lysate was then incubated for 45 minutes at room temperature. The reaction was then halted with 15 µl of SUPERase In RNase inhibitor. The preparation process then proceeded onto monosome removal.

2.10.2. Monosome purification

Illustra Microspin S-400 columns were used to purify monosomes from the purified protected mRNA. The S-400 columns were initially equilibrated by resuspending the resin and draining the buffer from the resin. 3 ml of 1X Mammalian Polysome buffer was allowed to pass through the column in 0.5 ml increments. Once all buffer had been drained, the columns were centrifuged at 600 g for 2 minutes. 100 µl of

protected sample was added to each column and the column was centrifuged at 600 g for 2 minutes. This was repeated once more for each sample. The flow through was collected in 1.5 ml microcentrifuge tubes. Both total and protected samples had the RNA purified from the cellular lysate using Phenol Chloroform RNA purification and RNA precipitation as described in Section 2.10.3. RNA was resuspended in 25 µl of nuclease free water. 1 µl of each sample was quantified using a nanophotometer (Section 2.3.3). The samples were then stored on ice ready for rRNA removal.

2.10.3. Phenol chloroform RNA purification and RNA precipitation

RNA samples were mixed with Acid Phenol:Chloroform at a 1:1 ratio. The mixture was vortexed for 30 seconds and then centrifuged at 14000 rpm for 5 minutes at -20^{o}C. The aqueous phase was then removed from the organic phase. This process was repeated two more times. Upon the third extraction the aqueous phase was mixed with $CHCl_3$:Isoamyl Alcohol (24:1) at a 1:1 ratio. The mixture was vortexed for 30 seconds and then centrifuged at 14000 rpm for 5 minutes at -20^{o}C. The aqueous phase was removed and an RNA precipitation was immediately performed.

For an RNA precipitation, each precipitation was performed with 400 µl (if lower, made up with nuclease free water) of each sample. If there was more than 400 µl of sample, multiple reactions were performed. To the 400µl of sample; 700 µl of isopropanol, 40 µl of 5 M ammonium acetate, 2µl of 10% SDS and 2µl of glycogen (20mg/ml) were added. The sample was briefly vortexed and stored at -20^{o}C overnight. The sample was centrifuged at 14000rpm for 20 minutes at -20^{o}C. The supernatant was removed and the pellet was washed twice in 800 µl of ice cold 80% ethanol. For each wash the ethanol and pellet were centrifuged at 14000 rpm for 20 minutes at -20^{o}C. The pellet was then dried and resuspended in nuclease free water.

2.10.4. rRNA removal

Ribo zero magnetic beads were used to remove rRNA from both protected and total RNA samples. 225 µl of magnetic beads were used for each sample. The magnetic beads were initially washed by placing the 225 µl magnetic beads in a 1.5ml microcentrifuge tube and placing the tube on a magnetic stand for 1 minute. The supernatant was removed and 225 µl of RNase free-water was added to the beads and resuspended by vortexing. The beads were then placed on a magnetic stand and left

for 1 minute. The process was repeated again and the beads were resuspended in 65 µl of Magnetic Bead Resuspension Solution. The beads were set aside at room temperature for later use.

For each total and protected sample, up to 5 µg of RNA (measured on a nanophotometer Section 2.3.3) was added to rRNA hybridizing probes in the mixture: Ribo-Zero Reaction Buffer, 4 µl; RNA sample, up to 28 µl; Ribo-Zero Removal Solution, 10 µl: RNase-free water, amount so sample reaches 40 µl. The mixture was incubated at 68^{o}C for 10 minutes and then left to incubate at room temperature for 5 minutes. The sample was then added to magnetic bead mixture and resuspended using a pipette to mix the sample 10 times and then vortexed for 10 seconds. The sample was then incubated at room temperature for 5 minutes and placed on the magnetic stand for a further minute. The supernatant was then removed and placed in a 1.5 ml microcentrifuge tube. The sample was then purified using phenol chloroform and precipitated with isopropanol overnight. The Protected samples were resuspended in 11 µl of RNase free water. Total RNA samples were resuspended in 20 µl of RNase free water. Samples if not used immediately for the next stage were stored at -80^{o}C.

2.10.5. Purification of protected samples

11µl of protected sample was added to 11 µl Gel Loading Buffer II. The ladder for the gel was prepared in a mixture of: 20/100 oligo ladder, 4µl; Nuclease-free water, 1 µl; Gel Loading Buffer II, 5 µl. Both the ladder and protected samples were incubated for 5 minutes at 95^{o}C. The samples were then loaded onto Novex 15% precast TBE urea-polyacrylamide gels and run at 150 V in TBE buffer until the bromophenol blue band had reached the bottom of the gel. The gel was then removed and stained in 50 ml SYBR Gold for 15 minutes. The gel was then placed under a blue light transilluminator. For each protected sample, the area corresponding to 28 to 35 nts was excised, this was performed even if no bands were visible. The excised gel band was cut up using a razor and placed into a 0.2 ml PCR tube with a hole punched in the bottom. This PCR tube was placed into a 0.5 ml microcentrifuge tube which also had a hole punched in the bottom by a needle. The tubes were placed in a 1.5 ml microcentrifuge tube. The tubes and gel were centrifuged at 12000 g for 2 minutes. The 0.5 and 0.2 ml tubes were removed and the following mixture was

added to the gel found in the 1.5 ml micro centrifuge tube: Nuclease-free water, 400 µl; 5 M ammonium acetate, 40 µl; 10% SDS, 2µl. The mixture was rocked at room temperature for 1 hour. A 1 ml pipette tip with its tip cut to increase the bore of the tip, was used to transfer the mixture to filter tubes. The whole sample was centrifuged at 2000 g for 3 minutes. The sample was precipitated as described in Section 2.10.3, samples were resuspended in 20 µl of RNase free water.

2.10.6. Fragmentation of total RNA samples and end repair of total and protected samples

Total RNA samples were fragmented in order to generate correct sized fragments for end repair and sequencing. Both protected and total samples had 7.5 µl PNK Buffer added to 20 µl of each sample. Protected samples were stored on ice and the total RNA samples were incubated at 94^{o}C for 25 minutes. The samples were then placed on ice. To both protected and total RNA samples the following reagents were added to the samples: Nuclease-free water 44.5 µl and PNK 3 µl. The samples were incubated at 37^{o}C for 1 hour. 25 µl of nuclease free water was added to each sample. The samples were then purified using Zymo RNA Clean and Concentrator Kit. The samples were purified according to manufacturers instructions. For step 1, 200 µl of RNA Binding Buffer was used and 450 µl of 100% ethanol was added in step 2. 10 µl of Nuclease Free water was used to elute the samples from the columns. From this point onwards both total and protected samples underwent the same preparation methods.

2.10.7. Ligation of 3´ Adapters

For all samples a ligation master mix was created with the following reagents used for each sample: ARTseq Ligation Buffer, 3.5 µl; 100 mM Dithiothreitol (DTT), 1µl; ARTseq Ligase, 1.5 µl. The master mix was kept on ice. For each of the samples, 1µl of ARTseq 3´ Adapter was added and the sample was incubated at 65^{o}C for 2 minutes and then held at 4^{o}C. 6 µl of master mix was added to each sample. The sample was briefly centrifuged to collate the mixture at the bottom of the tube. The sample was incubated at room temperature for 2 hours. 2 µl of ARTseq AR enzyme was added to the samples and mixed. The samples were incubated at 30^{o}C for 30 minutes.

2.10.8. Reverse transcription of samples

A RT master mix was created with the following reagents for each sample: ARTseq RT Reaction Mix, 4.5 µl; EpiScript RT, 1 µl; 100 mM Dithiothreitol (DTT), 1.5 µl; Nuclease-free water, 6 µl. 13 µl of each reaction mixture was added to both protected and total RNA samples and incubated at 50^{o}C for 30 minutes. 1 µl of ARTseq Exonuclease was added to each mixture. Each sample was placed in a thermocycler and then underwent the following programme: 37^{o}C for 30 minutes, 80^{o}C for 15 minutes and held at 4^{o}C. 1 µl ARTseq RNase Mix was added to each reaction and incubated at 55^{o}C for 5 minutes. Samples were then purified using an RNA Clean and Concentrator Kit (Zymo). The samples were purified according to manufacturers instructions. All samples were eluted in 10 µl of nuclease free water and stored at -20^{o}C.

2.10.9. Purification of cDNA

For both total and protected samples, 10 µl of sample was mixed with 10 µl of Gel Loading Buffer II. A 20/100 Oligo ladder was used as the size marker. 4 µl of ladder was mixed with 1µl of Nuclease-free water and 5 µl of Gel Loading Buffer II. All samples and ladder were incubated at 95^{o}C for 5 minutes and then cooled on ice. All samples were loaded onto Novex 10% precast TBE urea-polyacrylamide gels. The gel was run at 150 V in TBE buffer until the bromophenol blue had completely run off the gel. The gel was then stained in SYBR Gold for 15 minutes. The gel was then viewed under a blue light transilluminator. Bands were excised at 80-100 nts and 70-80 nts for total and protected samples respectively. The excised gel bands were then placed into a 0.2 ml PCR tube with a hole punched in the bottom. This PCR tube was placed into a 0.5 ml microcentrifuge tube which also had a hole punched in the bottom by a needle. The tubes were placed in a 1.5 ml microcentrifuge tube. The tubes and gel were centrifuged for 12000 g for 2 minutes. The 0.5 and 0.2ml tubes were removed and the following mixture was added to the gel found in the 1.5ml micro centrifuge tube: Nuclease-free water, 400 µl; 5M ammonium acetate, 40µl; 10% SDS, 2 µl. The samples were shaken for 1 hour at 37^{o}C. A 1ml pipette tip with its tip cut to increase the bore of the tip, was used to transfer the mixture to filter tubes. The whole sample was centrifuged at 2000 g for 3 minutes. The sample was precipitated and 2 µl of glycogen (20 mg/ml) and 700 µl of 100% isopropanol was

added with the supernatant to a new 1.5 ml microcentrifuge tube. The sample was mixed by vortexing and placed at –20^{o}C overnight. The sample was then washed in 80% ice cold ethanol and resuspended in 10 µl of nuclease free water.

2.10.10. Circularize cDNA

A master mix was created that contained each of the following reagents for each sample: ARTseq CL Reaction Mix, 4 µl; ATP, 2 µl; $MnCl_2$, 2 µl; CircLigase, 2 µl. 10µl of the master mix was mixed with 10 µl of each sample. The samples were then incubated at 4^{o}C for 2 hours.

2.10.11. Amplification of samples using PCR

For each sample, a master mix was created with the following reagents for each sample: Nuclease-free water, 5.33 µl; ARTseq Forward PCR Primer (Table 2.3), 0.66 µl; Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB), 8.33 µl. 14.32 µl of the master mix was added to 1.66 µl of sample and 0.66 µl of the specific index primer used for each sample (Table 2.3). The sample was amplified using a thermocycler as seen in Table 2.4.

Amplified samples were purified by native PAGE. Samples were mixed with 3 µl of bromophenol blue. A 10 bp ladder from IDT was used for each gel, 0.5 µl was mixed with 4.5 µl of water and 1 µl of bromophenol blue. All samples were run on a precast 8% Native PAGE TBE gel, 5 µl of xylene cyanol was added to an unused lane. The gel was run at 100 V in TBE buffer until the xylene cyanol band had run completely of the gel. A band at 140-160 bp in size was excised ensuring not to include the band at 113 bp, which was the amplified 3′ adaptor primer. The excised gel bands were then placed into a 0.2 ml PCR tube with a hole punched in the bottom. This PCR tube was placed into a 0.5 ml microcentrifuge tube which also had a hole punched in the bottom by a needle. The tubes were placed in a 1.5 ml microcentrifuge tube. The tubes and gel were centrifuged for 12000 g for 2 minutes. The 0.5 ml and 0.2 ml tubes were removed and the following mixture was added to the gel found in the 1.5 ml micro centrifuge tube: Nuclease-free water, 400 µl; 5 M ammonium acetate, 40 µl; 10% SDS, 2 µl. The samples were shaken for 1 hour at 37^{o}C. A 1 ml pipette tip with its tip cut to increase the bore of the tip was used to transfer the mixture to filter tubes. The whole sample was centrifuged at 2000 g for 3 minutes. The sample was

precipitated and 2 µl of glycogen (20 mg/ml) and 700 µl of 100% isopropanol was added with the supernatant to a new 1.5 ml microcentrifuge tube. The sample was mixed by vortexing and placed at –20°C overnight. The sample was centrifuged at 14000 rpm for 15 minutes. The sample was then washed in 80% ice cold ethanol and resuspended in 15 µl of nuclease free water. 1µl of the sample was then run on a 8% precast polyarcylamide TBE gel with same protocol as mentioned above. The sample was used for sequencing if there was no amplified 3′ adapter primer present. If the primer was present on the gel, the sample was PCR purified a second time. PCR reactions were scaled up by a factor of two in order to increase yield for sequencing if necessary.

Primer	Library to amplified	Sequence
Forward	All libraries	AATGATACGGCGACCACCGAGATCTACACGTTC AGAGTTCTACAGTCCGACG
Index 4	bRSV4P	CAAGCAGAAGACGGCATACGAGATTGGTCAGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Index 7	bRSV4T	CAAGCAGAAGACGGCATACGAGATGTAGCCGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Index 8	bRSV8P	CAAGCAGAAGACGGCATACGAGATTCAAGTGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Index 11	bRSV8T	CAAGCAGAAGACGGCATACGAGATGATTCGGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Index 1	MockP	CAAGCAGAAGACGGCATACGAGATCGTGATGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Index 2	MockT	CAAGCAGAAGACGGCATACGAGATACATCGGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Table 2.3: Primers used to amplify sequence libraries
Index and forward primers used for each sample to create the respective sequencing libraries. Specific index sequence for each primer is underlined.

Temperature in °C	Time in seconds	Cycles
98	30	1
94	15	9
55	5	
64	10	
4	Hold	Hold

Table 2.4: PCR conditions used to amplify sequence libraries
PCR conditions used to amplify sequence libraries using Phusion High-Fidelity DNA polymerase

2.10.12. Sequencing

Samples were sequenced using MiSeq Next Gen sequencer. Sequencing was performed by the University of Warwick Genomics department. For each MiSeq sequencing cartridge used, three samples were selected with optimal index primer combinations for sequencing. 2 nM of each sample was required for each sequencing cartridge. The purity and quantity of the sample was confirmed using Bioanalyser and Qubit DNA analyser. Raw next generation sequencing results were then stored on Base Space cloud server. For each cartridge, three libraries were sequenced based on the optimal combinations of index primers. For one cartridge bRSV4P, bRSV8P and MockP libraries were sequenced. The second cartridge contained bRSV4T, bRSV8T and MockT libraries.

2.10.13. Processing and analysis of sequencing from ribosomal profiling

Sequencing was stored on the cloud server BaseSpace. The quality of each sample was determined using the software FastQC provided on the Base Space server. Samples were downloaded off the server and processed using MAC operating system. All further processing was performed on a MAC operating system. All further codes that were used for scripts to process data were performed using BASH coding. Four different scripts were used to process the sequencing data. Fastx_toolkit was an open source software provided by the Hannon lab. The software trimmed the adaptor primers off the sequences reads and removed unwanted sequences that were artefacts generated through sequencing. Bowtie and SAMtools scripts were open source software available from John Hopkins University and were provided by Dr Nigel Dyer. Bowtie aligned the reads against each genome. SAMtools processed the aligned reads so they could be viewed using the software IGV. The script orfanalyser was a program created and provided by Dr Nigel Dyer. This software processed the genome that sequencing libraries were aligned to using Bowtie and SAMtools and so could be visualised in the software IGV. The results were viewed and analysed on the open source software IGV. Examples of scripts used can be seen in Appendix A. Calculation for adjust reads and adjusted percentages for both protected and total sequencing libraries are shown in Fig. 2.1.

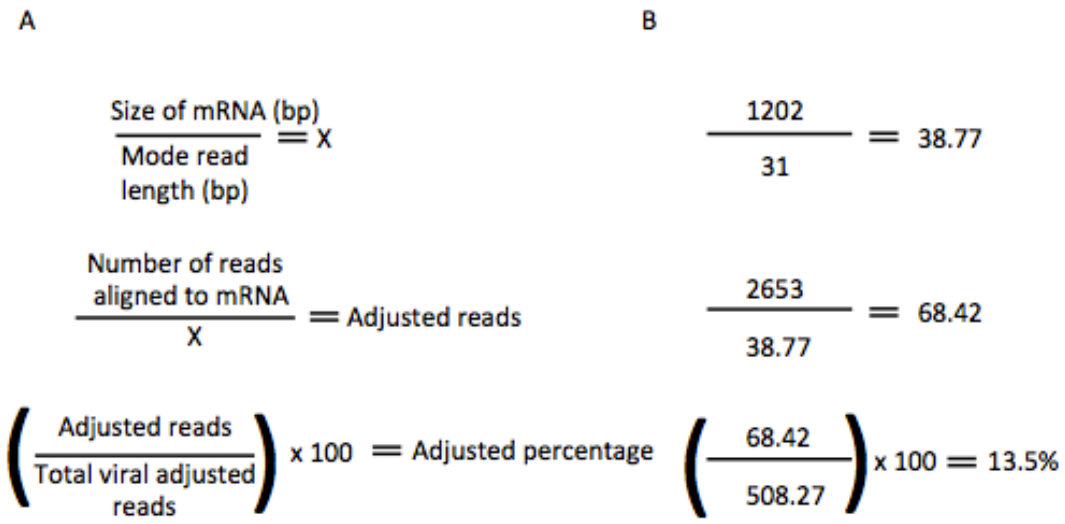


Fig. 2.1: Calculation for adjusted reads and adjusted percentage
(A) Calculation for adjusted reads and adjusted percentages for both protected and total sequencing libraries for each mRNA. (B) A working example of the calculation used to acquire the adjusted percentage of total viral translation for N mRNA in the library bRSV4P.

Chapter 3
Transcriptional and Translational Profiles of bRSV and hRSV

3.1. Introduction

Gene expression in members of the *Pneumoviridae* family is primarily controlled through the transcription of mRNAs and it is thought that primarily the levels of mRNAs transcribed dictate the level of viral proteins translated. With the exception of the mechanism of coupled translation termination/initiation for the initiation of translation of M2 ORF-2 (M2-2 ORF), little regulation occurs at the level of translation of viral mRNAs for this family.

The RNA dependent RNA polymerase (RdRP) complex is thought to control the levels of viral mRNAs through the generation of a transcriptional polar gradient. Initiation of transcription by the RdRP can only occur at the promoter located in the 3′ leader sequence. After binding to the RNA the RdRP is thought to begin transcription at the first gene start signal located in the NS1 gene. The RdRP continues to transcribe the gene until a gene stop signal is reached and then terminates transcription. After release of the nascent mRNA, the RdRP has two possible options; it can either dissociate from the genomic RNA after which it is able to reinitiate transcription by binding only at the promoter or it can move onto the next gene start signal and initiate transcription of the next gene in the genome. This occurs at every gene down the entire length of the genome, with a proportion of RdRPs disassociating at each gene stop signal. As a result, fewer mRNAs are transcribed from genes as the RdRP progresses towards the 5′ end of the genome. Evidence for the polar transcription model is reviewed in Section 1.3.2. It has been reported that transcriptional read-through also occurs in approximately 10% of occasions between genes creating polycistronic mRNAs (Fearns and Collins, 1999b).

Although the primary mechanism of regulation for gene expression in the *Pneumoviridae* family is thought to be through the polar transcription gradient, there is growing evidence to suggest that this is the not the sole factor in controlling the levels of viral mRNA abundance. As reviewed in Section 1.3.2, through the use of transcriptome sequencing, it has been reported that levels of mRNA abundance for several mRNAs (NS2, G and M) in the hRSV genome did not follow levels predicted by the polar gradient model. Instead these mRNAs had higher levels of mRNA abundance than mRNAs transcribed from genes found closer to the 3′ promoter, with the G mRNA (7^{th} gene from the 3′ promoter) having the highest levels of mRNA

abundance in the viral genome (Aljabr et al., 2016). However, these data do suggest that the levels of mRNA abundance for the majority of genes in the hRSV genome are controlled by the position of the gene relative to the 3′ promoter in a manner consistent with elements of the polar gradient model. Data on the transcription efficiency of the hRSV genes was derived from transcriptome sequencing libraries constructed from RNA harvested from hRSV infected cells at 24 and 48 hours post infection (Aljabr et al., 2016). Two factors are important in assessing these data. The first is that the data is determined using steady-state levels of mRNA which may not be a true reflection of transcription levels due to differences in mRNA stabilities. Secondly, the time points at which samples were taken are at the later stages of infections in cells, where cellular resources are likely to be at lower levels than at earlier stages of infection. The different metabolisms of the cells at late stage in infection may affect transcription levels as well as mRNA stabilities. Transcription at earlier time points may reflect a different situation.

The overall levels of mRNAs transcribed are also controlled by the switch from transcription to replication by the RdRP, which is orchestrated by the M2-2 protein through an unknown process. As the M2-2 protein is not present in packaged virions, the switch is dependent on the levels of M2-2 protein translated from the M2 bicistronic gene transcript and the precise levels of M2-2 protein dictate the 'switch', with over expression of the M2-2 protein leading to inhibition of viral replication (Cheng et al., 2005). While the levels of M2-2 protein translated in hRSV are controlled by the non-canonical mechanism of coupled translation termination/initiation it is currently not known which mechanism is used for initiation of M2 ORF-2 by the closely related bRSV. However, due to the location of the M2 ORF-2 (Section 4.1), it is likely that this mechanism will also be a non-canonical initiation mechanism. No additional non-canonical translation initiation mechanisms have been reported to be used by pneumoviruses.

With the advancements in next generation sequencing, it is now possible to build full transcriptional and translational profiles of organisms using the technique of ribosomal profiling developed by Ingolia et al., 2009. Ribosomal profiling works on the principle that all elongating ribosomes translating cellular mRNA cover a region of mRNA which is being translated. During elongation, the ribosome covers 9

nucleotides (which are positioned in the A, P and E sites of the ribosome), and a further 21 nucleotides surrounding this region (Ingolia et al., 2009; Steitz, 1969). This can be advantageous as the region covered by the ribosome (referred to as the footprint) is protected from degradation by RNAases.

Elongation can be paused using the antibiotic compound cycloheximide, a compound originally isolated from the *Streptomyces griseus* bacteria. It has been suggested that this compound binds to the E-site in the 60S ribosomal subunit whilst the E-site deacylated tRNA still resides there. This in turn inhibits the release of the E-site tRNA and therefore translocation of the ribosome is halted. However as a tRNA must be present in the E-site, the ribosome must undergo at least one cycle of translocation before translocation is halted meaning that cycloheximide is never capable of holding the ribosome on the initiation codon in the P-site (Schneider-Poetsch et al., 2010).

Cycloheximide is used in ribosomal profiling to pause all elongating ribosome *in vivo* (Section 2.10.1). The region of mRNA that is covered by each elongating ribosome is then isolated by lysing cells and digesting all other RNA present in the cell using RNases. The region of mRNA that is covered by each ribosome is protected from digestion from the RNase and a footprint is therefore created (Ingolia et al., 2009). RT-PCR amplification and further processing are then used to produce a next generation sequencing library for these footprints (Section 2.10.8-2.10.11).

By using next generation sequencing to sequence these footprints over an entire population of cells, it is possible to calculate the elongating ribosomal frequency on each mRNA in a cellular population based on the number of footprints aligned to that mRNA (Section 2.10.12). This gives valuable information on the levels of translation, frame of translation and reveals which genes are being translated and under what conditions this occurs (Ingolia et al., 2013; Ingolia et al., 2009). This can also be applied to measure viral gene expression during different stages of infection in a cell. A necessary control is to prepare an RNA extract in which the ribosomes are removed. This RNA is then subjected to limited fragmentation using an alkaline buffer followed by RT-PCR amplification prior to sequencing. This provides a control for potential bias occurring during the isolation of the short RNA fragments

and in the RT-PCR step during preparation of sequencing libraries. This control can also be used to establish the abundance of individual mRNAs in the cell.

The aim of this chapter was to use ribosomal profiling to build a transcriptional and translational profile of bRSV to investigate transcriptional and translational regulation. The data obtained here for bRSV was compared and analysed with hRSV transcriptional and translational profiles built and processed from obtained raw sequencing data that was generated previously by Dr Phillip Gould.

3.2. Ribosomal profile analysis of bRSV and hRSV transcription and translation

A detailed method of the ribosomal profiling technique used to build sequencing libraries for next generation sequencing from bRSV (Snook strain) infected MDBK cells harvested at 4 and 8 hours post infection (hpi) can be found in Section 2.10. The bRSV Snook strain was chosen to be used due to the strain being extensively studied in animals (Blodörn et al., 2014; Blodörn et al., 2015; Gaddum et al., 2003). This bRSV strain was also one of the only laboratory strains available that was pathogenic to cells (Thomas et al., 1998; Thomas et al., 1982). These two early time points of infection for both viruses were chosen as it has been reported (Bermingham and Collins, 1999) that the switch from transcription to replication of the viral genome began between 12-15 hpi. Therefore at the 4hr time point in particular it was expected that transcription was likely to be the primary function of the virus RdRP rather than replication. As the levels of viral transcription were likely to be near maximal, these were also optimal times to investigate whether regulation of translation could be detected.

Two libraries were created and sequenced for bRSV at each time point. The first (protected) comprised of a sequencing library where the sequences originated from the footprints belonging to elongating ribosomes translating mRNA. This sequencing library determined the location of all elongating ribosomes at the time point and was used to produce a translational profile. The second (total) consisted of sequences that originated from total cellular RNA after removal of ribosomes. This was the control for the ribosome protected libraries to ensure that there was no bias incorporated in later stages of the sequencing protocol and was also used to construct a transcriptional profile where the levels of mRNA could be quantified. Additional controls derived from total and protected RNAs extracted from mock infected cells were also prepared to ensure there was no previous viral contamination. No evidence of prior infection was detected and these data are not shown.

Raw next generation sequencing data generated from ribosomal profiling of Hep-2 cells infected with hRSV A2 strain were acquired from Dr Phillip Gould. Similar to bRSV, these hRSV infected cells had been harvested at 4 and 8 hpi and for each time point and each time point had the respective protected and total libraries sequenced.

Sequencing data from the analysis was stored on the cloud server BaseSpace. Table 3.1 gives the sequencing library name, the time point cells were harvested and the type of sequencing library. For completeness the equivalent data for the hRSV are also given (P. Gould and A. Easton, personal communication). BaseSpace was also used to test the quality of the raw sequencing data. All raw data was processed on an iMAC 2013, on a OS X Yosemite operating system. A workflow diagram of the scripts and software used to process the raw next generation sequencing data are shown in Fig. 3.1. All scripts were run in terminal, and bash commands were used to run the scripts. Full commands and variables used to execute each script are given in Appendix A.

Sample	Time point harvested post infection of virus	Protected/Total	Virus used	Transcriptional or Translational profile
bRSV8T	8 hours	Total	bRSV	Transcriptional
bRSV8P	8 hours	Protected	bRSV	Translational
bRSV4T	4 hours	Total	bRSV	Transcriptional
bRSV4P	4 hours	Protected	bRSV	Translational
hRSV8T	8 hours	Total	hRSV	Transcriptional
hRSV8P	8 hours	Protected	hRSV	Translational
hRSV4T	4 hours	Total	hRSV	Transcriptional
hRSV4P	4 hours	Protected	hRSV	Translational

Table 3.1: List of sequencing libraries used
Description of sequencing libraries used to create transcriptional and translational profiles

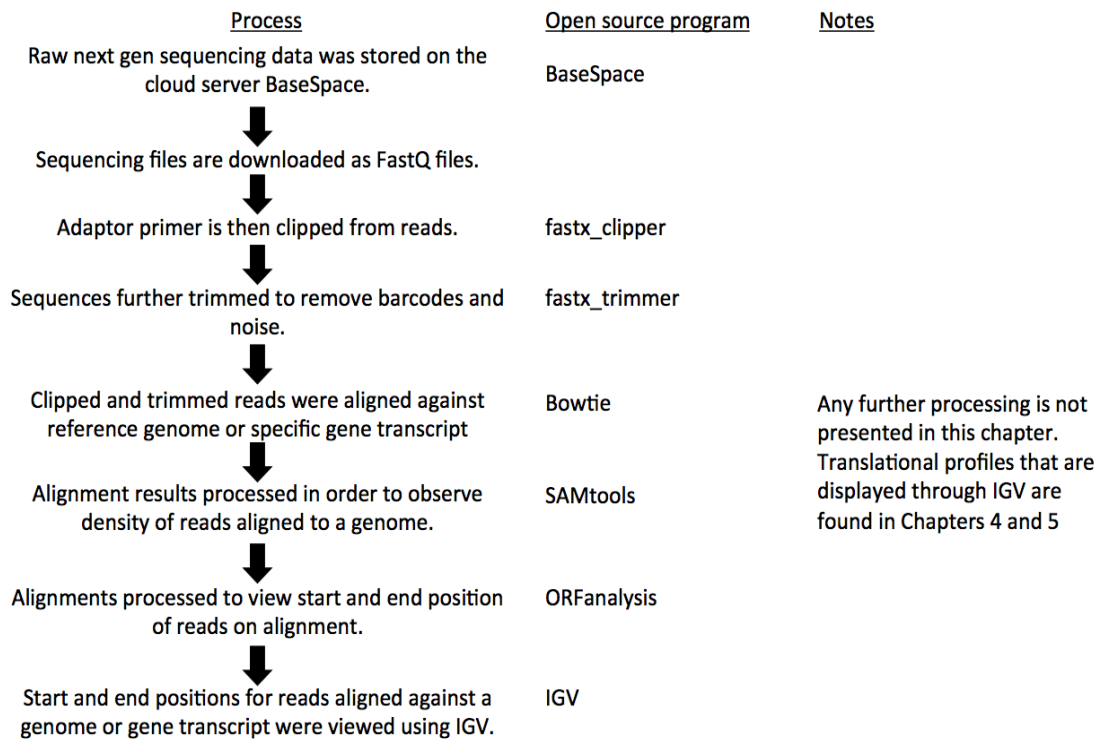


Fig. 3.1: Workflow diagram of processing raw next generation sequencing data
Workflow diagram of processing raw next generation sequencing data created through the technique of ribosomal profiling and open source software used to process the data.

3.3. Quality control of raw data

Quality control of raw next generation sequencing data was performed using the FastQC software on the cloud server BaseSpace. This software analysed the quality of each base along each read. On average, each read was 51nts before trimming. This length included the sequencing sample, the adaptor primer and sequencing bar codes. The software allocated a score to each base for each nucleotide based on the quality of each read. A score below 20 indicated a poor quality nucleotide. A score where the median was less than 20 and the lower quartile was less than 10 was considered completely unusable. A median score between 20 and 28 was considered to be of reasonable quality and could be used. The red line for each nucleotide indicates the median value, the yellow box represents the inter-quartile range and the whiskers represent the 10% and 90% points. The blue line represents the mean quality.

Fig. 3.2 shows a representative quality control (QC) plot for the library bRSV8P. QC plots for all libraries can be found in Appendix B1. All sequencing libraries scored a high quality score for each nucleotide in the 51 nucleotide fragment. All nucleotide positions had the bottom 10% of reads above a score of 28, considered to be an excellent quality of sequence read. The high quality of sequencing meant that it was possible to produce transcriptional and translational profiles in the confidence that no errors were created during the next generation sequencing of sample libraries due to poor quality of reads.

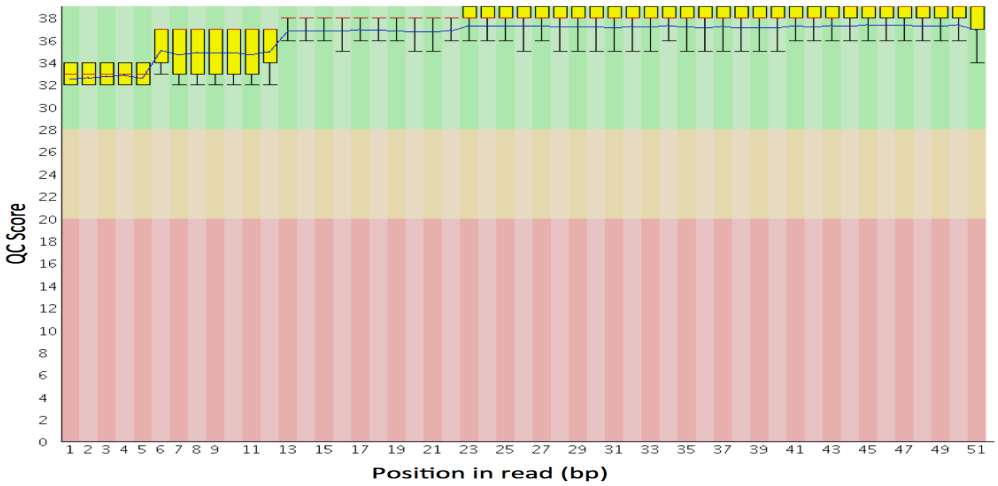


Fig. 3.2: QC plot of reads from sequencing library bRSV8P
QC plot, derived from FastQC software, of reads from bRSV sequencing library bRSV8P. The plot shows the quality of each nucleotide along unprocessed reads.

3.4.1.1. Identification of a transcription polar gradient during transcription for bRSV

Total sequencing libraries containing sequences originating from total RNA from infected cells were used to build a transcriptional profile of bRSV gene expression in MDBK cells harvested at 4 and 8 hpi. These samples contained virus genome and antigenome RNA as well as mRNAs, with mRNA representing the highest proportion of viral RNA present. Although there has been no investigation into the mRNA half-life for bRSV, previous studies on other members of the *Mononegavirales* order (VSV) have determined that the half-life of all viral mRNAs were the same (Pennica et al., 1979).

Alignment of sequencing libraries bRSV4T and bRSV8T to bRSV mRNAs

A detailed description of the processing of raw data is given in Section 3.2. The minimum read length aligned against each mRNA was 25nts. This was the recommended read size suggested by the manufacturer's protocol for the ARTseq Ribosomal Profiling kit used to prepare sequencing libraries. Alignments were performed using the Bowtie open source software (Ingolia et al., 2012). As the complete genome sequence for the bRSV Snook strain has not yet been reported, the sequence of the bRSV strain A51908 (accession: AF092942.1) was used to align the sequence library. Each bRSV mRNA sequence from bRSV A51908 strain was separately aligned against both sequencing libraries.

The L and M2 genes overlap by 67nts (Zamora and Samal, 1992) and unambiguously distinguishing the origin of reads aligned to the overlap region was therefore not possible. In addition, it has been shown for hRSV that a short polyadenylated RNA representing the overlap region of the M2 and L genes is produced as an independent transcript (Collins et al., 1987). For this reason the two more separate alignments were performed where the overlap region was not included so that transcription from the L and M2 genes could be distinguished (Collins et al., 1987).

As each mRNA is a different length, the number of reads aligned to each was normalised against the length of that mRNA. This was performed by taking the mode read length from each sequencing library, which was identified using the software

FastQC to plot the frequency of different sized read lengths as shown in Fig. 3.3 (30 bp for bRSV4T and 30 bp bRSV8T). The adjusted read frequency was then calculated by dividing the length of the mRNA by the mode read length. This figure was then divided into the number of reads aligned for that specific mRNA. This adjusted figure represented the actual number of mRNA strands belonging to each mRNA and the results for each mRNA are presented as a percentage of the adjusted total number reads aligned to all bRSV mRNAs for each sequencing library. The specific calculation can be found in Section 2.10.13. A proportion of all alignments originated from genomic and antigenomic RNA. However, these could not be separated from the mRNA. However, as replication has not begun at the time points chosen, both the genome and antigenome RNA will be present in very low quantities and therefore as each contains equimolar amounts of sequence from all genes they will not affect the final calculation.

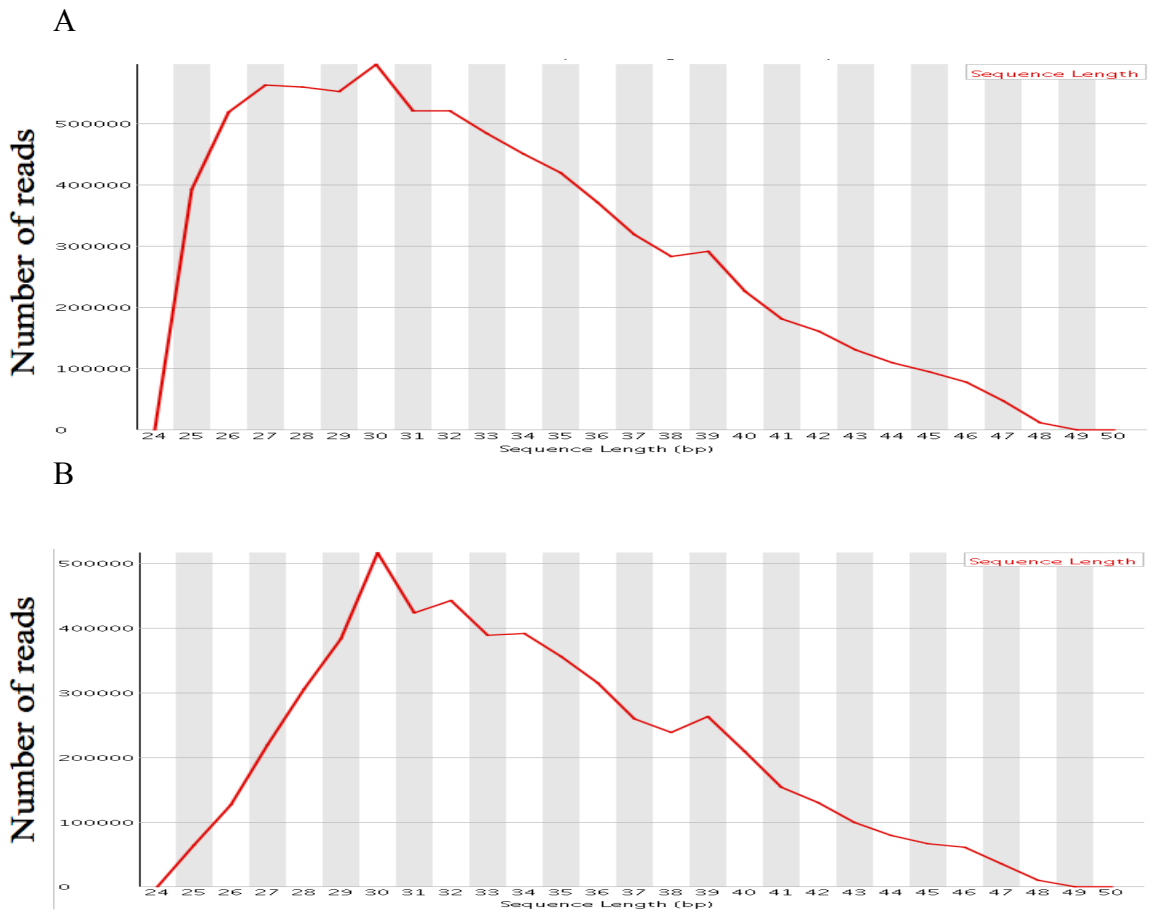


Fig. 3.3 Sequence length distribution of reads in libraries bRSV4T and bRSV8T
Analysis of libraries was performed using FastQC software. Libraries were analysed after the removal of sequences from adaptor primers and barcodes before analysis. (A) Sequence length distribution of reads in the sequencing library bRSV4T. The mode read length was 30bp. (B) Sequence length distribution of reads in the sequencing library bRSV8T. The mode read length was 30bp.

3.4.1.2. Transcriptional profile of bRSV at 4hr and 8 hr post infection

Sequencing libraries derived from RNA extracted at 4 and 8 hpi were used to build transcriptional profiles for bRSV. As shown in Table 3.2, there was a total of 5,548,383 reads and 7,897,845 reads above 25nts in size for the bRSV4T and bRSV8T libraries respectively. There were 3104 reads that aligned to all bRSV mRNAs for the bRSV4T library representing 0.06% of the total number of reads for this library. There were 53325 reads that aligned to all bRSV mRNAs for the bRSV8T library representing 0.68% of the total number for the library. This indicates a 11-fold increase in the levels of viral mRNA between the 4 and 8 hour time points. This may reflect the increasing levels RdRPs produced as the infection proceeds.

Fig. 3.4 shows the levels of bRSV viral mRNA abundance expressed as a percentage of the total number of adjusted reads aligned to all bRSV viral mRNAs at 4 and 8 hpi as described in Section 3.4.1.1. As seen in Fig. 3.4, the relative levels of mRNA for the first 6 genes of the viral genome at both time points resemble that of the predicted polar transcription gradient model. At 4 hpi, for the first three genes NS1 (14.5% of total viral mRNA), NS2 (15.4%) and N (14.7%), there was little fluctuation between the levels of mRNA. However at 8 hpi, the NS1 (18%) and N (17.9%) mRNA levels were higher compared to NS2 (14.1%) which did appear to change dramatically with time. This is not consistent with a polar gradient of transcription but may reflect the accumulated effects of the different stabilities of the mRNAs. Further analysis is required to explore this.

The levels of mRNA at 4 hpi from the N, M, P, SH and G genes decreased from 14.7% of the total level of viral mRNA for the N gene to 4.9% of the total level of viral mRNA for the G gene. However, this was not a linear decrease and there was only a very slight decrease observed between the levels of mRNA from the P (11.5%), M (10.9%), and SH (10.2%) genes with a much greater decrease of 5.3% between the SH and G genes. A similar pattern was observed with these 5 genes at 8 hpi. However the steepness of the gradient was greater with the levels of N gene mRNA at 17.9% of the total and only 0.9% of total mRNAs from the G gene. The decrease in the levels of mRNA between the P (15.3%), M (10.8%) and SH (9.4%)

genes were also much larger than seen at 4 hpi, with a decrease of 4.5% between the P and M mRNAs and a decrease of 1.4% between the M and SH mRNAs. At 8 hpi, the levels of the G gene mRNA, as a proportion of total virus mRNA was also significantly lower than at 4 hpi. There was also a significantly larger drop in levels of mRNA between the SH and G genes with a decrease of 8.5% at this gene junction. Unexpectedly, the G gene was also represented by the lowest level of mRNA from the entire genome at 8 hpi. This is not consistent with the polar gradient model of transcription and may again reflect differences in stability of the different virus mRNAs.

An increase in mRNA levels moving from the G gene to the M2 gene was seen for both time points. At 4 hpi there was an increase in mRNA levels by 1% between the F gene (5.9%) in comparison to the G gene. Although this increase was small, the trend of increasing mRNA levels moving further from the 3′ promoter continued with an increase of 3.7% between the F and the M2 (9.6%) genes. This trend was also observed at 8 hpi, with an increase in mRNA of 4.5% between the G and F (5.4%) genes. A smaller increase of 1.6% was also observed between the F and M2 (7%) genes. This increase in mRNA abundance along the genome is not consistent with a polar gradient of transcription. In addition, the levels of M2 mRNA were similar to those from genes further upstream, with mRNAs from the M2 gene at 9.6% and 7.2% for 4 and 8 hpi, respectively. This was similar to the mRNA levels of the SH gene (10.2% and 9.4%) seen at 4 and 8 hpi respectively.

The furthest gene from the 3′ promoter is the L gene that encodes the enzyme function of the RdRP. There was a 7.2% decrease in mRNA levels from the preceding M2 gene. At 4 hpi, L gene mRNA represented 2.4% of the total level of viral mRNAs. As well as having the lowest mRNA abundance, the reduction in level compared to the preceding gene was the greatest of all. While there was a similar drop in mRNA level at 8 hpi and the L gene was represented by 1.3% less mRNA, it was not the gene represented by least mRNA nor was it the greatest drop observed in comparison to the levels from the neighbouring upstream gene. The reason for the apparent difference in relative level of M2 mRNA is not known and, as before, further analysis will clarify the implications of these data.

The level of mRNA derived from of the 67nt overlap between the M2 and L gene is shown in Fig. 3.5. As can be seen, for the M2 gene at 4 hpi a large proportion of reads aligned to this overlap region, with 52% of the total aligned solely to the remainder of the M2 gene arising from this area. As it was not possible to assign these to the M2 or L genes, or to the 67nt polyadenylated RNA that is generated, this prevented accurate estimation of the increase in the level of mRNAs between the F and M2 genes at 4 hpi. The level was lower (6%) at 8 hpi. This did not affect the levels of mRNA abundance for M2 or L genes. Therefore it is possible to accurately estimate the level of mRNAs between the F and M2 genes at 8 hpi.

The data from this section would suggest that between certain genes in the genome, levels of mRNA abundance did not follow the polar gradient model of transcription at one or both time points. This either could be due to the divergence in mRNA half-life between different gene transcripts or the current model of transcription is not accurate.

	bRSV4T	**bRSV8T**
Total number of reads 25nts or above	5,548,258	7,897,845
Number of reads aligned to bRSV mRNAs	3,104	53,325
Percentage of reads aligned to bRSV mRNAs	0.06	0.66

Table 3.2: Number of reads in sequencing libraries bRSV4T and bRSV8T
Number of reads in the sequencing libraries bRSV4T and bRSV8T and total number of reads aligned to bRSV mRNAs.

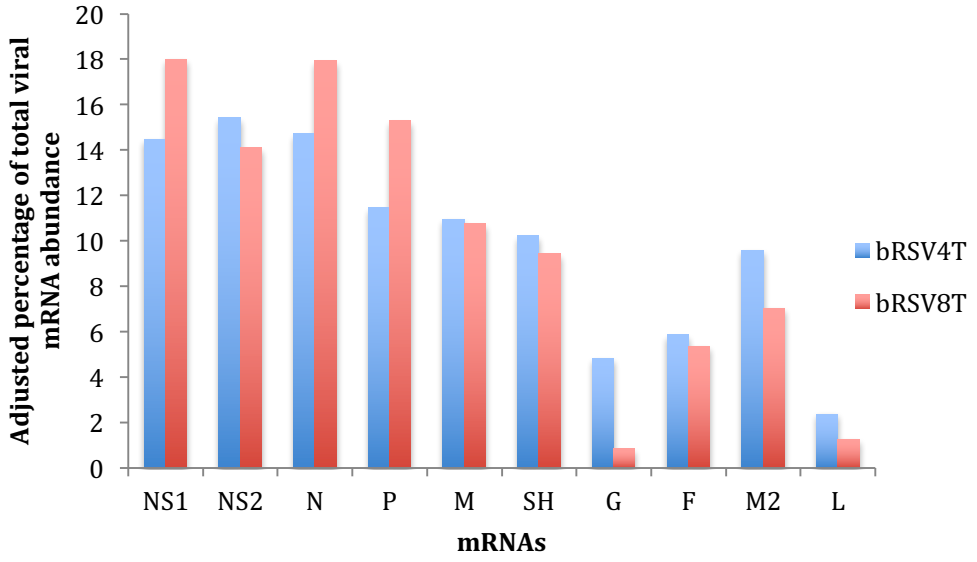


Fig. 3.4: Transcriptional profile of all bRSV mRNAs in infected MDBK cells harvested at 4 hpi and 8 hpi

Transcriptional profile of bRSV at 4hpi and 8hpi using the sequencing libraries bRSV4T and bRSV8T respectively. Transcriptional profile is displayed as the level of reads aligned to each mRNA as percentage of total reads aligned to bRSV viral genome. Reads have been normalised against the size of each mRNA.

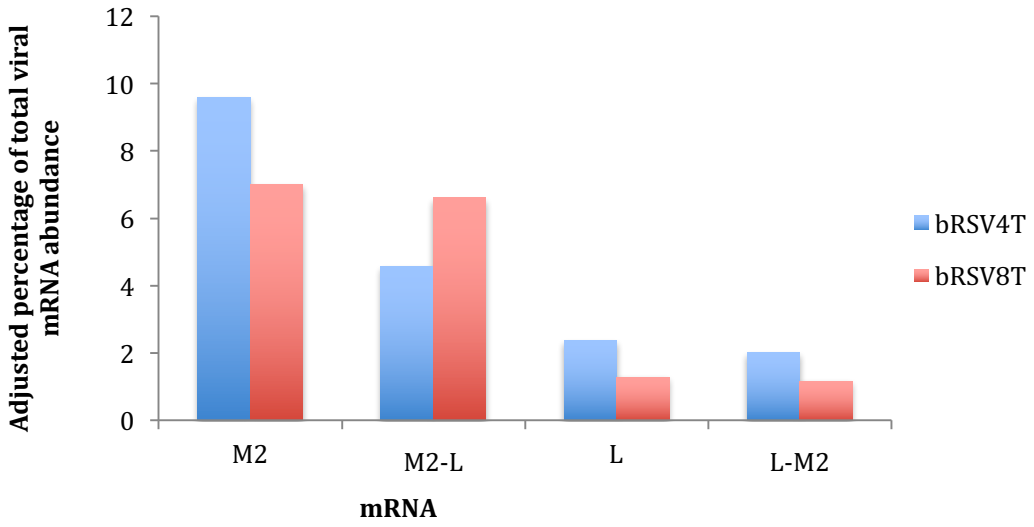


Fig. 3.5: Transcriptional profile showing the levels mRNA abundance of M2 and L mRNAs of bRSV at 4 hpi and 8 hpi.

Transcriptional profile showing the abundance of bRSV M2 and L mRNAs at 4hpi and 8hpi using the sequencing libraries bRSV4T and bRSV8T respectively. The transcriptional profile is expressed as the level of reads aligned to each mRNA as a percentage of total reads aligned to bRSV. Reads were normalised against the size of each mRNA. The figure also shows the adjusted percentage when the sequence from M2/L overlap was removed from the alignment of each mRNA sequence. M2-L represents the removal of the overlap sequence from M2 mRNA during alignment. L-M2 represents the removal of the overlap sequence from L mRNA during alignment.

3.4.2.1. Identification of a polar transcription gradient of hRSV genes

Sequencing libraries derived from total unprotected RNA were used to build transcriptional profiles of hRSV strain A2 by Dr Phillip Gould using the approach described above. As with bRSV the sequence libraries were created from RNA harvested at 4 hpi (hRSV4T) and 8 hpi (hRSV8T). As with bRSV, there is currently no data on the half-life of hRSV mRNAs but studies on other members of the mononegavirales order (VSV) suggested that the half-life of all viral mRNAs were the same (Pennica et al., 1979). All data analysis for the hRSV transcriptional profiling was carried out as part of this thesis.

Alignment of sequencing libraries hRSV4T and hRSV8T to hRSV mRNAs

A detailed description of the processing of raw data is given in Section 3.2. The minimum read length aligned against each gene was 25nts and alignments against the hRSV A2 genome (accession: KT992094.1) were performed using Bowtie open source software. Each hRSV mRNA sequence from the hRSV A2 strain was separately aligned against both sequencing libraries. In hRSV the M2 and L genes overlap by 68nts (Fearns and Collins, 1999a) and the analysis of this region was as before (Section 3.4.1.1). FastQC software was used to plot the read length distribution in each library. The mode length of 36 bp for hRSV4T and 34 bp hRSV8T is shown in Fig. 3.6. Reads were adjusted for gene transcript length using the same method as described in Section 3.4.1.1.

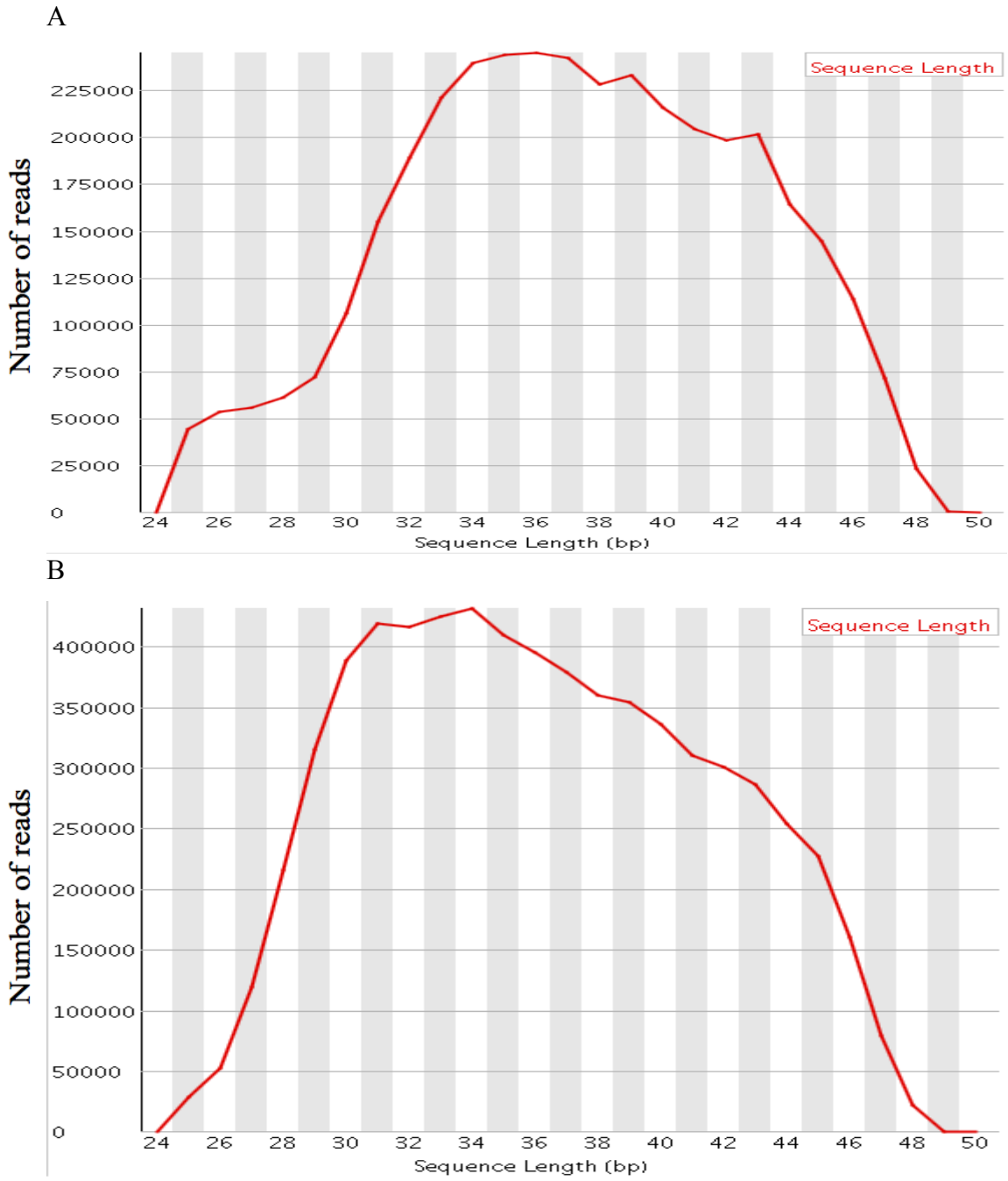


Fig. 3.6 Sequence read length distribution in sequencing libraries hRSV4T and hRSV8T
Analyses of libraries were performed using FastQC software. Libraries were analysed using the FastQC software after removal of adaptor primers and barcode sequences. (A) Sequence read length distribution in the sequencing library hRSV4T. The mode read length was 36bp. (B) Sequence read length distribution in the sequencing library hRSV8T. The mode read length was 34bp.

3.4.2.2. Transcriptional profile of hRSV at 4 and 8 hpi

The sequencing libraries hRSV4T and hRSV8T were used to create transcriptional profiles of hRSV at 4 and 8 hpi. The sequence reads from each library were aligned against each hRSV A2 gene. As seen in Table 3.3, there was a total of 3,733,622 reads in the sequencing library hRSV4T and 6,693,375 reads in the sequencing library hRSV8T that were above 25nts in size. At 4 hpi, 6,219 reads (0.17% of the total) were aligned to hRSV genes. At 8 hpi 18,713 (0.28% of the total) reads were aligned to hRSV genes. The read density suggested that there was a 16-fold increase in mRNA levels between 4 and 8 hpi which was a greater increase than was seen with bRSV. However, this may simply reflect differences in sample sizes for the two virus systems.

Fig. 3.7 shows hRSV viral mRNA abundance as a percentage of the total number of adjusted reads aligned to all hRSV viral mRNAs in hRSV infected cells at 4 and 8 hpi as described in Section 3.4.1.1. As expected the NS1 gene as the first gene in the genome was the most represented by mRNA levels of all genes (20.9% and 40.5% of total virus mRNAs at 4 and 8 hpi, respectively) and the L gene represented the least at both time points (0.6% and 0.4% of all viral mRNAs at 4 and 8 hpi, respectively) as the furthest gene from the 3´ promoter in the genome.

At 4 hpi, a decrease in mRNA levels occurred between the NS1 gene (gene 1) and the N gene (gene 3) with a decrease of 4.2% in levels of mRNA between the NS1 (20.9% of total virus mRNAs) and NS2 (16.7%) genes. A similar decrease of 4.3% was seen between the NS2 and N (12.4%) genes. However an increase of 2.6% was seen in mRNA levels between the N and P genes. A decrease in mRNA levels was seen sequentially in the P to G genes with a small decrease of 3% between the P (15%) and M (12%) genes, a larger decrease of 7.7% between the M and SH (4.2%) genes and a small decrease of 0.7% between the SH and G (3.6%) genes. Levels of mRNA increased between the G and M2 genes with an increase of 2.7% between the G and F (6.3% of virus mRNAs) genes and 2% increase between the F and M2 (8.3%) genes. These two increases in mRNA abundance observed between neighbouring genes were a divergence from the polar gradient of transcription. Finally, a decrease of 7.7% in the levels of mRNA was detected between the M2 and L genes.

A similar pattern of changing RNA levels was seen at 8 hpi, though the precise levels differed and there was no increase in mRNA levels between the N and P genes as was seen at 4 hpi. There was an increase that was not consistent with the polar gradient model in mRNA levels between the NS2 and N genes. The values obtained for the genes were: the NS1 gene expressed 40.5% of total virus mRNAs, the NS2 gene 10.5%, the N gene 15.3%, the P gene 8.9%, the M gene 8.9%, the SH gene 1.4%, the G gene 0.5%, the F gene 5.7%, the M2 gene 7.8% and the L gene 0.4%.

As with the bRSV M2/L gene overlap region separate sets of alignments were performed in which the overlapping sequence were removed from each gene and the remaining sequences aligned against the sequence library. For each mRNA with the removed overlap, a new the figure as a percentage of total viral mRNAs was calculated. As seen in Fig. 3.8, the proportion of reads found in this overlap is extremely small these did not affect the overall transcriptional profile observed for hRSV.

The data from this section would suggest that between certain genes in the genome, the levels of mRNA abundance did not follow the polar gradient model of transcription at one or both time points. This either could be due to the divergence in mRNA half-life between different gene transcripts or the current model of transcription is not accurate.

	hRSV4T	**hRSV8T**
Total number of reads 25nts or above	3,733,380	6,688,948
Number of reads aligned to hRSV mRNAs	6,219	18,713
Percentage of reads aligned to hRSV gene transcipts	0.17	2.8

Table 3.3: Number of reads in sequencing libraries hRSV4T and hRSV8T
Number of reads in the sequencing libraries hRSV4T and hRSV8T and total number of reads aligned to hRSV mRNAs.

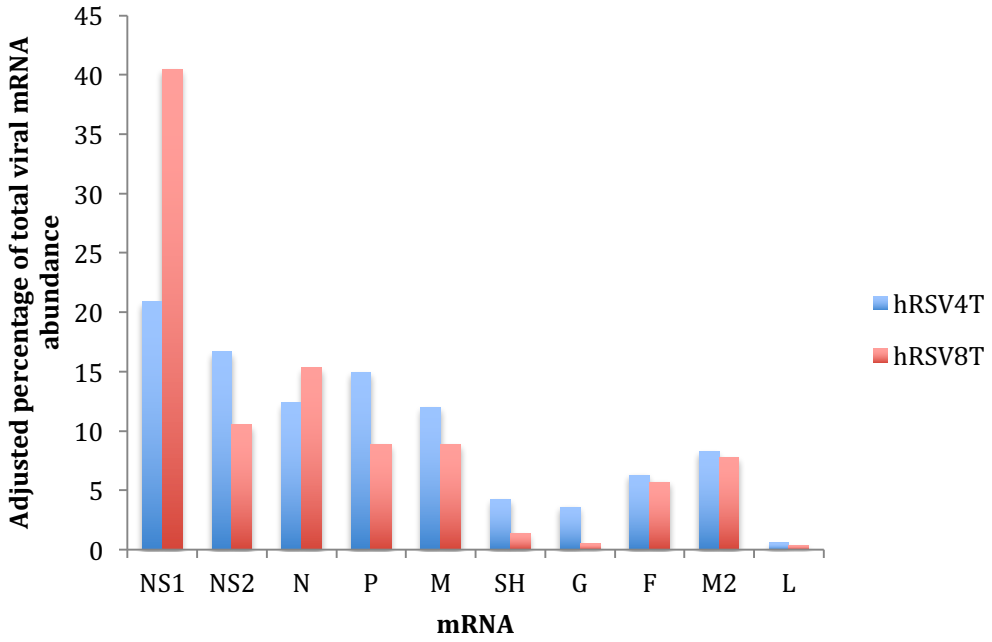


Fig. 3.7 Transcriptional profile of all hRSV mRNAs at 4hpi and 8hpi.
Transcriptional profile of hRSV at 4hpi and 8hpi using the sequencing libraries hRSV4T and hRSV8T respectively. Transcriptional profile is displayed as the level of reads aligned for each mRNA as percentage of total reads aligned to the hRSV viral genome. Reads have been normalised against the size of each mRNA.

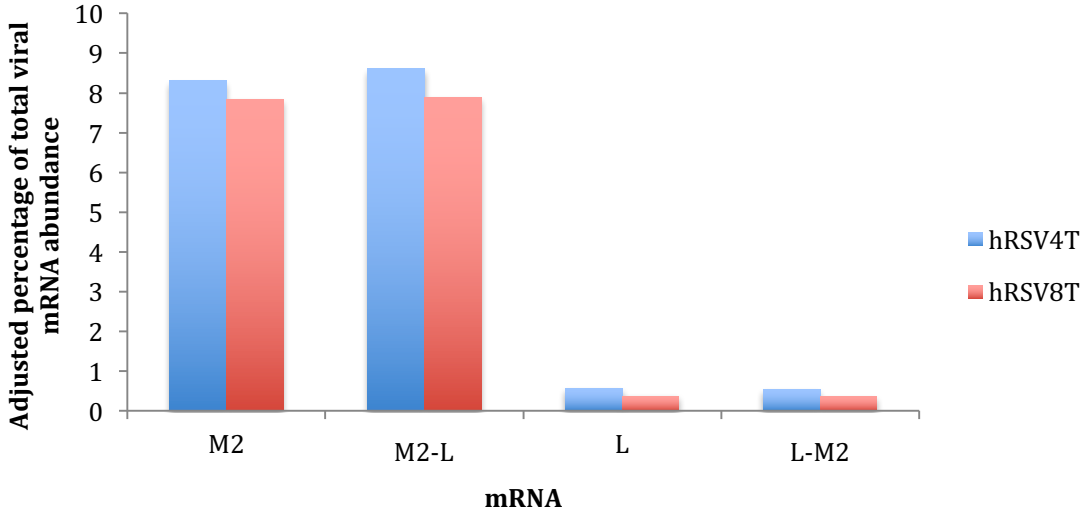


Fig. 3.8: Transcriptional profile showing the levels of mRNA abundance of M2 and L mRNAs in hRSV at 4 hpi and 8 hpi.
Transcriptional profile of mRNA abundance for the hRSV M2 and L mRNAs at 4hpi and 8hpi using the sequencing libraries hRSV4T and hRSV8T, respectively. The transcriptional profile is expressed as the level of aligned for each mRNA as percentage of total reads aligned to hRSV. Reads were normalised against the size of each mRNA. The figure also shows the adjusted percentage when the sequence from M2/L overlap is removed from the alignment of each mRNA sequence M2-L represents the removal of the overlap sequence from the M2 mRNA during alignment. L-M2 represents the removal of the overlap sequence from the L mRNA during alignment.

3.5.1.1. Identifying translational regulation during translation of bRSV viral mRNAs

The current model of regulation of gene expression in pneumoviruses is that the majority of the control is exerted through the level of mRNA abundance. There is no evidence that regulation of translation influences the level of gene expression, with the exception of the M2 ORF-2 in the M2 mRNA. The sequencing libraries bRSV4P and bRSV8P contained reads that originated from fragments of mRNA that were protected by ribosomes, present a snapshot of the location of all elongating ribosomes on virus mRNAs at 4 and 8 hpi, respectively. These data were used to build profiles of bRSV mRNA translation which could be interrogated for evidence of regulation of translation in bRSV. Regions within mRNA with high or low distributions of elongating ribosomes may also indicate areas of interest such as highly structural regions or internal initiation sites. With no translation regulation, the ribosome distribution for each viral mRNA should be proportionally equal to proportional levels of the respective individual mRNAs. Deviation from the expected profile would indicate that translational regulation was a factor in viral gene expression.

Alignment of sequencing library to bRSV mRNAs

A detailed description of the processing of raw data is given in Section 3.2. The minimum read length aligned against each mRNA was 25nts as recommended by the protocol. Alignments were performed using Bowtie open source software. As before, the genome sequence of the bRSV strain A51908 was used to align the sequencing libraries derived from RNA extracted from cells infected with bRSV (Snook strain). The sequencing libraries were aligned against each mRNA sequence from the virus genome except for the M2 mRNA. This was due to the presence of two ORFs which overlap each other by up to 17nts. To address this, two individual alignments were performed. The first used the sequence from the entire M2 ORF-1 including the stop codon and the second used the sequence beginning at the AUG at position 554 from the start of the M2 mRNA and continued up until the end of the M2 ORF-2. This AUG is the 5′ proximal putative start of the M2 ORF-2. As the two ORFs overlapped, it was therefore possible that these footprints could align twice in this region. The data however showed that there were only very small amount of reads aligned to the M2 ORF-2 suggesting that the majority of the reads arose from

ribosomes translating the M2 ORF-1. The overlap between the M2 and L genes does not present a similar problem, as M2 ORF-2 does not extend into the L mRNA sequence. However, a short 67 nt mRNA-like molecule is transcribed from the M2/L gene overlap region and as this has the L protein initiation codon and is therefore capable of making a short peptide, a proportion of ribosomes at the 5′ end of the L mRNA may arise from this RNA rather than the mRNA encoding the entire L protein. It was impossible to distinguish the origin of elongating ribosomes in this region. Therefore no attempts were made to alter the calculations for this mRNA.

As each mRNA in the viral genome differs in length, the number of reads aligned to each mRNA were normalised against the length of the individual mRNAs to provide the number of ribosomes per unit length of mRNA. First, the length of each mRNA was divided by the mode read length of each sequencing library as shown in Fig. 3.9 (31 bp for bRSV4P and 28 bp for bRSV8P calculated from the software FastQC). This value was the number of fragmented read lengths per mRNA length. The mRNA length was divided by the mode read length so levels of translation between different libraries could be compared. These values for each mRNA were then divided into the number of ribosome protected fragments aligned to the specific mRNA to generate the number of full length mRNAs that were completely covered in protected fragments (adjusted number of reads). These adjusted reads for each mRNA were expressed as a percentage of the adjusted total number of reads aligned to all viral mRNAs in each sequencing library. This percentage represented the adjusted percentage of total viral translation for each mRNA. The calculation and a working example are shown in Section 3.10.3.

Any change in translation compared to mRNA abundance was identified by measuring both the change in proportional viral translation to proportional viral mRNA abundance and the translational efficiency of each viral mRNA.

Looking at the change in proportional translation to proportional mRNA abundance for each viral gene transcript, identified if the respective levels of translation matched the levels of mRNA abundance. For each viral mRNA, this was calculated by taking the adjusted percentage of total viral translation (as discussed above), and comparing this to the adjusted percentage of total viral mRNA abundance (Section

3.4.1.2) from the respective total library harvested at that time point. If the proportion of viral translation and proportion of mRNA abundance were equal for a viral mRNA, then it was unlikely that there was any translation regulation and translation was controlled by viral mRNA abundance. However, a change in proportional translation to proportional mRNA abundance would suggest that translational regulation was influencing the levels of ribosomes translating that viral mRNA.

The translational efficiency for each viral mRNA was calculated by dividing the adjusted percentage of total viral translation (discussed above) by the adjusted percentage of total viral mRNA abundance for each viral mRNA (Section 3.4.1.2). A translational efficiency of 1 indicated that there was no translational regulation. A translational efficiency between 0 and 1 would indicate that the translation is being downregulated. Correspondingly, a translational efficiency value above 1 would indicate upregulation to increase translation. In this way a translational efficiency value of 1 would indicate that translation of the mRNA is linked to abundance and that there is little probability of it being affected by translational regulation.

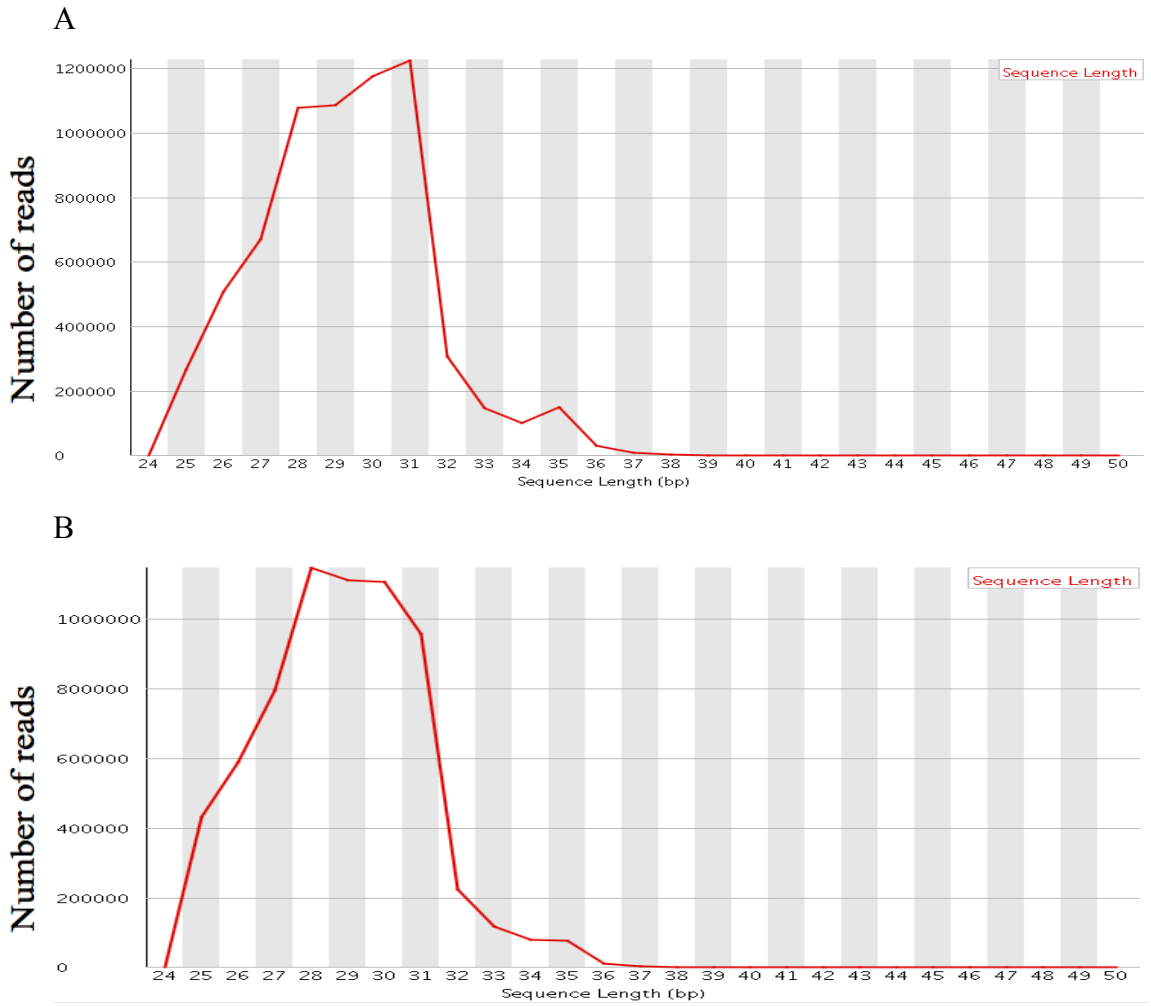


Fig. 3.9: Sequence length distribution of reads in sequencing libraries hRSV4T and hRSV8T
Analysis of libraries were performed using FastQC software. Libraries were analysed using FastQC software after removal of adaptor primers and barcode sequences. (A) Sequence length distribution of reads in the sequencing library bRSV4P. The mode read length was 31 bp. (B) Sequence length distribution of reads in the sequencing library bRSV8P. The mode read length was 28 bp.

3.5.1.2. Translational profile of bRSV

For both sequence libraries bRSV4P and bRSV8P, derived from ribosome-protected RNA, reads of 25 nts or longer were aligned against each mRNA (including the M2 ORF-1 and M2 ORF-2s) from the bRSV strain as before. A total of 6,775,850 reads were obtained for the bRSV4P library prepared from RNA extracted at 4 hpi and 6,665,582 reads for the bRSV8P library for RNA extracted at 8 hpi. Although rRNA removal was performed during the preparation of each sequencing library (Section 3.10.4), 2.5% and 5.2% of reads still aligned to Bos taurus rRNA in the bRSV4P and bRSV8P libraries respectively.

When aligned to bRSV mRNAs, 0.2% of the reads aligned to the bRSV4P library and 1.02% aligned to the bRSV8P library. The increase in level of ribosome-associated RNAs is likely to be due to the increase in virus mRNA levels detected between 4 and 8 hpi described in Section 3.4.1.2. However, the 10-fold increase in mRNA levels between 4 and 8 hpi was matched by a 5.1-fold increase in the levels of translation. The reason for the difference is not known. Fig. 3.10A and Fig. 3.11A shows the change in proportional translation to proportional mRNA abundance for each viral mRNA at 4 and 8 hpi, respectively. The translational efficiencies for each viral mRNA at 4 and 8 hpi, were displayed in Fig. 3.10B and Fig. 3.11B respectively.

M2 ORF-2 was not considered in this analysis due to the unique nature of the M2 bicistronic mRNA. Translation of this ORF in hRSV is known to use a non-canonical mechanism and therefore it is likely that translation of bRSV M2 ORF-2 is regulated by factors other than mRNA abundance. The number of reads aligned to this ORF were also so small that omission did not affect the calculations. The adjusted percentage of total viral translation for the M2 ORF-1 was included and the value for ribosome-protected reads aligned to the M2 ORF-1 were adjusted to the size of the M2 ORF-1 alone rather than the entire M2 mRNA. For both the translational efficiency and the change in proportional translation to mRNA abundance, the adjusted percentage of total viral mRNA for the M2 mRNA was used to calculate both figures.

Fig. 3.10 shows the data from the RNAs prepared at 4 hpi. Fig. 3.10A shows the change in proportional translation to proportional mRNA abundance for each viral mRNA and the black line represents the point where the proportion of translation is equal to the proportion of mRNA abundance for each mRNA. As no repeats were able to be performed for these time points, the mean change in translation against mRNA abundance of all mRNAs in the viral genome is shown as the two red lines on each figure. This enabled identification of viral mRNAs where the change in translation greatly exceeded the mean variation, which would support the possibility that the change was not due to random variance and was caused by regulation of translation. Fig. 3.10B shows the translational efficiencies for each viral mRNA.

At 4 hpi, the NS1, NS2, P and G mRNAs all had higher proportional levels of translation than expected from their levels of mRNA abundance, with translational efficiencies of 1.22, 1.17, 1.3 and 1.63 respectively. All of these values were higher than the average change strongly suggesting that they were subject to translational upregulation.

The SH mRNA and M2-1 ORF had lower levels of proportional translation compared to the levels of their mRNAs with translational efficiencies of 0.71 and 0.41, respectively. These changes in translation were also greater than the average change, suggesting that both of these mRNAs are translationally downregulated

For the L mRNA, the change in the translation in comparison to mRNA abundance was 1.2%, which was below the average change (2.59%). However proportionally, this was a 50% drop in translation compared to mRNA abundance and the translational efficiency was significantly lower (0.5) than that determined for the SH mRNA. This suggests that the decrease in translation was not caused by variation and was likely due to downregulation of translation. The low proportional change can therefore be accounted for the low levels of translation and mRNA abundance instead of a lack of regulation.

For all other viral mRNAs from the N, F and M genes, there was little change in translation compared to mRNA abundance. Further replicates of the analysis will be required to establish if these small changes are significant.

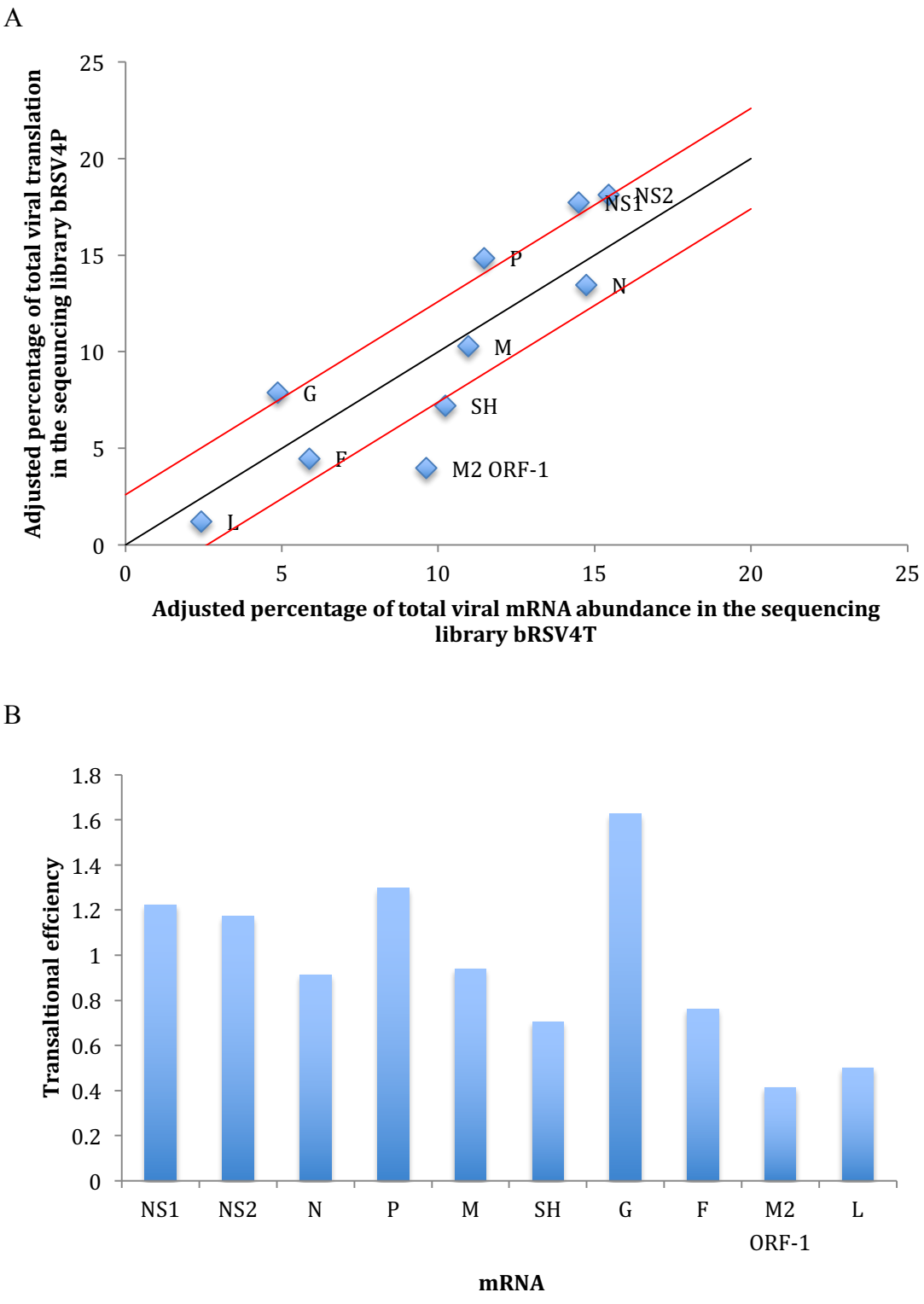


Fig. 3.10: Comparison of proportional translation against proportional mRNA abundance and translational efficiencies of bRSV viral mRNAs harvested at 4 hpi

(A)Adjusted levels of translation as a percentage of total reads aligned to all bRSV viral mRNAs against adjusted levels of mRNA abundance as a percentage of total reads aligned to bRSV viral mRNAs. The black line represents the points where levels of mRNA abundance and translation were proportionally equal. The red lines represent the mean change in translation from the levels of mRNA abundance. (B). Translational efficiency of viral mRNAs created from libraries where RNA was harvested at 4 hpi from sequencing libraries bRSV4P and bRSV4T

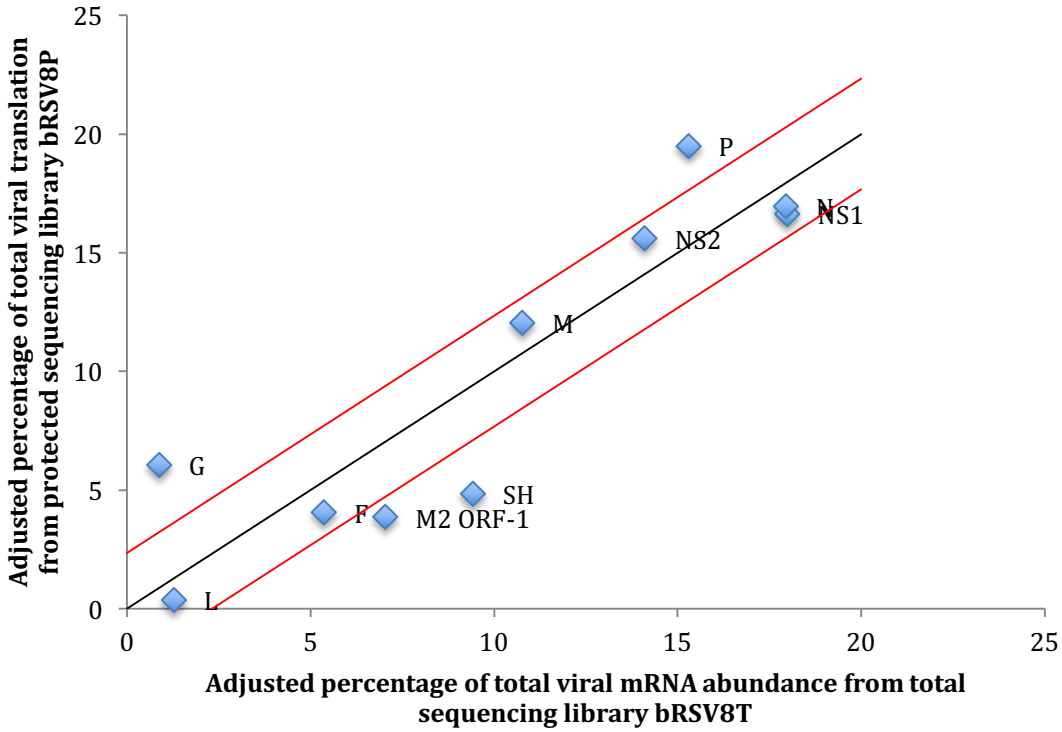
Fig.3.11A and 3.11B show the comparison between proportional translation to proportional mRNA abundance and the translational efficiencies for viral mRNAs extracted at 8 hpi. At 8 hpi, the P and G mRNAs had higher levels of proportional translation than their proportional levels of mRNA abundance with translational efficiencies of 1.27 and 7 respectively. Both of these proportional changes were higher than the average change, strongly suggesting that they were subject to translational upregulation.

The SH mRNA and M2-1 ORF had lower levels of proportional translation compared to the levels of their mRNAs with translational efficiencies of 0.51 and 0.55, respectively. These changes in translation were also greater than the average, suggesting that both of these mRNAs are translationally downregulated at 8 hpi.

As seen at 4 hpi, the L mRNA exhibited a low translational efficiency of 0.27 though this proportional change was within the average change. However, given the very low translational efficiency value, the data suggests that translation of L mRNA is downregulated.

For the other viral mRNAs from the NS1, NS2, N, M and F genes, although all gave translational efficiencies greater or lower than 1 they were very close to unity. This suggests that these mRNAs are not subject to translational regulation. However, as for the equivalent mRNAs at 4 hpi, replicate samples will be required to determine if the small changes are significant.

A



B

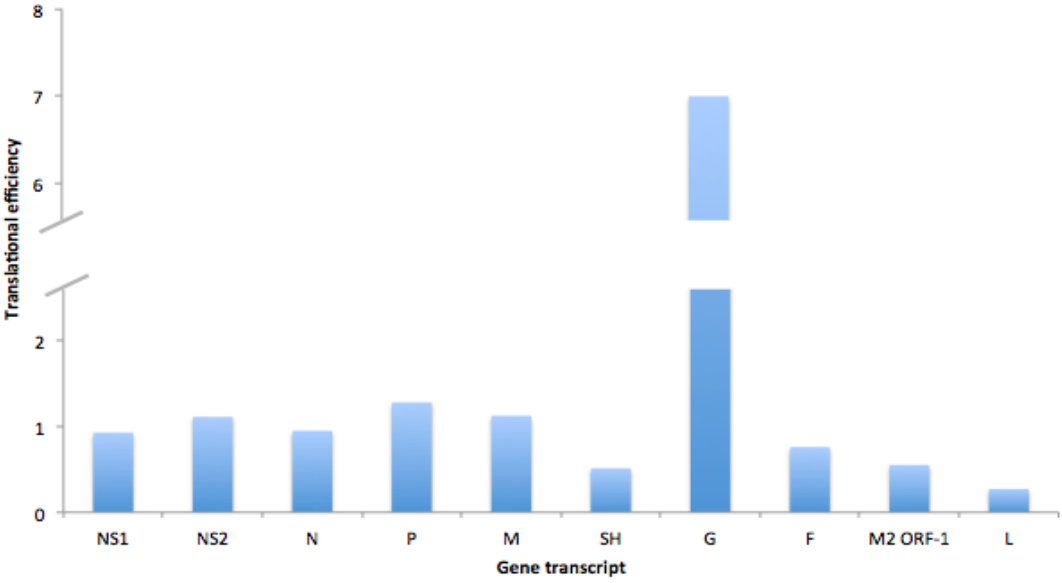


Fig. 3.11: Comparison of proportional translation against proportional mRNA abundance and translational efficiencies of bRSV viral mRNAs harvested at 8 hpi

A) Adjusted levels of translation as a percentage of total reads aligned to all bRSV viral mRNAs against adjusted levels of mRNA abundance as a percentage of total reads aligned to all bRSV viral mRNAs. The black line represents the points where levels of transcription and translation were proportionally equal. The red lines represent the mean change in translation from the levels of transcription. (B) Translational efficiency of viral mRNAs created from libraries where RNA was harvested at 8 hpi from sequencing libraries bRSV8P and bRSV8T

3.5.1.3. Comparison of translational profiles between 4 and 8 hpi for bRSV

In total five viral mRNAs appeared to be upregulated due to translational regulation. The P and G mRNA were upregulated at both 4 and 8 hpi. For the P mRNA, the increase in translation was similar at both time points. However, in contrast, for the G mRNA, there was a 4.4-fold increase in translation efficiency between 4 and 8 hpi.

The SH, L mRNA and M2 ORF-1 were all downregulated at both time points. The levels of downregulation appeared to increase for both the L and SH mRNAs, with a 0.71-fold and 0.54-fold decrease for SH and L mRNAs respectively over both time points. For the M2 ORF-1, only a small change in the levels of translation was observed between the two time points with downregulation observed in both samples. Translation of NS1 and NS2 mRNAs were clearly upregulated at 4 hpi, but not regulated at 8 hpi.

3.5.2.1. Translational profile of hRSV

To investigate potential translational regulation in hRSV gene expression the sequencing libraries hRSV4P and hRSV8P (provided by Dr Phillip Gould) were analysed as described in Section 3.5.1.1.

Alignment of sequencing library to hRSV mRNAs

Sequences with a read length of 25 nts or above were used as recommended by the protocol and the data were processed as before in Section 3.5.1.1, with alignments performed using the Bowtie open source software. The sequence for hRSV strain A2 genome was used to align each hRSV mRNA sequence (except the M2) against both sequencing libraries. The M2 mRNA used the same method as described in Section 3.5.1.1 to align the separate ORFs and to account for the M2 ORF-1/ORF-2 overlap. The mode read length of each sequencing library is shown in Fig. 3.12 and was 29 bp for hRSV4P and 29 bp for hRSV8P. The same methods as described in Section 3.5.1.1, were used to calculate both the translational efficiency and to compare the change in proportional translation to the proportional levels of mRNA abundance for each viral mRNA.

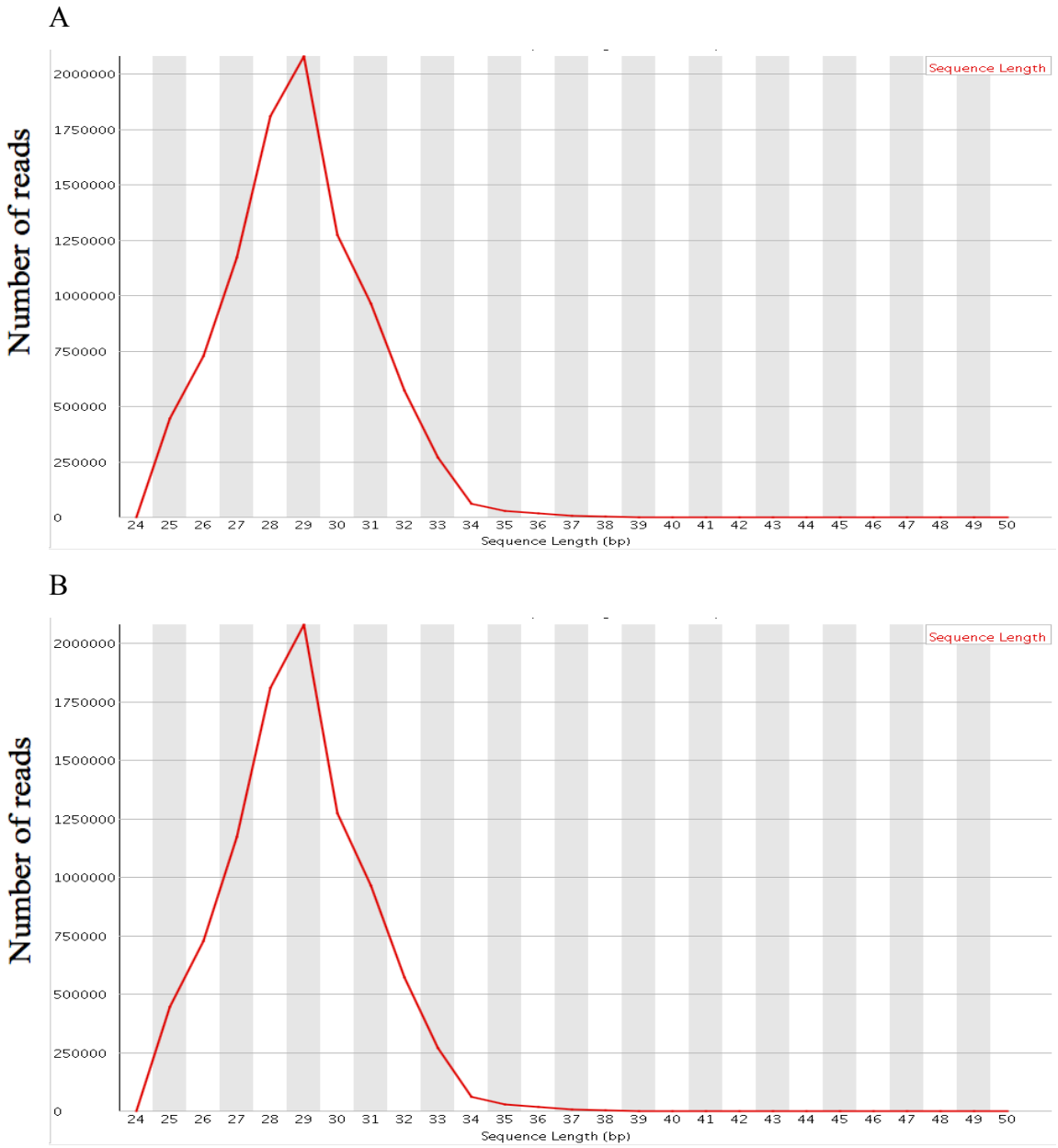


Fig. 3.12: Sequence length distribution of reads in sequencing libraries hRSV4T and hRSV8T
Analysis of libraries were performed using FastQC software. Libraries were inputted into FastQC software after reads had been trimmed of adaptor primers and barcodes. (A) Sequence length distribution of reads in the sequencing library hRSV4P. The mode read length was 29 bp. (B) Sequence length distribution of reads in the sequencing library hRSV8P. The mode read length was 29 bp.

3.5.2.2. Translational profile of hRSV from sequencing libraries hRSV4P and hRSV8P

A total of 9,461,174 reads was obtained for the sequencing library hRSV4P and 10,083,760 for the sequencing library hRSV8P. Although rRNA removal was performed during the preparation of each sequencing library (Section 3.10.4), 33% and 33.9% of reads still aligned to Homo sapien rRNA in the hRSV4P and hRSV8P libraries respectively.

When aligned to hRSV mRNAs, 79,060 (0.84%) and 628,988 (6.24%) reads aligned to the hRSV4P and hRSV8P libraries respectively, suggesting a 7.5 fold increase in translation of hRSV mRNAs between 4 and 8 hpi. As for the analysis of bRSV mRNAs, reads aligned to the levels of translation of M2 ORF-2 were omitted for the subsequent analysis.

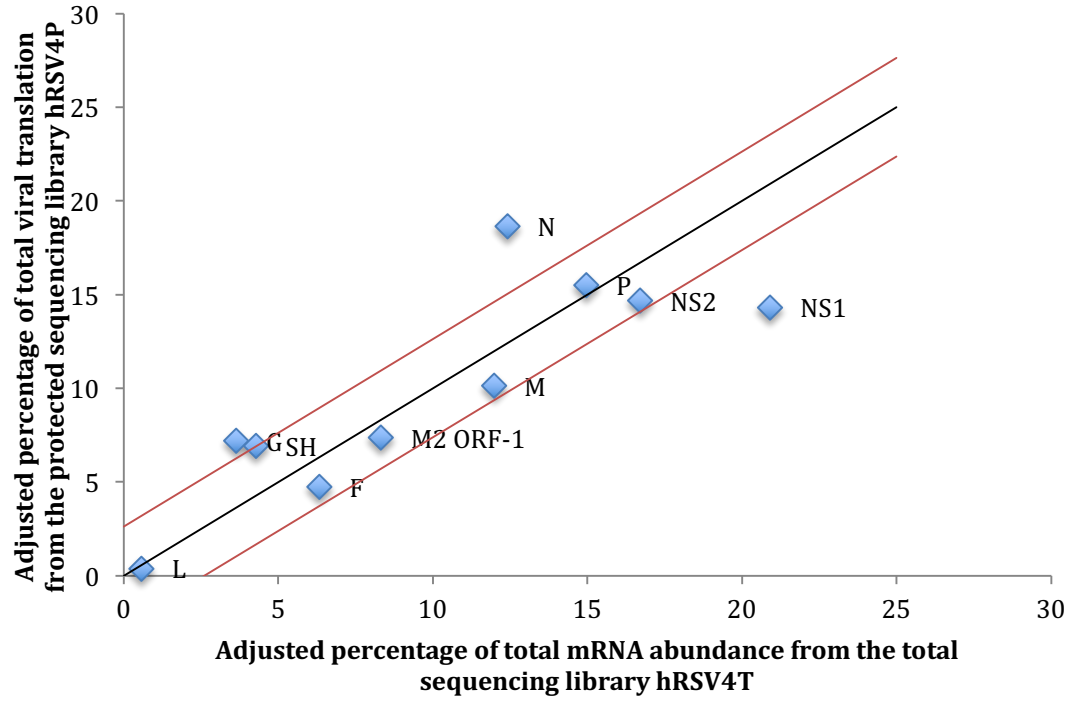
At 4 hpi, the N and G mRNAs had higher levels of proportional translation than the proportional levels of their mRNA abundances with translational efficiencies of 1.5 and 1.98, respectively (Figs. 3.13A and Fig. 3.13B). Both mRNAs had higher changes in translation that the average, suggesting that they were subject to translational upregulation. For the SH mRNA, there was an increase in proportional translation in comparison to the level of the mRNA. While this was less than the average change the translational efficiency was higher than that of the N mRNA (1.63) suggesting that translation of SH mRNA was regulated. This proportional change being lower than the average change was likely due to the low levels of translation and mRNA abundance. As although a large proportional increase was observed, suggesting regulation, in real terms this was only a small increase which was smaller than the average change.

The NS1 mRNA alone had a reduced level of translational efficiency of 0.69. This change in proportional translation was also greater than the average change, suggesting that this mRNA was subject to translational control leading to reduced translation.

As seen with bRSV, the hRSV L mRNA demonstrated a low translational efficiency with a value of 0.68 even though the proportional change was smaller than the

average change. This was also similar to the translational efficiency observed in the downregulated NS1 mRNA. This suggests that translation of the L mRNA is downregulated at 4 hpi. For other viral mRNAs from the NS2, P, M, SH, F and M2 genes, the values of translational efficiencies did not also significantly differ from unity suggesting that their translation was determined by mRNA abundance alone.

A



B

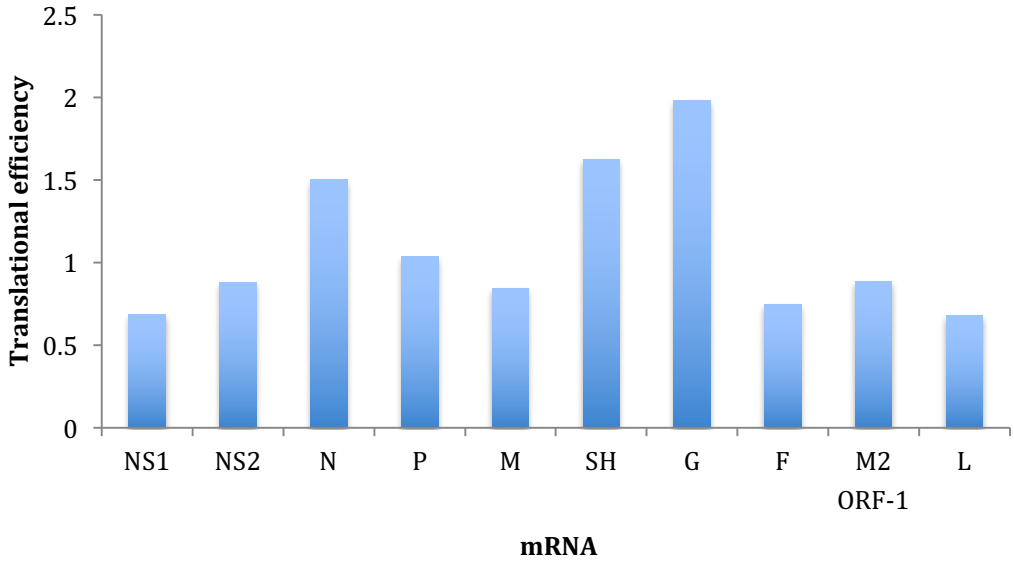


Fig. 3.13: Comparison of proportional translation against proportional mRNA abundance and translational efficiencies of hRSV viral mRNAs harvested at 4 hpi

(A) Adjusted levels of translation as a percentage of total reads aligned to all hRSV viral mRNAs against adjusted levels of levels of mRNA abundance as a percentage of total reads aligned to all hRSV viral mRNAs. The black line represents the points where levels of mRNA abundance and translation were proportionally equal. The red lines represent the mean change in translation from the levels of mRNA abundance. (B) Translational efficiency of viral mRNAs from the hRSV sequencing libraries hRSV4P and hRSV4T.

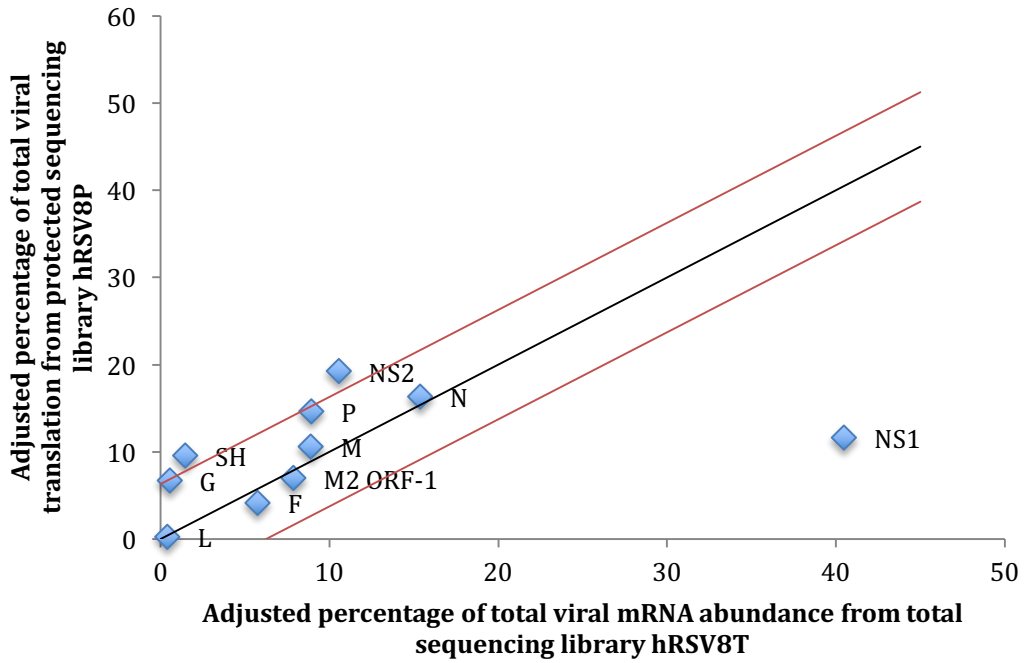
Fig.3.14A and 3.14B display the comparison between proportional translation to proportional mRNA abundance and the translational efficiencies for sequencing viral mRNA harvested at 8 hpi respectively (libraries hRSV8P and hRSV8T)

At 8 hpi, the NS2 and SH mRNAs had higher proportional changes in translation than the average change, with translational efficiencies of 1.83 and 6.57 respectively. This indicates that translation of these mRNAs was being upregulated. As observed for the SH mRNA at 4hpi, there were high translational efficiencies for the G and P mRNAs (1.64 and 13.03, respectively) though neither had a greater proportional change than the average change seen with all mRNAs in this sample. These high translational efficiencies suggest that translation of both P and G mRNAs were upregulated. The G mRNA also had the highest translational efficiency of both time points.

The NS1 mRNA was the only mRNA that had significantly reduced translational efficiency with a value of 0.29, the proportional change was also greater than the average change. This indicates that the NS1 mRNA was translationally downregulated. As seen at 4 hpi, there was a decrease in the translational efficiency (0.47) for the L mRNA though the change was lower than the average change. As described at 4hpi, this low translational efficiency likely means that downregulation of translation for the L mRNA was observed.

For the M, F and M2 mRNAs, the values did not exceed the average change and were close to unity. This suggests that for these mRNAs at this time point there was no translational regulation and the translation was determined primarily by mRNA abundance.

A



B

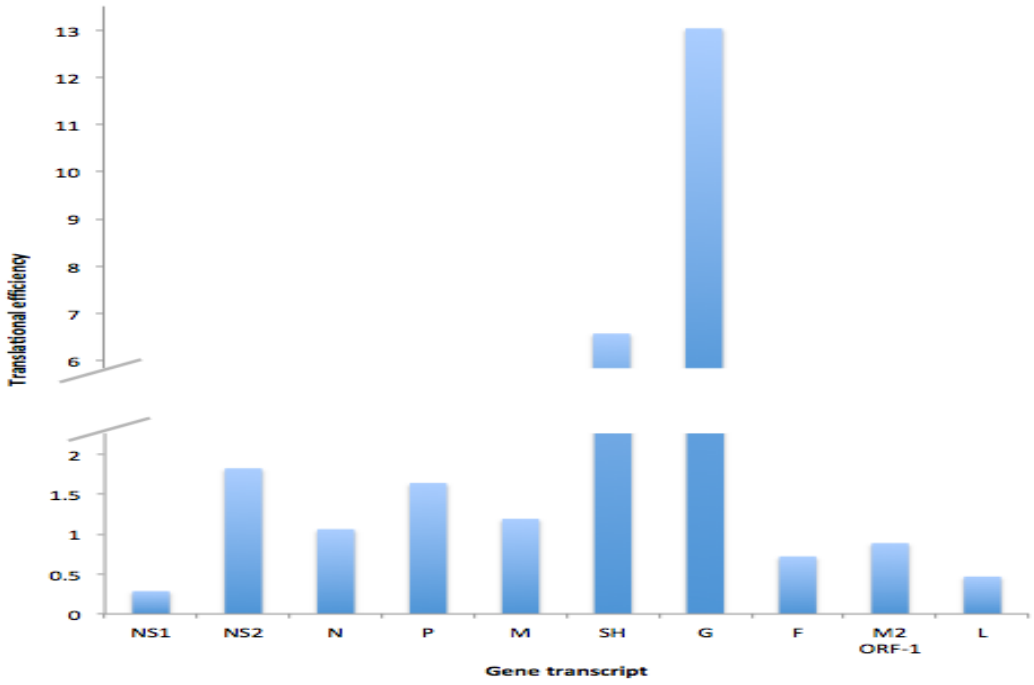


Fig. 3.14: Comparison of proportional translation against proportional mRNA abundance and translational efficiencies of hRSV viral mRNAs harvested at 8 hpi

(A) Adjusted levels of translation as a percentage of total reads aligned to all hRSV viral mRNAs against adjusted levels of mRNA abundance as a percentage of total reads aligned to all hRSV viral mRNAs. The black line represents the points where levels of mRNA abundance and translation were proportionally equal. The red lines represent the mean change in translation from the levels of mRNA abudance. (B) Translational efficiencies of mRNAs from the hRSV sequencing libraries hRSV8P and hRSV8T..

3.5.2.3. Comparison of translational profiles between 4 and 8 hpi for hRSV

A 6.6-fold increase in translational efficiency was seen for the G mRNA, which was the highest increase for all viral mRNAs between the two points. This suggests that there was a significant increase in translation of this mRNA at 8 hpi. There was also an increase in translation efficiency for the SH mRNA from 4 to 8 hpi, with a 4.03-fold increase between the two time points.

The NS1 and the L mRNAs were both poorly translated at both time points and the degree of downregulation increased for both the L and NS1 mRNAs, with a 0.42 and 0.62-fold decrease for NS1 and L mRNA, respectively between 4 and 8 hpi.

It is also interesting that translation of N and NS2 viral mRNAs were clearly upregulated at one time point but not regulated at the other with the N mRNA regulated at 4 hpi and the NS2 mRNA regulated at 8hpi. For the other viral mRNAs, the levels of translational regulation did not exceed the mean change indicating that they were not subject to translational regulation.

3.6. Conclusions

A summary of the levels of mRNA detected for each virus at 4 and 8 hpi are shown in Table 3.4. The transcriptional profiles of both viruses followed similar trends in the levels of mRNA abundance. For genes located between the NS1 and G genes in the genome at 4 hpi, there was a decreasing gradient of mRNA abundance. At 8 hpi, this trend was also observed in both viruses with the exception of the NS2 mRNA abundance which was at lower levels that the adjacent N gene mRNA. This is interesting, as the lower NS2 mRNA level was not observed for both viruses at 4 hpi and may indicate that factors other than the RdRP or that the mRNA half-life are influencing transcription changes for this mRNA at later stages of infection. The hRSV SH mRNA had significantly lower levels of mRNA in comparison to levels observed for the bRSV SH mRNA which may be due to different transcription levels in the two viruses, or that the mRNA half-lives for these two mRNAs differ.

Progressing along the virus genome from the G to M2 genes in both viruses there was a sequential increase in the levels of mRNA abundance in the same order from the 3´ promoter genome at both time points which is not consistent with the polar gradient model of transcription. However, the analysis was based on steady-state levels of mRNA and did not take account of mRNA half-lives, which may play a role in determining the final levels of each mRNA in infected cells.

The relative abundances of the M2 and L mRNAs were consistent with the model of polar transcription. |It was also seen that the abundance of the G mRNA in both viruses were extremely low at 8 hpi. Again this is not consistent with the polar gradient of transcription but may be due to differential mRNA stability.

Virus and time point		Virus gene									
		NS1	**NS2**	**N**	**P**	**M**	**SH**	**G**	**F**	**M2**	**L**
bRSV	4 hr	14.5	15.4	14.7	11.5	10.9	10.2	4.9	5.9	9.6	2.4
	8 hr	18.0	14.1	17.9	15.3	10.8	9.4	0.9	5.4	7.0	1.3
hRSV	4 hr	20.9	16.7	12.4	15.0	12.0	4.3	3.6	6.3	8.3	0.6
	8 hr	40.5	10.5	15.3	8.9	8.9	1.4	0.5	5.7	7.8	0.4

Table 3.4: Summary of mRNA levels at 4 and 8 hpi for bRSV and hRSV
The levels of viral mRNA abundance for each gene are expressed as a percentage of the total virus-specific mRNAs at 4 and 8hpi. Percentage has been adjusted for size of each viral mRNA.

A summary of the translational efficiency of mRNA detected for each virus at 4 and 8 hpi are shown in Table 3.5. In both viruses there appears to be translational regulation occurring for various viral mRNAs. For both bRSV and hRSV translational profiles, only the G mRNA appears to be upregulated in both viruses at both time points, contrasting with the reduction in observed levels of mRNA. There also appears to be upregulation in both viruses throughout the first 8 hr of infection for the P mRNA with the exception for translation of the hRSV P mRNA at 4hpi. There are also examples of upregulation observed for other viral mRNAs at one time point in one virus but not the other. For example this is seen with the NS2 and NS1 mRNAs at 4 hpi for bRSV and the NS2 mRNA at 8 hpi for hRSV.

The L mRNA is the only mRNA from which translation is reduced in both viruses at both time points with this being greater at the later time point. For bRSV, the SH and M2-1 mRNAs were poorly translated at both time points as was the situation for the hRSV NS1 mRNA. It is interesting to note that the SH mRNA was upregulated at both time points for hRSV and downregulated for both time points in bRSV. This may suggest that translational regulation is necessary in both viruses to produce the required levels of the proteins, which could be attributed to the differences in the levels of mRNA abundance between the two viruses.

Virus and time point		Virus gene									
		NS1	**NS2**	**N**	**P**	**M**	**SH**	**G**	**F**	**M2 ORF-1**	**L**
bRSV	4hr	1.22	1.17	0.94	1.3	0.94	0.71	1.63	0.76	0.42	0.50
	8hr	0.92	1.11	0.94	1.27	1.12	0.51	7	0.76	0.55	0.27
hRSV	4hr	0.69	0.88	1.5	1.04	0.85	1.63	1.98	0.75	0.89	0.68
	8hr	0.29	1.86	1.06	1.64	1.19	6.57	13.03	0.72	0.89	0.47

Table 3.5: Summary of translation efficiencies mRNA levels at 4 and 8 hpi for bRSV and hRSV
Translational efficiencies for viral mRNAs for hRSV and bRSV at 4 and 8hpi.

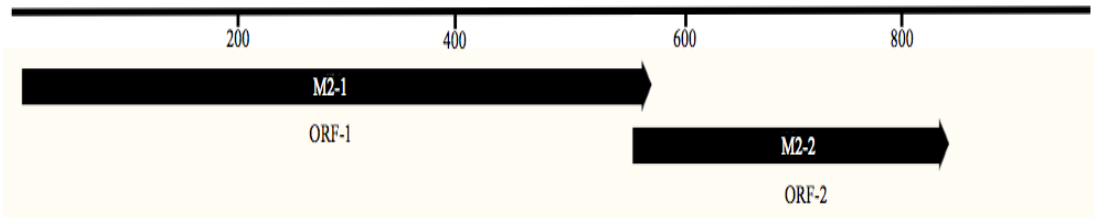
Chapter 4
Characterisation of bRSV M2 ORF-2 Initiation Mechanism

4.1. Introduction

The method of initiation of translation for ORF-2 of the M2 gene transcript of two viruses from the *Orthopneumovirus* genus has been investigated (Ahmadian et al., 2000; Gould and Easton, 2007). For both hRSV and PVM the method of initiation reported was that of coupled translation termination/initiation. There is currently no data on the method of initiation of translation of ORF-2 in the M2 gene transcript of bRSV.

The M2 gene of bRSV is the ninth gene from the 3′ end of the viral genome. The gene is 960 nts in size and is translated into a bicistronic gene transcript, which contains two open reading frames. The two ORFs overlap each other by 17 nts from position 554 to 570 on the M2 gene transcript (Fig. 4.1A). The first ORF (M2 ORF-1) which is 570 nts in size is translated to the protein M2-1. The second ORF (M2 ORF-2), which is 282 nts in size, is translated to the smaller protein M2-2. M2 ORF-2 is in the second reading frame relative to the reading frame of M2 ORF-1 (M2-1 ORF). There are two potential initiation sites located in the M2 ORF-1/ORF-2 overlap region for the initiation of M2 ORF-2 (M2-2 ORF) (Fig. 4.1B) and these are separated by six nucleotides (beginning at positions 554 and 563 from the start of the M2 gene transcript). The second start codon at position 563 is 2 nucleotides from the ORF-1 stop codon. Neither start codon has a perfect Kozak sequence; with the 5′-proximal start codon consisting of UGUUAAUGA and the 3′ start codon consisting of ACAAAAUGA. Although other viruses within this genus that have been investigated can use all of the start codons as potential start codons in their overlapping regions, it is not known in bRSV which of these initiation sites is used.

A



B

549-UGUUAAUGAACAAAAUGAA**UAA**-570

Fig. 4.1: Schematic of bRSV gene transcript and the M2 ORF-1/ORF-2 overlap region
(A) Schematic of the bRSV M2 gene transcript highlighting both ORF-1 (M2 ORF-1) (position 10-570) and ORF-2 (M2 ORF-2) (position 554-835). (B) Sequence of the 3′ end of M2 ORF-1 from position 549-570 of the M2 gene transcript. Sequence includes the M2 ORF-1/ORF-2 overlap region from position 554 to 570. Potential AUG sites for initiation of M2 ORF-2 are underlined beginning at positions 554 and 563. The terminating stop codon for ORF-1 is highlighted in bold at position 570.

For the *Pneumoviridae* family, hRSV has been used as the main model for the investigation of the function of each gene in the genome. This includes the functions of the M2-1 and M2-2 proteins which are involved in the regulation of virus RNA synthesis. The role of the M2-1 protein has been primarily investigated through the use of minigenomes. These studies have shown that the M2-1 protein acts as an antitermination factor during transcription of the viral genome (Collins et al., 1996; Fearns and Collins, 1999b; Hardy and Wertz, 1998). Conversely to the M2-1 protein, the role of the M2-2 protein is in the switch from transcription to replication.

hRSV is the most closely related virus in the *Orthopneumovirus* genus to bRSV in terms of sequence homology of the M2 gene. hRSV uses the mechanism of coupled translation termination/initiation for the initiation of translation of M2 ORF-2 (M2-2 ORF). The nucleotide sequence homology between the two viruses for the entire M2 gene is 68.02% and 76.14% for M2 ORF-1 (M2-1 ORF). Although the M2 ORF-1/ORF-2 overlap regions in both viruses do not share any consensus sequence, there is some sequence homology between the overlap region in the bRSV M2 mRNA and upstream sequences in the hRSV M2 mRNA.

It is unclear what mechanism is used by bRSV for the initiation of translation of the M2 ORF-2. It is possible that bRSV uses the mechanism of coupled translation termination/initiation similar to that used by hRSV. The nucleotide sequence homologies are very high and although the two viruses do not share the same sequences within the overlap region, PVM, which also uses coupled translation termination/initiation for the initiation of M2 ORF-2 (M2-2 ORF), has a completely different overlap region compared to hRSV. This suggests that although the overlap sequence is different this mechanism still may be used in bRSV.

It has been shown that hRSV uses a highly structured region within the M2 ORF-1 in the M2 mRNA as a necessary component for coupled translation (P. Gould and A. Easton, personal communication). Comparison of the sequences has identified a high level of sequence homology (78.45%) between hRSV and bRSV in the region surrounding the highly structured element that is essential for initiation of translation for the hRSV M2 ORF-2 (positions 390-518 in both viruses). This highly homologous sequence may suggest that a similar or related structural region may exist in the bRSV M2 mRNA. If this is the case, it would strengthen the likelihood that bRSV uses a mechanism similar to that used by hRSV to translate the M2 ORF-2. However, it is also possible that a coupled translation mechanism used for the bRSV M2 mRNA may differ from that observed in other members of the *Orthopneumovirus* genus.

The aim of this chapter is to explore the mechanism of initiation of translation of the bRSV M2 ORF-2.

4.2.1. Construction of a bRSV M2 gene reporter construct

To analyse the mechanism of translation of the bRSV M2 ORF-2 it was necessary to generate a reporter plasmid. For this plasmid, the sequence for the M2 gene transcript from bRSV strain Snook was isolated and cloned into a suitable plasmid.

For each different technique used to create reporter plasmids, a detailed methodology can be found in Section 2. Sequences for all primers used in all PCR reactions in all sections can be found in Appendix C.

Total RNA was isolated from 30 µl of a bRSV virus stock preparation, containing $1x10^6$ pfu/ml, using a QIAGEN viral RNA extraction kit and samples were purified according to the manufacturer's instructions (Section 2.6.1). The RNA was reverse transcribed into DNA by RT PCR using random primers for extension (Section 2.6.2). The M2 gene transcript was amplified by PCR (Section 2.3.1) using the primer pair M2 GTF and M2 GTR. The amplicon generated was 985 nts in size, and contained a Kpn I and Sph I site in the 5´ and 3´ ends, respectively. The amplicon was then digested with the endonuclease pair Kpn I/Sph I (Section 2.3.4 and Section 2.3.5) and the isolated DNA fragment was ligated into the vector pBluescribe (Section 2.3.7) which had also been digested with the endonuclease pair Kpn I and Sph I (Section 2.3.4 to Section 2.3.6). The plasmid was then transformed into *E.coli* DH5α bacteria and plated onto AMP/IPTG/X-Gal LB agar plates (Section 2.3.8). The bacteria were grown overnight at 37^{o}C and colonies were selected on the basis of blue/white selection. Each selected colony was grown overnight at 37^{o}C (Section 2.3.9) and small-scale plasmid preparation was carried out (Section 2.3.13). Purified plasmids were digested with endonucleases Kpn I and Sph I to confirm the presence of the correctly ligated insert (967 nts) (Section 2.3.4). The plasmids were sequenced using the primers T7-91 and T3 down to confirm that the sequence was correct (Section 2.3.10). Sequencing results from the plasmid, named pbRSVM2, is in Appendix B2.

4.3.1. Insertion of a reporter gene into the bRSV M2 ORF-2

To allow measurement of the level of expression from the bRSV M2 ORF-2 a reporter gene was inserted into ORF-2, in an analogous way to those used to study the hRSV and PVM systems (Ahmadian et al., 2000; Gould and Easton, 2007). This approach was based on the premise that translation of M2 ORF-2 was not reliant on sequence downstream of the ORF-1/ORF-2 overlap region. Therefore it was possible to replace the ORF-2 sequence with a reporter gene that could be quantified when expressed.

The reporter plasmid pbRSVWC (Fig. 4.2) was generated to measure bRSV M2 ORF-2 expression. The reporter plasmid contained a T7 promoter, which allowed a T7 RNA polymerase to transcribe the gene in the plasmid. This reporter plasmid contained a gene that was transcribed into a bicistronic gene transcript. The gene transcript transcribed, was in the same format as the bRSV M2 gene transcript. ORF-1 of the pbRSVWC gene transcript contained the bRSV M2 ORF-1 sequence. This also included the sequence for the bRSV M2 ORF-1/ORF-2 overlap region. This overlap region contained the initiation sites used for the translation of ORF-2 in the mRNA. The coding sequence of the CAT reporter gene, lacking the CAT AUG initiation codon, was fused downstream of the overlap region. This sequence replaced the M2 ORF-2 sequence downstream of the overlap region and was in the same reading frame as the M2 ORF-2 sequence to create the ORF-2 for the plasmid's bicistronic mRNA. Following transfection of the plasmid into mammalian cells, the bicistronic mRNA would be transcribed by a T7 RNA polymerase through initiation at the T7 promoter. The T7 RNA polymerase was expressed by infecting the mammalian cells with vaccinia VTF7-3 virus that carried the T7 RNA polymerase gene. Initiation of translation of ORF-2 would occur in the ORF-1/ORF-2 overlap region and therefore translation of ORF-2 would be controlled by the bRSV M2 gene transcript sequence. This led to the reporter CAT protein being translated, which could then be quantified using an ELISA assay. The mRNA transcribed from pbRSVWC also contained a 30 nt 5′ UTR, of which the last 9 nts were from the bRSV M2 5′ UTR; the additional UTR sequence came from the plasmid vector.

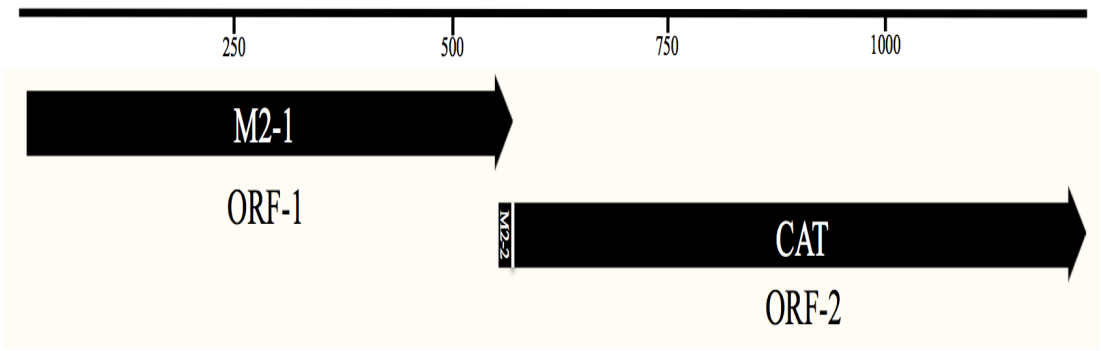


Fig. 4.2: Schematic of the gene transcript transcribed from the plasmid pbRSVWC
Schematic representation of the gene transcript transcribed from the plasmid pbRSVWC. Including the two ORFs, ORF-1 (M2 ORF-1) and ORF-2 (M2 ORF-2/CAT ORF) which includes the ORF-1/ORF-2 overlap region. For all diagrams M2-1 represents M2 ORF-1 sequence, M2-2 represents M2 ORF-2 sequence and CAT represents CAT ORF sequence.

Construction of the plasmid pbRSVWC

To construct the reporter plasmid pbRSVWC, two fragments were generated by PCR. The first fragment (bRSV fragment), was used to create the last 9nts of the 5′ UTR and the entire of ORF-1 in the reporter plasmid. This also included the ORF-1/ORF-2 overlap region .The second fragment (CAT fragment), was used to fuse the CAT coding sequence downstream of the ORF-1/ORF-2 overlap region to create ORF-2. The two fragments were ligated together to create the insert for the reporter plasmid. The bRSV fragment was generated by PCR using the primer pair bRSVWTF/bRSVWTR from the plasmid pbRSVM2 (Section 2.3.1). This amplicon included the M2 gene transcript sequence from the 5′ end of the M2 gene transcript to the 3′ end of the ORF-1/ORF-2 overlap region (residues 1-570 of the bRSV M2 gene transcript). The 595 nts amplicon contained a Kpn I site and Xba I site at its 5′ and 3′ ends, respectively. The CAT fragment was amplified using the primer pair CATF/CATR from the plasmid pWildCAT (Section 2.3.1). The amplicon included the CAT coding sequence from residue 4-660 from the plasmid pWildCAT. The amplicon, which was 678 nts in size contained an Xba I and an Sph I site at its 5′ and 3′ ends respectively. The bRSV fragment was then digested with the endonuclease pair Kpn I/Xba I and the CAT fragment was digested with the endonuclease pair Xba I/Sph I (Section 2.3.4 and Section 2.3.5). The two fragments were double ligated into the plasmid pBluescribe (Section 2.3.7), which had been digested with endonucleases Kpn I and Sph I (Section 2.3.4 to Section 2.3.6). The ligated plasmid was then transformed into *E.coli* DH5α bacteria and plated onto AMP/IPTG/X-Gal

LB agar plates (Section 2.3.8). The bacteria were grown overnight at 37^{o}C and colonies were selected using blue/white selection (Section 2.3.9). Each picked colony was grown overnight at 37^{o}C and small-scale plasmids purification was performed as described in Section 2.3.13. Purified plasmids were digested with endonucleases Kpn I and Sph I to confirm the presence of the correctly ligated insert of 1237 nts. Plasmids were sequenced using the primers T7-91 and T3 down to confirm the expected sequence as described in Section 2.3.10.

4.3.2. Expression from the bRSV M2 ORF-2 in HEp-2 cells

HEp-2 cells were individually transfected with either plasmid pbRSVWC, plasmid pWildCAT or empty pBluescribe vector plasmid as described in Section 2.4.1. The plasmid pWildCAT contained the hRSV M2 gene with the CAT reporter gene in place of the M2 ORF-2 (Ahmadian et al., 2000) and was provided by Dr Phillip Gould. This plasmid was the original model for the reporter system for pneumovirus coupled translation and was used as a positive control. Cells were infected with the vaccinia VTF7-3 virus. This virus expressed the T7 polymerase protein, which was used to transcribe gene transcripts from the transfected plasmids. Transfected cells were incubated for 48 hours and then lysed and cellular material harvested. An ELISA assay was then performed on the cellular lysate to measure the levels of CAT protein expression as described in Section 2.4.2. All samples were normalised against the assay results from cells transfected with empty vector pBluescribe plasmid. This provided the background signal generated in the ELISA. For each sample, levels of CAT protein expressed have been visualised as a percentage of the levels achieved with the control plasmid pWildCAT.

For all transfected plasmids in this thesis, each transfection for an individual plasmid was repeated three times. Each repeat was the mean of three separate transfections of each plasmid. Therefore a total of nine separate transfections were performed for each plasmid. The mean of all three repeats was calculated and used as the value presented for the level of CAT protein expression expressed by each plasmid. The standard deviation was calculated from the values of the three repeats.

As shown in Fig. 4.3, plasmid pbRSVWC expressed levels of CAT protein from ORF-2 at 19.17% of that of the control plasmid pWildCAT. The expression of CAT

protein shows that there was successful initiation of translation of ORF-2 and that the reporter system can be successfully used to investigate the mechanism of initiation of translation for ORF-2 of the M2 gene transcript in bRSV. The data also indicate that in the bRSV M2 gene transcript, sequences within ORF-2 downstream of the ORF-1/ORF-2 overlap region are not critical for initiation of translation. However, it is possible that sequences within ORF-2 may enhance levels of expression.

There was a statistically significant difference in the levels of expression observed between the hRSV plasmid pWildCAT and the bRSV plasmid pbRSVWC ($p=0.013$). It has also been observed that PVM expresses lower levels of CAT reporter protein in this system than that of hRSV (Gould and Easton, 2007).

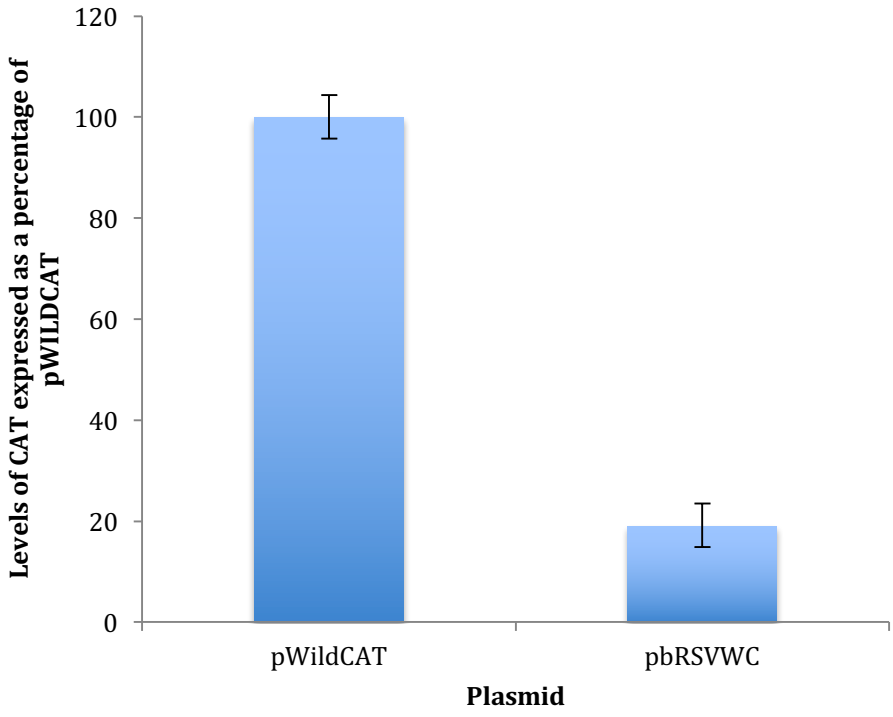


Fig. 4.3: CAT protein expression from the bRSV M2 ORF-2
Levels of expression of CAT protein from plasmids pWildCAT and pbRSVWC individually transfected into HEp-2 cells. Levels of expression were normalised against the empty vector pBluescribe and are expressed as a percentage of the levels detected with the positive control plasmid pWildCAT. Error bars indicate standard deviation.

4.4.1. Identification of the initiation site for translation of the bRSV M2 ORF-2

The availability of a reporter system to measure the level of translation for the bRSV M2 ORF-2 in Section 4.3, allowed the creation of mutants to investigate the mechanism by which M2 ORF-2 is accessed for translation. Two initiation codons are located in the ORF-1/ORF-2 overlap region of the M2 mRNA that could be the sites of initiation of translation for ORF-2 (beginning at positions 554 and 563 from the M2 gene transcript). It is currently unknown if either or both start codons act as the initiation codon for initiation of translation of M2 ORF-2. As a first step in establishing the mechanism of translation initiation, mutations were introduced into the ORF-1/ORF-2 overlap region to mutate each of the potential start codons to determine if either or both start codons were used.

The mutations were introduced into the plasmid pbRSVWC, (Fig. 4.4). The plasmid pbRSV5´AUG had the AUG beginning at position 554 and located at the 5´ end of the ORF-1/ORF-2 overlap region mutated by mutating AUG to CUG. In a similar way the plasmid pbRSV3´AUG had the 3´ AUG located at position 563 mutated by mutating AUG to CUG. The plasmid pbRSVDouble had both start codons mutated by mutating both AUGs in the overlap region to CUG.

bRSV ORF-1/ORF-2 overlap region AUGAACAAAAUGAA**UAA**
pbRSV5´AUG CUGAACAAAAUGAA**UAA**
pbRSV3´AUG AUGAACAAACUGAA**UAA**
pbRSVDouble CUGAACAAACUGAA**UAA**

Fig. 4.4: Mutations made to reporter plasmids in the ORF-1/ORF-2 overlap region
Mutations found in the ORF-1/ORF-2 overlap region for the plasmids pbRSV5´AUG (A to C position 554), pbRSV3´AUG (A to C position 563) and pbRSVDouble (A to C position 554 and 563). Sites of mutations are highlighted in red. Potential AUG sites for initiation of ORF-2 are underlined. The terminating stop codon for ORF-1 is highlighted in bold.

Construction of plasmids pbRSV5´AUG, pbRSV3´AUG and pbRSVDouble

For all three plasmids the method for construction was identical except for the reverse primer used to generate the mutations in the bRSV fragment. For each plasmid two fragments were generated by PCR that were ligated together as the insert into the vector plasmid. The bRSV fragment, which comprised of the bRSV M2 5´ UTR and the entire of the bRSV M2 ORF-1 sequence, was amplified from the

plasmid pbRSVWC by PCR using the forward primer bRSVWTF and the respective reverse primer (Table 4.1), which incorporated the specific mutation (Section 2.3.1). For each plasmid, the 594 nts amplicon contained a Kpn I site and a Xba I site at its 5′ and 3′ ends, respectively. The CAT fragment comprised of the reporter gene CAT and was amplified from the plasmid pbRSVWC and using the primer pair CATF and CATR (Section 2.3.1). The 678 nts amplicon contained an Xba I and an Sph I site at its 5′ and 3′ ends, respectively. The bRSV fragment was digested with the endonuclease pair Kpn I and Xba I and the CAT fragment was digested with the endonuclease pair Xba I and Sph I (Section 2.3.4 and Section 2.3.5). The two fragments were double ligated into the plasmid pBluescribe (Section 2.3.7), which had been digested with endonuclease pair Kpn I and Sph I (Section 2.3.4 to Section 2.3.6). The ligated plasmid was then transformed into *E.coli* DH5α bacteria and plated onto AMP/IPTG/X-Gal LB agar plates (Section 2.3.8). The bacteria were grown overnight at 37^{o}C and colonies were selected using blue/white selection (Section 2.3.9). Each selected colony was grown overnight at 37^{o}C and small scale plasmid preparations were carried out (Section 2.3.13). Purified plasmids were digested with endonucleases Kpn I and Sph I to confirm the presence of the correctly ligated insert 1237 nts. After confirmation of the correct sized fragments, plasmids were sequenced using the primers T7-91 and T3 down to confirm the presence of the correct sequence as described in Section 2.3.10.

Plasmid	Forward Primer	Reverse Primer
pbRSV5′AUG	bRSVWTF	5′AUGR
pbRSV3′AUG	bRSVWTF	3′AUGR
pbRSVDouble	bRSVWTF	DoubleR

Table 4.1: Primers used in PCR reaction for bRSV fragment for generating the plasmids pbRSV5′AUG, pbRSV3′AUG and pbRSVDouble
Forward and reverse primers used to generate each bRSV amplicon for the plasmids pbRSV5′AUG, pbRSV3′AUG and pbRSVDouble. Sequences for primers are detailed in Appendix C.

4.4.2. Expression of ORF-2 AUG mutations in HEp-2 cells

The three expression plasmids containing mutated ORF-2 AUG initiation codons were individually transfected into HEp-2 cells. Control cells were individually transfected into HEp-2 cells with either pbRSVWC or the empty vector pBluescribe. Plasmids were transfected into HEp-2 cells as described in Section 2.4.1. The

plasmid pbRSVWC acted as the positive control. As stated before, levels of CAT protein expressed were quantified using an ELISA assay as described in Section 2.4.2. All samples were normalised against the assay results from transfected pBluescribe plasmid. This provided the background signal generated in the ELISA. For each sample, levels of CAT protein expressed were displayed as a percentage of the level obtained with plasmid pbRSVWC.

As shown in Fig. 4.5, mutation of both start codons in the overlap region in the plasmid pbRSVDouble reduced levels of CAT protein expression of ORF-2 to 0.17% of the control. This significant difference ($p=0.003$) from the control indicates that initiation of translation of the bRSV M2 ORF-2 is completely inhibited and the site for initiation is located within the M2 ORF-1/ORF-2 overlap. This is also suggests that there is no non-AUG initiation for M2 ORF-2. Mutation of the start codon beginning at position 554 in the plasmid pbRSV5´AUG increased levels of CAT protein expression from ORF-2 to 104.9% of the control pbRSVWC. Although the level of expression was higher compared to the control there was no statistical difference between the two ($p=0.8$). This was likely due to the high standard deviation (STD) observed through repeated triplicates. These data indicates that the AUG located in the overlap region beginning at the position 554 is not the start codon for initiation of translation of the M2 ORF-2. Mutation of the start codon beginning at position 563 in the plasmid pbRSV3´AUG reduced levels of CAT protein expression to 5.4% of the control. This significant difference compared to the control ($p=0.003$) indicates that this start codon is the primary site for initiation of translation of the second ORF. There was also a significant difference between the level of expression of pbRSVDouble and pbRSV3´AUG ($p=0.002$) indicating that translation was not fully inhibited by the mutation of the start codon beginning at position 563. This may suggest that a small degree of initiation can take place at the alternate start codon beginning at position 554.

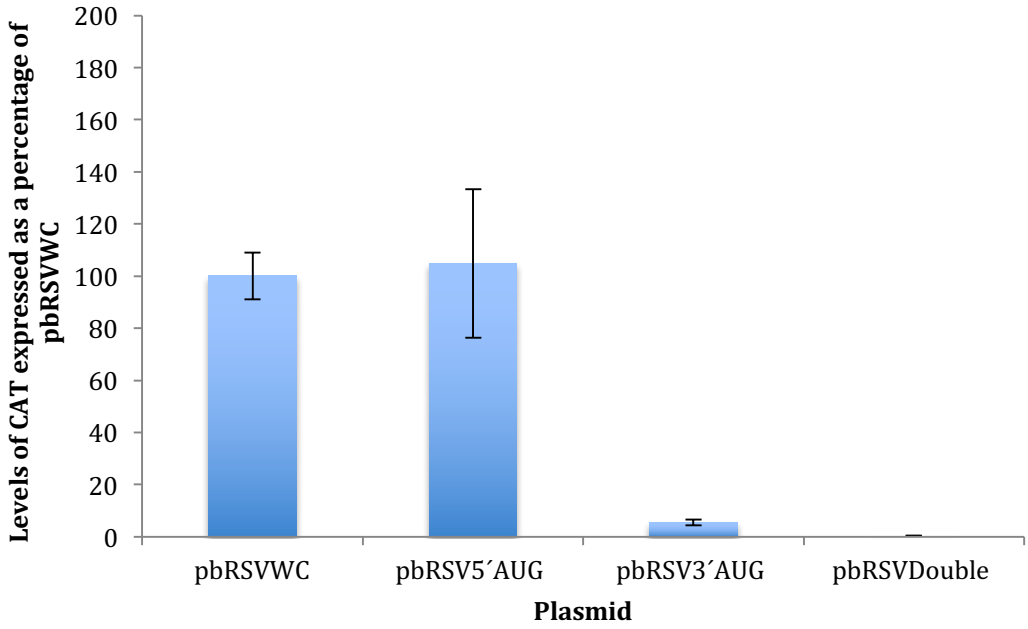


Fig. 4.5: CAT reporter protein expression from ORF-2 AUG initiation codon mutants
Levels of expression of CAT protein expressed from wild type and mutant plasmids transfected into HEp-2 cells. Mutations made to these plasmids are described in Section 4.4.1. Results were normalised against the empty vector pBluescribe and are expressed as a percentage of the level achieved with the positive control plasmid pbRSVWC. Error bars indicate standard deviation.

4.5.1.1. Analysis of translation initiation of the bRSV M2 ORF-2

The close phylogenetic relationship of the members of *Orthopneumovirus* genus raised the possibility that the mechanism for initiation of the M2 ORF-2 used by bRSV was also coupled translation termination/initiation. For hRSV and PVM the mechanism is reliant on the termination of elongating ribosomes that are translating M2 ORF-1 which then reinitiate translation of the M2 ORF-2 following translocation upstream from the M2 ORF-1 stop codon. Mutation of the M2 ORF-1 terminating stop codon leading to termination at a point further in the 3′ direction reduces protein expression to negligible levels in a length dependent manner (Ahmadian et al., 2000; Gould and Easton, 2007).

To determine if translation in the bRSV M2 ORF-2 uses a process of coupled translation termination/initiation, the plasmid pbRSVStop was generated. This plasmid, which was in the same format as the reporter plasmid pbRSVWC, had the ORF-1 (M2 ORF-1) terminating stop codon at position 568 mutated from UAA to CAA. The nearest in-frame 3′ stop codon downstream was 42 nts away. This increased the size of the ORF-1/ORF-2 overlap region to 50 nts (if overlap region began at position 563 at the preferred initiation codon). If coupled translation was used as the mechanism for the initiation of M2 ORF-2, then the additional distance would have a large negative effect on CAT protein expression.

Construction of the plasmid pbRSVStop

For the construction of the plasmid bRSVStop, the method for cloning and sequencing of the plasmid was the same as used in Section 4.3.1. The bRSV fragment for the plasmid was amplified by PCR from the plasmid pbRSVWC using the primer pair bRSVWTF/bRSVSTOP. The second fragment CAT fragment was also amplified by PCR from the plasmid pbRSVWC, using the primer pair CATF/CATR.

4.5.1.2. Expression of plasmid with M2 ORF-1 terminating stop codon mutation in HEp-2 cells

HEp-2 cells were transfected with either plasmid, pbRSVStop pbRSVWC or empty pBluescribe vector plasmid as described in Section 2.4.1. The plasmid pbRSVWC

acted as the positive control. Levels of CAT protein expressed were quantified using a CAT ELISA assay as described in Section 2.4.2. All samples were normalised against the assay results from transfected pBluescribe plasmid that provided a background signal generated in the ELISA. For each sample, levels of CAT protein expressed were displayed as a percentage of the levels of CAT protein expressed from the plasmid pbRSVWC.

As shown in Fig. 4.6, mutation of the ORF-1 (M2 ORF-1) termination codon at position 568 in the plasmid pbRSVStop reduced levels of CAT protein expression to 93.3% of the control. There was no significant difference in expression of the CAT protein between the plasmid pbRSVStop and the control ($p=0.54$), indicating that mutation of the M2 ORF-1 stop codon had no significant effect on the initiation of translation of M2 ORF-2. This indicates that, in contrast to the situation for hRSV and PVM, coupled translation termination/initiation is not used for initiation of translation of the bRSV M2 ORF-2.

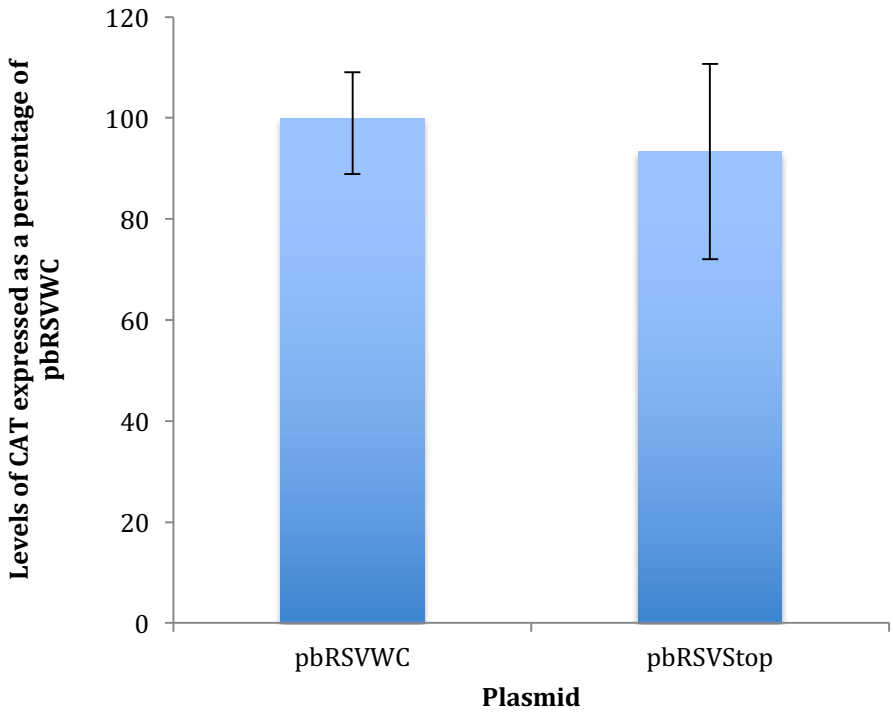


Fig. 4.6 CAT protein expression from M2 ORF-1 terminating stop codon mutants
Levels of expression of the CAT protein from the plasmid pbRSVSTOP transfected into HEp-2 cells. Mutations made to this plasmid are described in Section 4.5.1.1. Results were normalised against the empty vector pBluescribe and are expressed as a percentage of the level achieved with positive control plasmid pbRSVWC. Error bars indicate standard deviation.

4.5.2.1. Effect of increasing the length of the M2 ORF-1 on initiation of translation of the bRSV M2 ORF-2

In Section 4.5.1, it was shown that mutation of the M2 ORF-1 terminating stop codon that resulted in the length of the M2 ORF-1 increasing by 42nts, did not significantly alter the levels of CAT protein expression from ORF-2. This suggests that the mechanism used by bRSV is not that of coupled translation termination/initiation. However, as the process of coupled translation termination/initiation is length dependent and the distance from the M2 ORF-1 termination codon and the M2 ORF-2 initiation codon is important, it is possible that the extension of the ORF-1/ORF-2 overlap region in the plasmid pbRSVStop was too short to prevent reinitiation. In the equivalent hRSV and PVM systems used to test for coupled translation, the M2 ORF-1 terminating stop codon was mutated and the M2 ORF-1 lengths were extended by 40 nts (Ahmadian et al., 2000) and 138 nts (Gould and Easton, 2007) respectively. These increased lengths of M2 ORF-1 prevented ORF-2 expression. Although the M2 ORF-1 in pbRSVStop was extended by 42nts, this may have not been a large enough distance for a ribosome to fail to reinitiate. This may be due to the close proximity of the preferred start codon to the mutated stop codon, resulting in a smaller distance for the reinitiating ribosome to scan compared to plasmids used in hRSV and PVM.

In the equivalent study in hRSV (Ahmadian et al., 2000), two expression plasmids containing the mutated ORF-1 (M2 ORF-1) stop codon were used to test for coupled translation. As mentioned above, both of the plasmids had the ORF-1/ORF-2 overlap region extended by 40 nts. In a feature unlike the plasmid pbRSVStop, which had the ORF-1/ORF-2 overlap region extended with CAT reporter coding sequence, the two plasmids used for hRSV had overlap regions increased with hRSV M2 ORF-2 sequence that was taken from downstream of the hRSV M2 ORF-1/ORF-2 overlap region. For the first plasmid, an additional 44 nts of hRSV M2 ORF-2 sequence was added. For the second plasmid an additional 120 nts of hRSV M2 ORF-2 sequence was added resulting in an extra 80 nts of hRSV M2 ORF-2 sequence added past the plasmids ORF-1 stop codon. Any stop codons in the same frame as ORF-1 in the additional sequences were mutated to ensure that the ORF-1/ORF-2 overlap region was extended as desired and in both plasmids, no expression of ORF-2 was detected. However, it is important to note that no controls were used for either plasmid with

the addition of equivalent M2 ORF-2 sequences where the controls would have the ORF-1 terminating stop codon present. It is therefore possible that as well as the extension of the ORF-1/ORF-2 overlap region being too short to inhibit the potential mechanism of coupled translation in the plasmid pbRSVStop, additional M2 ORF-2 sequence may be required for the mechanism to operate.

4.5.2.2. Analysis of length dependence in the bRSV ORF-1/ORF-2 overlap region

To investigate the importance of length dependence in the bRSV ORF-1/ORF-2 overlap region and the possible role of the 5′ proximal M2 ORF-2 sequences for initiation of translation of M2 ORF-2, four additional plasmids were created. These plasmids were in a similar format to the two hRSV plasmids used to test for coupled translation of hRSV M2 ORF-2 as described in Section 4.5.2.1. The newly synthesised bRSV-based plasmids had additional bRSV M2 ORF-2 sequence added between the 3′ end of the ORF-1/ORF-2 overlap region and the 5′ end of the CAT reporter coding sequence in ORF-2. The constructs used the CAT reporter system as before to measure ORF-2 expression. Two of the newly generated plasmids, pbRSVWC42 and pbRSVStop42, had an additional 42 nts of bRSV M2 ORF-2 (residues 571-612 from the beginning of the bRSV M2 gene transcript) added (Fig. 4.7A), whilst plasmids pbRSVWC120 and pbRSVStop120 had an additional 120nts of bRSV M2 ORF-2 (residues 571-690 from the beginning of the bRSV M2 gene transcript) added (Fig. 4.7B). In the plasmid nomenclature, “WC” refers to the control in which no mutations were made to the terminating ORF-1 (M2 ORF-1) stop codon. “Stop” refers to equivalent plasmids where the ORF-1 stop codon has been mutated from UAA to CAA. The number refers to the amount of bRSV M2 ORF-2 sequence that was added downstream of the ORF-1/ORF-2 overlap region. All plasmids had any stop codons present in the additional M2 ORF-2 sequence that were in the same frame as ORF-1 mutated out.

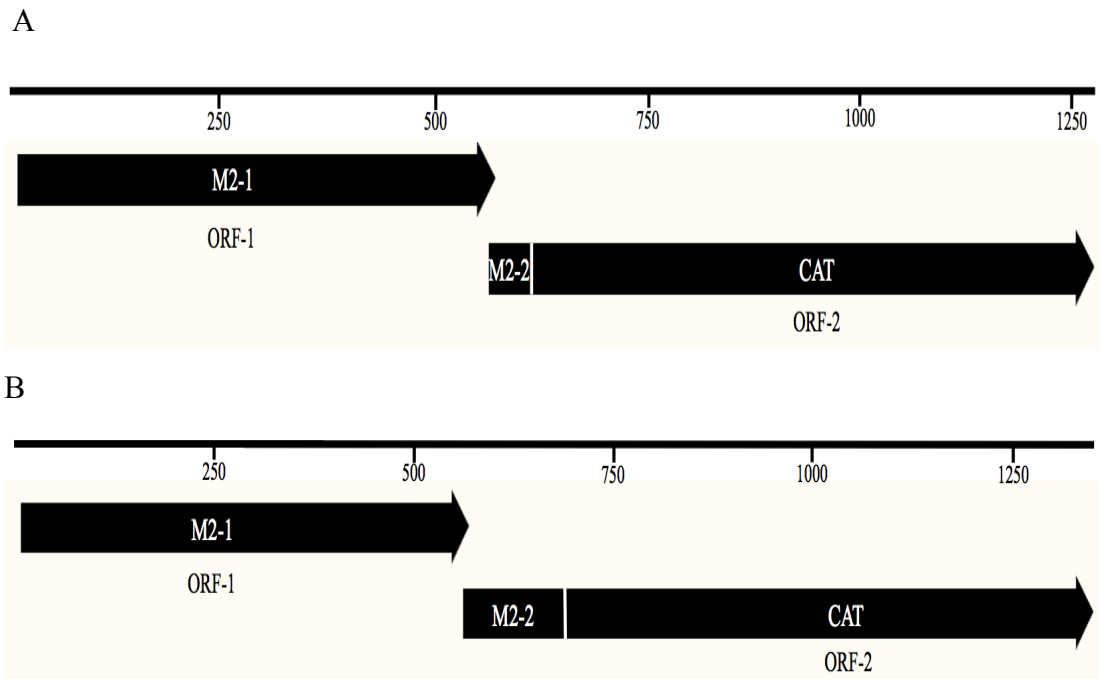


Fig. 4.7: Schematic of gene transcripts for plasmids used to investigate the effect of increasing the size of the M2 ORF-1/ORF-2 overlap region on the translation of ORF-2

Schematic representation of the gene transcript transcribed from the plasmids pbRSVWC42 and pbRSVWC120. Including both ORFs for each plasmid. The second ORF also includes the additional ligated M2 ORF-2 sequence. (A) Schematic for the plasmids pbRSVWC42, where 42nts of additional M2 ORF-2 sequence was ligated to the 5′ end of the CAT coding sequence. Stop codons in the same frame as ORF-1 (M2 ORF-1) in this sequence were removed. (B) Schematic for the plasmid pbRSVWC120, where 120nts of additional M2 ORF-2 sequence was ligated to the 5′ end of the CAT coding sequence. Stop codons in the same frame as ORF-1 (M2 ORF-1) in this sequence were removed.

The plasmids pbRSVStop42 and pbRSVStop120 were used to investigate if increasing the size of the ORF-1/ORF-2 overlap region by 84 nts and 162 nts respectively affected the levels of ORF-2 CAT protein expression. Significant reduction or inhibition of CAT protein expression would suggest that the mechanism of coupled translation is used.

For each of the plasmids pbRSVStop42 and pbRSVStop120, a control was generated that was identical with the exception of the ORF-1 terminating stop codon being present in the usual position (plasmids pbRSVWC42 and pbRSVWC120 for the plasmids pbRSVStop42 and pbRSV120 respectively). These were used to identify if a significant change in expression was registered when the ORF-1 terminating stop codon was mutated and the ORF-1/ORF-2 overlap region was extended. These plasmids were also used to investigate if additional M2 ORF-2 sequences improved the efficiency of initiation of translation in comparison to the reporter plasmid pbRSVWC where only the M2 ORF-2 sequence in the ORF-1/ORF-2 overlap region was present.

Construction of plasmids pbRSVWC120 and pbRSVStop120

Inserts for plasmids pbRSVWC120 and pbRSVStop120 were generated by a three-step PCR. This allowed the introduction of multiple point mutations to mutate stop codons in the additional 120 nts M2 ORF-2 region that were in the same frame as ORF-1 in the plasmid. Initially two amplicons were generated by PCR from the plasmid pbRSVWC for each plasmid. One was a bRSV fragment, which contained the bRSV M2 5′ UTR and the entire bRSV M2 ORF-1 sequence. The second was a fragment, which contained the CAT reporter coding sequence. For the bRSV fragment, the reverse primer that would anneal to the ORF-1/ORF-2 overlap region contained an additional 38nts of M2 ORF-2 sequence with the appropriate point mutations. For the CAT fragment the same principle was used except the forward primer was used to add additional M2 ORF-2 sequence to the 5′ end of the fragment. A second PCR step was performed by using each fragment as a template to increase the length of the M2 ORF-2 sequence by 42 nts to either the 3′ or 5′ end of the bRSV and CAT fragments respectively. By the second PCR step, the two PCR fragments contained a region of sequence homology located at the 3′ and 5′ ends for

the bRSV and CAT fragments respectively. A third PCR step was performed where both fragments were placed in the same reaction. This generated a single amplicon where the two fragments were joined due to the region of sequence homology. External primers were used to amplify the product.

Each plasmid had a separate bRSV fragment generated with the additional M2 ORF-2 sequence added to the fragment through the three PCR stages. As well as the additional mutations and the addition of the M2 ORF-2 in the first and second PCR steps, each of these fragments had either the ORF-1 stop codon retained or mutated as necessary. The same CAT fragment was used for both plasmids that would have the additional M2 ORF-2 sequence added to the fragment through the first two PCR stages.

The primers used for each PCR step for each plasmid can be seen in Table 4.2 and the sequences are in Appendix C. For all PCR steps the same PCR protocol was used (Section 2.3.1). Products from the previous stage were added in equimolar quantities. The template used for each plasmid at each PCR step can be seen in Table 4.3. After each stage of PCR, the amplicon was purified using a QIAGEN PCR clean up kit before being used for the next PCR step (Section 2.3.15). At the final PCR step for each plasmid, the appropriate bRSV fragment generated from the second PCR step and the CAT fragment generated in the second PCR step were used as templates to join the two fragments together. The primer pair bRSVF/CATR was used to amplify the joined fragments for both reactions. Both final joined amplicons generated by PCR contained a Kpn I and Sph I sites at the 5′ and 3′ ends respectively. The joined fragments for each plasmid were then digested with the endonuclease pair Kpn I/Sph I (Section 2.3.4 and Section 2.3.5). The fragments were separately ligated into the plasmid pBluescribe (Section 2.3.7) which was digested with endonuclease pair Kpn I/Sph I (Section 2.3.4 to Section 2.3.6). The ligated plasmids were then separately transformed into *E.coli* DH5α bacteria and plated onto AMP/IPTG/X-Gal LB agar plates (Section 2.3.8). The bacteria were grown overnight at 37^{o}C and colonies were selected using blue/white selection (Section 2.3.9). Each picked colony was grown overnight at 37^{o}C and small-scale plasmid preparations prepared (Section 2.3.13).Purified plasmids were digested with endonucleases Kpn I and Sph I to confirm the presence of the correctly ligated insert (1457 nts). After confirmation of

the correct sized fragments each plasmid was sequenced using primers T7-91 and T3 down to ensure the mutations had been inserted (Section 2.3.10).

PCR Step	Primers for bRSV fragment for plasmid pbRSVWC120	Primers for bRSV fragment for plasmid pbRSVStop120	Primers for CAT fragment
1st	bRSVF/bRSV120R1WC	bRSVF/bRSV120R1ST	CAT120F1/CATR
2nd	bRSVF/bRSV120R2	bRSVF/bRSV120R2	CAT120F2/CATR
3rd	bRSVF	bRSVF	CATR

Table 4.2: Primers used for three-step PCR for the construction of plasmids pbRSVWC120 and pbRSV120
Primers used for each round of three-step PCR used to create the final insert for plasmids pbRSVWC120 and pbRSVStop120.

PCR Step	Template used for each PCR stage for the bRSV fragment for the plasmid pbRSVWC120	Template used for each PCR stage for the bRSV fragment plasmid for the plasmid pbRSVStop120	Template used for each PCR stage for CAT fragment
1st	pbRSVWC	pbRSVWC	pbRSVWC
2nd	Amplicon generated above	Amplicon generated above	Amplicon generated above
3rd (joining of bRSV and CAT fragments)	Amplicon generated above + CAT fragment generated in 2nd stage of PCR	Amplicon generated above + CAT fragment generated in 2nd stage of PCR	

Table 4.3: Templates used for three-step PCR for the construction of the plasmids pbRSVWC120 and pbRSV120
Templates used for each round of three-step PCR used to create the final insert for plasmids pbRSVWC120 and pbRSVStop120.

Construction of plasmids pbRSVWC42 and pbRSV42

For each plasmid two fragments were generated by PCR that would act as inserts to be ligated into the new vector. The bRSV fragment, containing the entirety of bRSV sequence to be ligated into the new plasmids, was amplified by PCR using the primer pair bRSVWTF/bRSV42 (Section 2.3.1). This was amplified from either the plasmid pbRSVWC120 or pbRSVStop120 as required depending on whether the terminating stop codon for ORF-1 needed to be kept or removed. Each amplicon contained a Kpn I and Xba I site at the 5′ and 3′ ends, respectively. The CAT fragment used for each plasmid contained the CAT reporter coding sequence and was amplified by PCR with the primer pair CATF/CATR from the plasmid pbRSVWC120 (Section 2.3.1). The amplicon contained an Xba I and a Kpn I site at its 5′ and 3′ ends, respectively. The bRSV fragment was then digested with the endonuclease pair Kpn I and Xba I and the CAT fragment was digested with the endonuclease pair Xba I and Sph I (Section 2.3.4 and Section 2.3.5). The two fragments for each plasmid were simultaneously ligated into the plasmid pBluescribe (Section 2.3.7) that had been digested with endonuclease pair Kpn I and Sph I (Section 2.3.4 to Section 2.3.6). The ligated plasmid were transformed separately into *E.coli* DH5α bacteria and plated onto AMP/IPTG/X-Gal LB agar plates (Section 2.3.8). The bacteria were grown overnight at 37^{o}C and colonies were selected through blue white selection (Section 2.3.9). Each picked colony was grown overnight at 37^{o}C and small-scale plasmid preparations were performed (Section 2.3.13). Purified plasmids were digested with endonucleases Kpn I and Sph I to confirm the presence of the correctly ligated insert (1321nts). Upon confirmation of the correct sized fragments, plasmids were sent off for sequencing using the primers T7-91 and T3 down (Section 2.3.10). This allowed sufficient coverage for the entire insert sequence to ensure the correct sequence was ligated.

4.5.2.3. Expression of plasmids used to investigate the effect of increasing the length of the ORF-1/ORF-2 overlap region in HEp-2 cells

The four expression plasmids described in Section 4.5.2.2 were transfected separately into HEp-2 cells as described in Section 2.4.1. Control cells were transfected with either pbRSVWC, pbRSVStop or the empty vector pBluescribe. The levels of CAT protein expressed were quantified using a specific ELISA assay as

described in Section 2.4.2. All data were normalised against the assay results from cells transfected with pBluescribe plasmid that provided the background signal generated in the ELISA. The plasmids pbRSVWC and pbRSVStop acted as positive controls. For each sample levels of CAT protein were expressed as a percentage of the levels of CAT protein expressed from the plasmid pbRSVWC.

As seen Fig. 4.8, addition of an extra 42 nts of bRSV M2 ORF-2 (M2-2 ORF) sequence downstream of the ORF-1/ORF-2 overlap region in the plasmid pbRSVWC42 reduced levels of CAT protein expression to 38% of the control. This significant reduction in comparison to the control ($p=0.0003$) indicated that the M2 ORF-2 sequence past the M2 ORF-1/ORF-2 overlap region did not enhance initiation of translation of the bRSV M2 ORF-2. The reduction in CAT protein expression is likely to be due to the addition of the 42 extra nts to ORF-2 creating a new basal level of CAT protein expression for this newly generated plasmid.

Mutation of the ORF-1 (M2 ORF-1) stop codon at position 568 and increasing the size of the overlap region to 92 nts in size, with the addition of 42 nts of bRSV M2 ORF-2 sequence downstream of the ORF-1/ORF-2 overlap region in the plasmid pbRSVStop42 reduced levels of CAT protein expression to 33.5% of the control. Although there was a significant difference between this and the control ($p=0.0002$), expression of CAT protein was not reduced to the very low levels seen in the equivalent plasmid constructs for analysis of hRSV and PVM M2 ORF-2 expression. The significant reduction in expression however, is not due to a decrease in the ORF-2 initiation efficiency, as there is no significant reduction between the expression of the plasmid pbRSVStop42 and the plasmid pbRSVWC42 ($p=0.097$); the plasmid pbRSVWC42 was the relevant control and expressed CAT protein at basal levels for the new format of plasmid. Taken together with the previous results it is therefore likely that coupled translation termination/initiation is not the mechanism used for initiation of translation of bRSV M2 ORF-2.

Increasing the size of the addition of the bRSV M2 ORF-2 (M2-2 ORF) sequence downstream of the ORF-1/ORF-2 overlap region to 120 nts in the plasmid pbRSVWC120, reduced levels of CAT protein expression to 0% of the control. This was also observed in the plasmid pbRSVStop120, which had ORF-1/ORF-2 overlap

region extended to 170 nts through the mutation of the ORF-2 terminating stop codon in addition to the addition of 120 nts of M2 ORF-2 sequence downstream of the ORF-1/ORF-2 overlap. This indicates that expression of the CAT protein in both plasmids was either being inhibited or the expressed reporter protein was failing to bind to the antibodies in the ELISA due to the addition of 40 amino acids to the CAT protein. This addition may have caused a conformational change that rendered it less efficiently detected. It is therefore unlikely that the inhibition of CAT protein expression in the plasmid pbRSVStop120 was due to the mechanism of coupled translation as expression of CAT protein was not recorded by the ELISA in both plasmids.

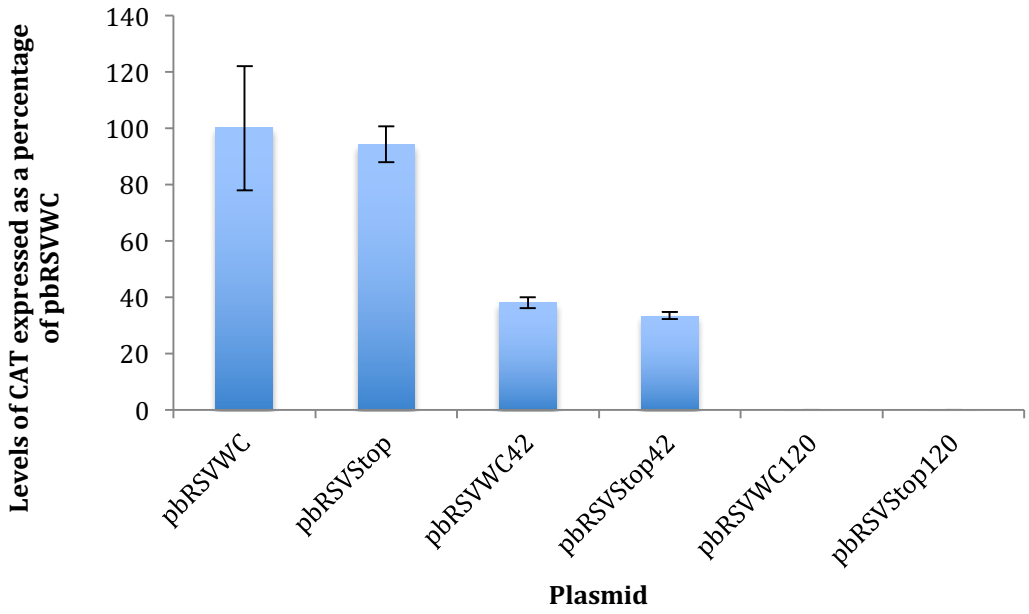


Fig. 4.8: CAT protein expression from plasmids investigating the effect of an increased M2 ORF-1/ORF-2 overlap region on the translation of ORF-2
Levels of expression of CAT protein expressed from plasmids pbRSVWC42, pbRSVStop42, pbRSVWC120 and pbRSVStop120 as described in Section 4.5.2.2 were individually transfected into HEp-2 cells. Results were normalised against the empty vector pBluescribe and are expressed as a percentage of the level achieved with positive control plasmid pbRSVWC. Error bars indicate standard deviation.

4.6.1. Investigation of the translational profile of the bRSV M2 mRNA using ribosomal profiling

Over recent years the technique of ribosomal profiling have enabled the precise definition of positions of ribosomes on mRNA (Ingolia, 2010; Ingolia et al., 2012; Ingolia et al., 2009). Ribosomal profiling works on the principle that all elongating ribosomes translating cellular mRNA cover a region of mRNA which is being translated. As well as covering the 9 nucleotides which are positioned in the A, P and E sites of the ribosome an elongating ribosome also extends over a further 21 nucleotides (Ingolia et al., 2009; Steitz, 1969). The region of mRNA covered by the ribosome (referred to as the footprint) is protected from degradation by RNases. The elongation phase of translation can be paused using the compound cycloheximide. Therefore by digesting all cellular RNA with RNases in lysed cells where translation is inhibited with cyclocheximide, only RNA footprints that are protected by paused elongating ribosomes are left. These footprints can be sequenced by next generation sequencing techniques. The sequence library generated can then be used to map the location of all elongating ribosomes in a cellular population at a specific time point. This allows the identification of specific regions of interest on a mRNA where high or low distributions of elongating ribosomes are present. The frame of each elongating ribosome can also be identified allowing the identification of ORFs in gene transcripts.

Ribosomal profiling analysis of the hRSV M2 mRNA demonstrated that a high frequency of elongating ribosomes were observed around the highly stable secondary structure located in the M2 ORF-1 that is essential for the mechanism of coupled translation (P. Gould and A. Easton, personal communication). Based on this observation it was suggested that the function of the structure is to stall elongating ribosomes, creating a ‘gap in ribosomal traffic’ at the 3′ end of ORF-2 and so allowing terminating ribosomes to move in a 5′ direction to reinitiate translation at one of the ORF-2 start codons. The high number of stalled ribosomes at a specific nucleotide position in the vicinity of the structure, a predominant peak was observed in the ribosome profile in this region. This peak was the greatest in the M2 mRNA and constituted 5.6% of all elongating ribosomes translating ORF-1 (P. Gould and A. Easton, personal communication). Taking the hRSV system as a model, ribosomal profiling was used to build a translational profile of the bRSV M2 ORF to identify

any regions of interest that could help in identifying if the mechanism of initiation of translation for the M2 ORF-2 was that of coupled translation.

Two ribosomal profile sequencing libraries were generated from MDBK cells that had been infected with bRSV Snook strain. One library was generated using RNA extracted at 4 hpi and the other at 8 hpi (sequencing libraries bRSV4P and bRSV8P, respectively). These sequence libraries contained data only originating from the protected footprints of elongating ribosomes. A detailed description of the methodology of ribosomal profiling and the sequencing of these libraries is given in Section 2.10. The processing pathway to analyse the raw sequencing data is in Section 3.2. All profiles are presented using the software IGV, and the spectra represent the distribution of elongating ribosomes on the M2 mRNA. Each point represents the nucleotide that each read aligned to the M2 gene transcript begins at.

For the RNA harvested at 4hpi (bRSV4P), a total of 13757 sequence reads out a total of 6774318 reads aligned to the bRSV genome. Of the 13757 reads 407 were obtained from the M2 mRNA. For the RNA harvested at 8 hpi (bRSV8P), a total of 67977 reads of a total of 6648581 aligned to the bRSV genome and of these, 1871 reads were from the M2 mRNA. The bRSV8P sequencing library was used to generate a translational profile of the M2 mRNA due to the higher level of reads aligned to the gene.

The data presented in Fig. 4.9 shows the distribution of ribosomes along the M2 mRNA in bRSV infected cells. From this profile, it is possible to identify specific regions in which there is a high or low distribution of elongating ribosomes. It is clear from this profile, that translation of M2 ORF-1 (M2-1 ORF) is significantly higher than that of M2 ORF-2 (M2-2 ORF). The large peaks located at the 5′ end of the M2 gene transcript represents the ribosomes surrounding the initiation codon of ORF-1. This is a frequent observation in such analyses as there is bias created during preparation of such sequencing libraries which favours the amplification of reads at the 5′ end of a mRNA. There were no significantly large peaks observed within the bRSV M2 mRNA. While there were peaks observed that indicated a higher number of reads aligned inside the region of bRSV M2 ORF-1 sequence that shared sequence homology to the hRSV critical structure these represented only a small proportion of

the total number of ribosomes translating ORF-1 and all were considerably less than the 5% observed in the hRSV M2 ribosomal profile data (P. Gould and A. Easton, personal communication). This further suggests that a coupled translation process analogous to that used by hRSV is not used by the ribosome for initiation of translation of the bRSV M2 ORF-2.

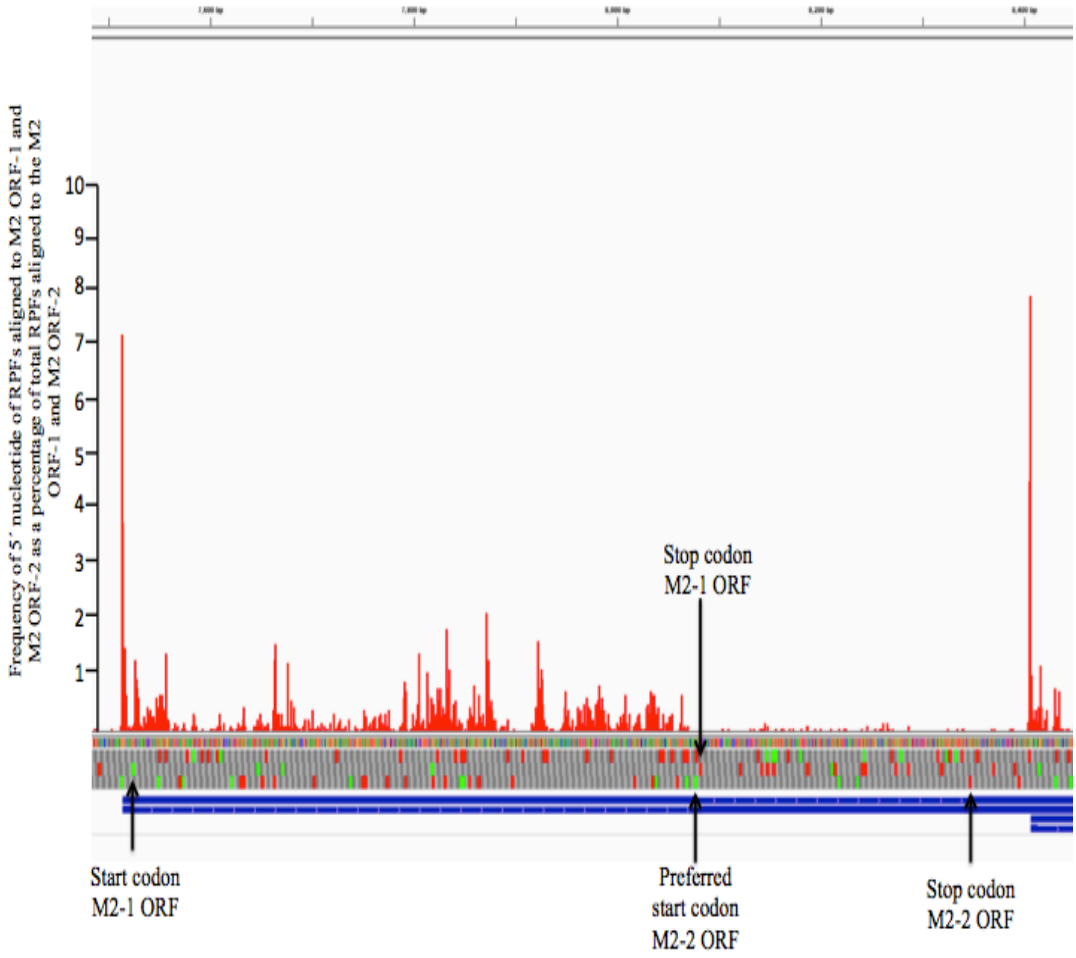


Fig. 4.9: Translational profile of the bRSV M2 gene transcript from MDBK cells infected with bRSV 8hpi.
Translational profile of the bRSV gene transcript from MDBK cells infected with bRSV 8hpi from the sequencing library bRSV8P. Location of all putative start (green) and stop (red) codons in all three frames in the M2 gene transcript are found below the spectra. Locations of the preferred start and stop codons for the M2 ORF-1 and M2 ORF-2 are identified.

4.6.2. Investigation of linked translation of bRSV M2 ORF-1 and ORF-2 using ribosomal profiling

As the initiation of translation of M2 ORF-2 is directly linked to the level of translation of M2 ORF-1 in viruses that use coupled translation, it is possible to use the method of ribosomal profiling to measure if translation of M2 ORF-2 is directly linked to the translation of M2 ORF-1 in bRSV. By measuring the percentage increase in translation between two time points for both M2 ORF-1 and M2 ORF-2, it is possible to compare the percentage increases from both ORFs in order to identify if the translation of the two ORFs are linked. A similar increase or decrease between two time points for both ORFs may suggest that translation of M2 ORF-2 is linked to translation of M2 ORF-1 therefore providing evidence that coupled translation may be used as the mechanism.

The two ribosomal profile sequence libraries bRSV4P and bRSV8P were used to measure the percentage increase in translation of both M2 ORFs at 4 and 8 hpi. The alignment software Bowtie was used to align the libraries to both ORFs in the bRSV M2 gene transcript. In order to calculate the percentage increase in translation for both ORFs, the number of reads aligned to each M2 ORF at 8 hpi (library bRSV8P) were taken and divided into the number of reads that were aligned to each M2 ORF at 4 hpi (library bRSV4P).

As described in Chapter 3, some of the sequences arising from the overlap region obtained will align to both ORFs. In the current literature there is a consensus that the bRSV ORF-1/ORF-2 overlap region overlaps each other by 17 nts due to the assumption that the ribosome initiates translation at the 5´ proximal AUG at position 554. However, evidence presented here in Section 4.4.2 shows that the AUG at position 563 is the preferred site of initiation, which reduces the overlap region to 8nts. Alignments of both sequencing libraries to this region identified a maximum of 1.1% (library bRSV8P) of reads aligned to the M2 mRNA sequence located in the M2 ORF-1/ORF-2 overlap region. As the percentage of sequences from the overlap region were sufficiently small to ignore, the analyses were performed using sequences derived from both ORFs that both included the ORF-1/ORF-2 overlap region.

Fig. 4.10 shows the percentage increase in the levels of translation between 4 and 8 hpi for the bRSV M2 ORF-1 and bRSV M2 ORF-2. For the M2 ORF-1 there was a 4.89-fold increase in the levels of translation between 4 hpi and 8 hpi. For the M2 ORF-2 there was a 1.35-fold increase in translation between 4hpi and 8 hpi. Although a statistical analysis was unable to be performed due to ribosomal profiling not being repeated for each time point, there was a significantly larger increase in translation in M2 ORF-1 compared to M2 ORF-2. This significant difference in the percentage increase suggests that the translation of M2 ORF-2 is not linked to the levels of translation of M2 ORF-1. This would suggest that as translation is not linked, the mechanism of coupled translation is unlikely to be used as the mechanism of initiation for translation of M2 ORF-2.

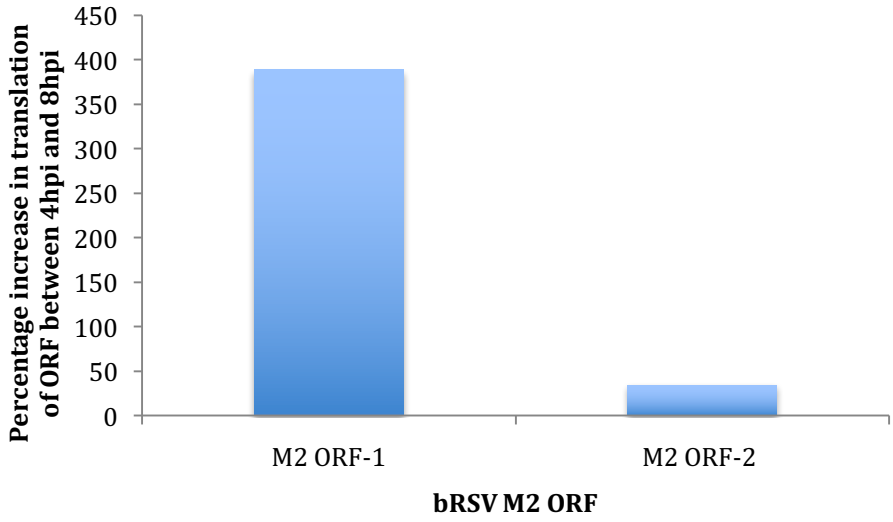


Fig. 4.10: **Efficiency of translation of the bRSV M2 ORF-1 and M2 ORF-2 at 4 hpi and 8 hpi**
Increase in translation displayed as a percentage increase between 4 hpi and 8hpi for M2 ORF-1 and M2 ORF-2. The bRSV M2 ORF-1 and M2 ORF-2 sequence were aligned to the bRSV4P library for the levels of translation at 4hpi. The M2 ORF-1 and M2 ORF-2 sequence were aligned to the bRSV8P library for the levels of translation at 8hpi.

4.7. Conclusions

Several conclusions could be drawn from the investigation into the initiation mechanism of bRSV M2 ORF-2. From these investigations it is clear initiation of the bRSV M2 ORF-2 is not reliant on sequence downstream of the M2 ORF-1/ORF-2 overlapping region. Replacing sequence downstream of the M2 ORF-1/ORF-2 overlap region with the CAT reporter gene, still lead to successful translation of the second ORF. This allowed the generation of a reporter system that was capable of measuring ORF-2 expression and thereby allowing the investigation of the mechanism used for initiation of translation M2 ORF-2. Furthermore sequence downstream of the M2 ORF-1/ORF-2 overlap region had no bearing on the efficiency of the mechanism. These observations were all similar characteristics observed for the translation of M2 ORF-2 in other members of the *Pneumoviridae* family.

It can also be concluded that initiation of bRSV M2 ORF-2 is reliant on the two cognate start codons in the ORF-1/ORF-2 overlap. There was no indication that translation of ORF-2 could be initiated from cognate or non-cognate start codons located outside of the M2 ORF-1/ORF-2 overlap region as translation was inhibited when both cognate start codons in the overlap region were mutated. It was also clear that the primary initiation site for M2 ORF-2 was located at position 563. Initiation of translation could also occur at the start codon at position 554 but only at a 20-fold lower level if the start codon at position 563 was not present. Mutation of the start codon at 554 also improved efficiency for initiation for the start codon at 563.

From this chapter it is also apparent that unlike other members in the *Pneumoviridae* family, bRSV does not use the mechanism of coupled translation termination/initiation for initiation of translation of M2 ORF-2. Data presented in this chapter revealed that translation of M2 ORF-2 was not reliant on termination of translation of ORF-1 in the vicinity of the initiation site. Mutating the terminating stop codon of M2 ORF-1 so that the M2 ORF-1/ORF-2 overlap was increased in size to 92 nts, had no effect on the efficiency of translation of M2 ORF-2. The size that the bRSV M2 ORF-1/ORF-2 overlap was increased was significantly larger than the size of the increase in the M2 ORF-1/ORF-2 overlap region required to inhibit translation of M2 ORF-2 in other members of the *Pneumoviridae* family (hRSV

(Ahmadian et al., 2000) and APV (Gould and Easton, 2007)). This would suggest that the mechanism in bRSV is not distance dependent on the M2 ORF-1 terminating stop codon, which is a characteristic universal to all examples of coupled translation.

Although there was no detectable expression of the CAT protein when the ORF-1/ORF-2 was increased to 170 nts with the mutation of the ORF-1 stop codon and addition of 120 nts of M2 ORF-2 sequence (pbRSVStop120), the control with the ORF-1 terminating stop codon present also had no detectable expression of CAT protein. This suggests that the inhibition of ORF-2 translation in the plasmid pbRSVStop120 was not due to the mutation of the ORF-1 stop codon and was likely due to other unknown circumstances.

Further evidence of an alternate mechanism used by bRSV for initiation of translation of M2 ORF-2 came from the use of ribosomal profiling. By building a translational profile of the M2 gene transcript, it was clear that there were no significant examples of ribosomes stalling in a similar manner observed in hRSV (P. Gould and A. Easton, personal communication). As this was essential for translation of M2 ORF-2 in hRSV, the lack of a significant peak would suggest a different mechanism was used. In bRSV it was also observed that there was no relationship between translation of M2 ORF-1 and translation of M2 ORF-2. No proportional increase was observed in the translation of M2 ORF-2 when an increase in M2 ORF-1 translation was observed between 4 hpi and 8 hpi. This indicated that translation of M2 ORF-2 was not linked to translation of M2 ORF-1. In hRSV, translation of M2 ORF-2 is linked to translation M2 ORF-1, with translation of M2 ORF-2 always remaining at ~3% of translation of M2 ORF-1. As this was not observed in bRSV, it is clear that translation of M2 ORF-2 is not linked to translation of M2 ORF-1 and does not use the mechanism of coupled translation. However, these data from this chapter does not point to another mechanism.

Chapter 5
Identification of Non-Canonical Mechanism used for Initiation of Translation of M2 ORF-2

5.1. Introduction

Data presented in Chapter 4 demonstrates that initiation of translation of M2 ORF-2 (M2-2 ORF) in bRSV is not affected by termination of translation of M2 ORF-1 (M2-1 ORF). This is a defining character of the mechanisms of coupled translation termination/initiation that use a TURBS sequence or an upstream structured region, and both are reliant on termination of translation of ORF-1. It is therefore unclear what mechanism is used for initiation of translation of M2 ORF-2. As the M2 ORF-1 separates the m^7G cap from the M2 ORF-2 initiation site by a distance of 562 nts, it is impossible for standard canonical initiation to be used by the ribosome for the initiation of M2 ORF-2. Therefore a non-canonical initiation mechanism must be used for the initiation of translation of the bRSV M2 ORF-2.

It is possible that a proportion of scanning ribosomes may bypass a number of initiation sites to reach the initiation site of M2 ORF-2 using leaky scanning. As discussed in Section 1.6.2, leaky scanning is used by ribosomes to initiate translation at alternate internal initiation sites that have stronger Kozak sequences than the AUG codon with a weak Kozak sequence located closest to the m^7G cap. In addition, the distance the ribosome can scan through leaky scanning can range from a few nucleotides to over 800 nts (Fütterer et al., 1997). However, in most examples, ribosomes tend to only bypass two or three AUG codons over a short distance, as the efficiency of scanning decreases with distance from the m^7G cap. Between the first start codon on the M2 mRNA and the first start codon located in the ORF-1/ORF-2 overlap region, there are 14 alternate start codons that the scanning ribosome could initiate at upstream of the ORF-1/ORF-2 overlap region. As seen in table 5.1, these 14 start codons vary in degrees of strength of their Kozak sequence. In comparison to the other AUG codons in M2 ORF-1 (M2-1 ORF), the preferred AUG codon at position 564 in the ORF-1/ORF-2 overlap region has a relatively strong Kozak sequence. However there are three start codons upstream with similar strength Kozak sequences (beginning at position 35, 332 and 335). It would be unusual for the ribosome to bypass all 14 codons along a length of 562 nts to initiate translation of ORF-2. In viruses where there is a significant distance bypassed by the scanning ribosome, such as in the rice tungro bacilliform virus (895 nts), there are only three codons between the m^7G cap and the initiation site. One of these codons is also used for initiation of another ORF through leaky scanning (Fütterer et al., 1997). On the

balance of these factors it is therefore unlikely that bRSV uses leaky scanning for initiation of ORF-2

Position of A in Start Codon	Nucleotide at position -3 of start codon	Nucleotide at position +4 of start codon	Notes on Start codon
10	A	U	Start codon for ORF-1 (M2-1 ORF)
35	A	A	
59	U	G	
69	A	U	
96	U	G	
107	C	C	
133	U	C	
157	U	G	
224	A	C	
304	G	A	
332	A	A	
335	A	A	
524	U	G	
554	U	A	First start codon found in ORF-1/ORF-2 overlap region
563	A	A	Preferred start codon for M2 ORF-2 (M2-2 ORF)

Table 5.1: Sites of AUG codons in the bRSV M2 ORF-1
Sites of AUG codons in the bRSV M2 ORF-1 (M2-1 ORF) including key Kozak nucleotides surrounding each AUG.

Three additional potential non-canonical translation initiation mechanisms are options that bRSV may use for the initiation of translation of the M2 ORF-2. Programmed ribosomal frameshifting (PRF) involves the use of a slippery sequence and a structured element to allow the elongating ribosome to shift frame to translate another ORF. The result of PRF is the generation of a fusion protein, although this has never been reported in any *Pneumoviridae* M2 mRNA products making this mechanism an unlikely candidate. Ribosomal shunting involves a ribosome bypassing large sections of sequence, usually in the 5′ UTR after translating a short ORF. Although this mechanism is unlikely to be used in the classical sense as the bRSV M2 mRNA 5′ UTR is only 9 nts long, there are examples of shunting that use complementary sequences to the 18S rRNA to bypass sections of sequence (Yueh and Schneider, 2000). Internal ribosome entry sites (IRES) use structured elements to initiate translation in a cap-independent manner. These structured regions are capable of directly loading a ribosome onto the mRNA without involvement of the 5′cap. This mechanism may be used by bRSV for initiation of translation of M2 ORF-2, as there is no data on whether initiation of translation is cap-dependent. A more detailed review of these mechanisms can be found in Section 1.6.

The site of initiation of M2 ORF-2 is also of interest. While there are two start codons in the ORF-1/ORF-2 overlap region, only the AUG codon closest to the stop codon of M2 ORF-1 is the preferred initiation site. It is possible that the Kozak sequence is a factor in initiation or the specific position of the AUG in relation to some other critical sequence such as a slippery sequence or an IRES. As discussed in Section 4.1, the context of the Kozak sequence surrounding the preferred AUG codon used for initiation of translation of M2 ORF-2 is not perfect, with an optimal A nucleotide located in position -3 in the Kozak sequence but a non-optimal A nucleotide in the +4 position. The context of the remaining Kozak sequence is also far from perfect (ACAAAAUGA). In comparison to the upstream AUG codon (position 554) Kozak sequence (UGUUAAUGA), the preferred AUG codon has stronger consensus sequence with a stronger nucleotide located at position -3 (Table 5.1). Further work identifying the importance of the Kozak sequence and the specific position is needed before a conclusion can be drawn.

The aim of this chapter is to explore the mechanism used by bRSV to initiate translation of the M2 ORF-2 (M2-2 ORF) and to investigate why the specific site of initiation is so important for successful initiation of translation of ORF-2

5.2.1. Investigation into whether a programmed ribosomal frameshifting mechanism is used by bRSV for the initiation of translation of the M2 ORF-2

Programmed ribosomal frameshifting (PRF) is a non-canonical mechanism used to translate multiple proteins from a polycistronic mRNA strand. As reviewed in Section 1.63, PRF allows an elongating ribosome to alter its translating reading frame by either slipping backwards or forwards by 1 or 2 nucleotides. Examples of PRF are common in many different virus families and many tend to shift the elongating ribosome to the -1 reading frame. There are two requirements for PRF; a slippery sequence with the consensus sequence of X_XXY_YYZ (where _ separates in-frame triplets, X can be any three identical nucleotides, Y can represent a triplicate of A or U and Z represents A, C or U) and a hairpin loop which is thought to act to slow the elongating ribosome down allowing PRF to occur. As a result of this shift, a fusion protein is formed from the combined translated ORFs. A method to test for PRF is to use western blot analysis to investigate the presence of fusion proteins that are created as result of the frame shift by the elongating ribosome.

Fig. 5.1 shows the sequence surrounding the bRSV M2 ORF-2 initiation codon. As shown in Section 4.4.2, mutation of the AUG codon at position 563 to CUG reduced ORF-2 expression to negligible levels. Thus, if PRF is used in bRSV M2 ORF-2 initiation, the site of the slippery sequence is likely to be in close vicinity upstream of this codon. As seen in Fig. 5.1, there is no sequence that resembles the consensus slippery sequence. However, it is possible for frame shifting to occur even without the full slippery sequence, meaning that is possible that this mechanism could still be used.

545-UAAUUGUUAAUGAACAAAAUGAA**UAA**-570

Fig. 5.1: Sequence surrounding the bRSV M2 ORF-2 initiation site
Sequence surrounding the ORF-2 (M2 ORF-2) preferred initiation site from position 545-570. Preferred AUG codon is underlined and the termination codon for M2 ORF-1 (M2-1 ORF) is highlighted in bold.

While western blot analysis of the M2 gene products of hRSV found no evidence for a fusion product that would be a natural consequence of PRF (Ahmadian et al., 2000) this has never been investigated for bRSV. Currently there is no available antibody specific for the bRSV M2-1 or M2-2 proteins. To identify whether a fusion protein

was produced western blot analysis was performed using the protein derived from cells transfected with the expression plasmids that express CAT protein from ORF-2 used to investigate bRSV M2 ORF-2 expression as described in Section 4.5.2.3. Western blot analysis was performed as described in Section 2.2.1 and Section 2.2.2 with the antibody Anti-CAT (Goat Anti-CAT IgG Fc (HRP) Abcam) as the primary antibody. For each sample a standard amount of protein (100μg) was used, which was established by using a bicinchoninic acid assay (BCA) assay to quantify protein concentrations in lysed samples as described in Section 2.2.3.

The protein from cells transfected with the plasmid pbRSVWC42 was used to investigate if any fusion proteins were present. Protein from cells transfected with the plasmids pbRSVWC120, pbRSVStop120 were used as negative controls as these plasmids did not express CAT protein (Section 4.5.2). Material from cells transfected with plasmid pBluescribe was also used as an additional negative control. Finally, the material from cells transfected with the hRSV plasmid pWildCAT was used as a positive control for CAT. For all samples except those from cells transfected with plasmid pWildCAT, the same quantity of protein was analysed. For the material from cells transfected with pWildCAT plasmid, a 1:10 dilution (10 μg) was used in view of the significantly higher levels of CAT protein expressed by this plasmid when compared to bRSV reporter plasmids.

A theoretical fusion protein of CAT and M2-1 proteins would have a mass of approximately 46 kDa. As shown in Fig. 5.2, intense bands with a mass of approximately 25 kDa were observed in lanes 1 and 5 which were from cells transfected with plasmids pbRSVWC42 and pWildCAT, respectively. These products are the expected size for CAT protein that was not fused to the M2-1 protein. No protein of the size predicted for a fusion protein was seen in any track. As expected, no CAT protein was observed in the negative controls from cells transfected with plasmids pbRSVWC120 (lane 2), pbRSVStop120 (lane 3) or pBluescribe (lane 4). The absence of expressed CAT protein in lanes 2 and 3 for pbRSVWC120 and pbRSVStop120, respectively, indicates that there was no expression of the CAT protein from the ORF-2 with these constructs as expected.

With no reason to assume that a fusion protein would be efficiently cleaved to form an intact CAT protein and the absence of a fusion protein product, the data indicates that ribosomal frameshifting is not used to initiate translation of M2 ORF-2 in bRSV. This is further supported by the evidence in Section 4.4.3, where the AUG located at position 563 is essential for efficient initiation of translation for M2 ORF-2. PRF also requires a hairpin loop directly downstream of the initiation site, and, for all plasmids used in previous chapters, expression from ORF-2 was seen even when the bRSV M2 ORF-2 (M2-2 ORF) directly downstream of the ORF-1/ORF-2 overlap region was replaced with the CAT gene.

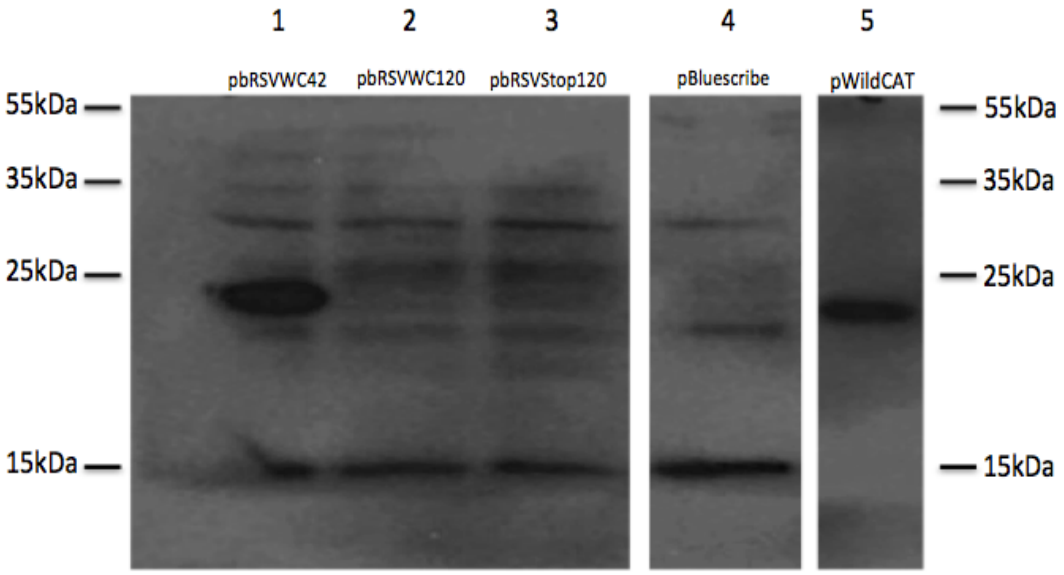


Fig. 5.2: Western blot analysis of CAT protein from cellular lysates of transfected plasmids
Western blot analysis of CAT protein from HEp-2 cellular lysates that were individually transfected with reporter plasmids. Lane 1: pbRSVWC42, lane 2: pbRSVWC120, lane 3: pbRSVStop120, lane 4: pBluescribe and lane 5: pWildCAT. All lanes orginated from the same gel, white gaps indicate lanes that were not adjacent to each other.

5.3.1.1. Investigation of initiation of the M2 ORF-2 in bRSV at the AUG at position 563

It is unclear why ribosomes have a strong preference to initiate translation at the AUG located at position 563 rather than the AUG located at position 554 in the ORF-1/ORF-2 overlap region. As indicated above, a possible reason for the preference is that the Kozak sequence surrounding the AUG codon plays a role in the initiation process with the preferred AUG site at position 563 having a stronger upstream consensus sequence as seen in Fig. 5.3 (+4 position for both AUG codons is the same). To investigate whether the sequence upstream of the preferred initiation codon plays a role in initiation of ORF-2, a plasmid named pbRSVAUGSeqS was created. This plasmid was based on the standard reporter system described in Chapter 4 to measure ORF-2 expression. The plasmid was designed to direct synthesis of a bicistronic mRNA, where the two ORFs overlapped. The bicistronic mRNA contained the entire bRSV M2 ORF-1 sequence in ORF-1 and the bRSV M2 ORF-1/ORF-2 overlap region and the CAT reporter coding sequence in ORF-2. All mutations made to create the plasmid pbRSVAUGSeqS were introduced into the native bRSV M2 ORF-1 sequence. Mutations in the native bRSV M2 ORF-1 sequence in ORF-1 for the plasmid pbRSVAUGSeqS are described below. For all other plasmids used in this chapter to investigate ORF-2 expression, the same reporter system was used. Unless otherwise stated these plasmids were in the same format as pbRSVWC and mutations made to investigate ORF-2 expression were made to the native bRSV M2 ORF-1 sequence located in ORF-1 of the plasmid's gene.

For the plasmid pbRSVAUGSeqS, the upstream Kozak sequence of the preferred AUG at position 563 (residues 557-562) was mutated to the same sequence used for the Kozak sequence for the AUG codon at position 554 (residues 548-553) (Fig. 5.3). The AUG codon at position 554 in the ORF-1/ORF-2 overlap region that is not used for expression of ORF-2 was also mutated. The aim was to determine whether the introduction of a weaker Kozak sequence (as used by the AUG codon at position 554) for the preferred AUG codon at position 563, would alter CAT protein expression. A decrease in expression would indicate that the sequence upstream of the AUG at position 563 was critical for its use as an initiation site.

bRSV M2 Sequence UAAUUGUUAAUGAACAAAAUGAA**UAA**

pbRSVAUGSeqS UAAUUGUUACUGUUGUUAAUGAA**UAA**

Fig. 5.3: Mutations introduced into the sequence upstream of AUG (563) for expression of ORF-2

The mutations in the M2 ORF-1/ORF-2 overlap region for the plasmids pbRSVAUGSeqS are highlighted in red. AUG initiation sites for M2 ORF-2 (M2-2 ORF) inside the ORF-1/ORF-2 overlap region are underlined. The stop codon for the ORF-1 (M2-1 ORF) is highlighted in bold. The sequence displayed is from position 545 to position 570

Construction of the plasmid pbRSVAUGSeqS

Two-Step PCR

For the construction of the plasmid pbRSVAUGSeqS, a two-step PCR method was used to create the insert with the mutations that were to be ligated into the vector plasmid. Two amplicons were created for the first PCR step. Each amplicon was individually amplified by PCR using specific primer pairs. Both amplicons contained a region of sequence that was homologous to the other amplicon and this was incorporated by either the forward or the reverse primer at the 5′ or 3′ ends respectively depending on the primer pair used. This region of overlapping sequence contained the mutations that were introduced into the sequence. Both amplicons were purified and placed into a PCR reaction. As the two amplicons contained a region of sequence homology, the two fragments annealed during the second PCR step. The joined amplicons were then amplified with external primers. The resulting joined fragment was then ligated into the digested vector plasmid. A detailed methodology for each technique used in this chapter can be found in Section 2. Sequences for all primers used in all PCR reactions in all sections can be found in the Appendix C.

For the first PCR step, one amplicon (named bRSV) contained the bRSV 5′ UTR and the bRSV M2 ORF-1 and the second (named CAT) contained the bRSV M2 ORF-1/ORF-2 overlap and the CAT reporter coding sequence. The two primer pairs used to generate the first two amplicons in the first PCR stage were bRSVWTF/AUGSeqSR (bRSV amplicon) and AUGSeqSF/CATR (CAT amplicon). Each amplicon was generated by PCR from the plasmid pbRSV5′AUG (Section 2.3.1), which already had the AUG codon at position 554 mutated to CUG. Before the second PCR step, both products were cleaned using a QIAGEN PCR Purification Kit as described in Section 2.3.15. The two products were added in equal molar

quantities and were 'fused' via the second PCR step using the primers bRSVWTF/CATR (Section 2.3.1). The final product contained a Kpn I and Sph I site at the 5′ and 3′ ends, respectively. The fragment was digested with the endonuclease pair Kpn I/Sph I (Section 2.3.4 and Section 2.3.5) and ligated into the plasmid pBluescribe (Section 2.3.7) that was previously digested with endonuclease pair Kpn I/Sph I (Section 2.3.4 to Section 2.3.6). The ligated plasmid was transformed into *E.coli* DH5α bacteria and plated onto AMP/IPTG/X-Gal LB agar plates (Section 2.3.8). The bacteria were grown overnight at 37^{o}C and colonies were selected through blue white selection (Section 2.3.9). Each picked colony was grown overnight at 37^{o}C and small scale plasmid purification was performed as described in Section 2.3.13. The purified plasmids were digested with endonucleases Kpn I and Sph I to confirm the presence of the correct insert (1237 nts). After confirmation the sequence was confirmed using primers T7-91 and T3 down (Section 2.3.10).

5.3.1.2. The role of upstream sequences in expression of the bRSV M2 ORF-2 from the AUG at position 563

The plasmid pbRSVAUGSeqS and control plasmids pbRSVWC, pbRSV3′AUG, pbRSV5′AUG and pBluescribe were individually transfected into HEp-2 cells as described in Section 2.4.1. Plasmids pbRSVWC and pbRSV5′AUG were positive controls and pbRSV3′AUG was a negative control for CAT protein expression (sequences for plasmids pbRSVWC, pbRSV5′AUG and pbRSV3′AUG are found in Section 4.4). Levels of CAT protein expressed were quantified using a CAT ELISA assay as described in Section 2.4.2. All samples were normalised against levels from transfected empty vector pBluescribe plasmid to account for any background signal generated in the ELISA. For each sample levels of CAT protein were expressed were as a percentage of CAT protein expression from plasmid pbRSVWC.

As seen Fig. 5.4, introducing a weaker Kozak sequence (as used by the AUG codon at position 553) for the AUG codon at position 563 and mutating the upstream sequence of the AUG codon located at position 554 in the plasmid pbRSVAUGSeqS reduced levels of CAT protein expression to 13.1% of that of the control. There was a significant reduction in CAT protein expression from the plasmid pbRSVAUGSeqS in relationship to both the positive controls pbRSVWC

(p=0.000120) and pbRSV5′AUG (p=0.0047). CAT protein expression from the plasmid pbRSVAUGSeqS was statistically significantly higher by 9.1% in comparison to the negative control pbRVS3′AUG (p=0.0015). These data indicate that the upstream sequence of the preferred AUG site beginning at position 563 is important for efficient initiation of ORF-2 as mutating this upstream sequence to a Kozak sequence of poorer context, significantly reduces levels of expression.

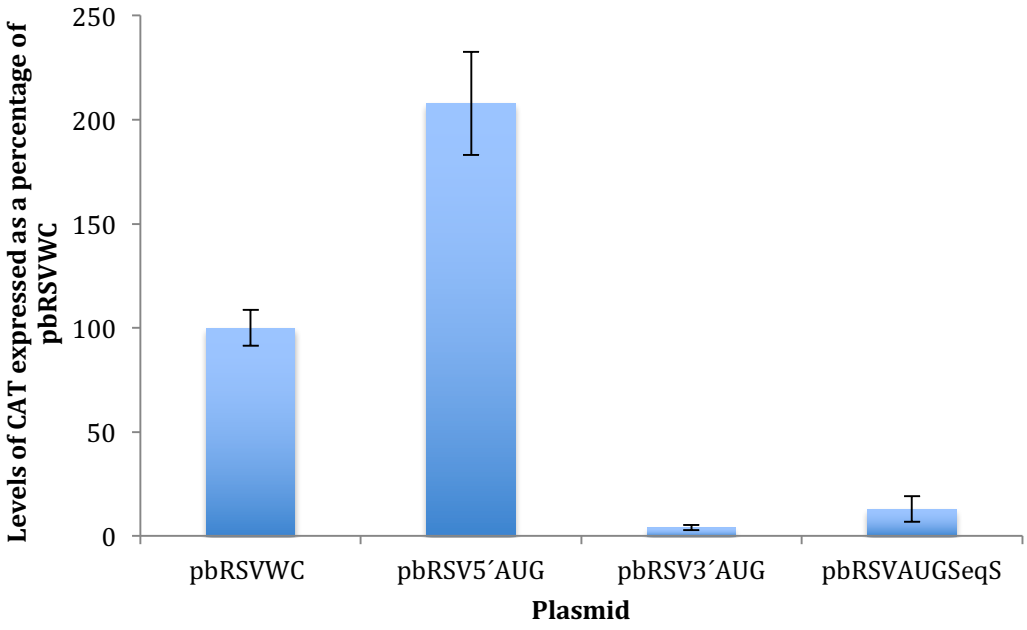


Fig. 5.4: CAT protein expression of plasmids used to investigate factors leading to initiation of translation at the AUG codon at position 563
Levels of expression of CAT protein expressed from plasmids individually transfected into HEp-2 cells. Mutations made to these plasmids are described in Section 5.3.1.1. Results were normalised against the empty vector pBluescribe and are expressed as a percentage of the positive control plasmid pbRSVWC. Error bars indicate standard deviation.

5.3.2.1. The role of the Kozak sequence for translation initiation from the AUG codon located at position 563

The data from Section 5.3.1.2 demonstrated that the sequence within residues 548-562 upstream of the preferred AUG codon at position 563 plays a role in the initiating ribosome choosing this site for initiation of translation of the bRSV M2 ORF-2. However, the data did not establish whether this is due to the key Kozak nucleotides surrounding the AUG codon. In this region the strongest nucleotide of the Kozak sequence is an A residue located at position 560 (at position -3 relative to

the start codon). The -3 position in the Kozak sequence is one of the key residues and the optimal nucleotide for this position is an A nucleotide. Therefore to further investigate whether the Kozak sequence is relevant for the initiation of the bRSV M2 ORF-2, three plasmids (pbRSVΔ560G, pbRSVΔ560C and pbRSVΔ560U) were created. These plasmids used the CAT reporter system to measure ORF-2 expression as used previously. For the three plasmids the ORF-1 sequence was unaltered while the A residue at position 560 was mutated to one of the three other nucleotides (Fig. 5.5) to determine if translation of CAT protein was reduced with the reduction in Kozak sequence strength. For plasmid nomenclature, G, C and U at the end of the plasmid name denotes the nucleotide that the A nucleotide at position 560 was mutated to.

bRSV M2 Sequence	UAAUUGUUAAUGAACAAAAUGAA**UAA**
pbRSVΔ560G	UAAUUGUUAAUGAACGAAAUGAA**UAA**
pbRSVΔ560C	UAAUUGUUAAUGAACCAAAUGAA**UAA**
pbRSVΔ560U	UAAUUGUUAAUGAACUAAAUGAA**UAA**

Fig. 5.5: Mutations made to the nucleotide at position 560 in the ORF-1/ORF-2 overlap region
Mutations in the ORF-1/ORF-2 overlap region for the plasmid pbRSVΔ560G, pbRSVΔ560C and pbRSVΔ560U. Sites of mutations are highlighted in red. AUG sites inside the ORF-1/ORF-2 overlap region are underlined. The stop codon for ORF-1 is highlighted in bold. The sequences displayed, begin at position 545 and end at position 570 from the start of the bRSV mRNA sequence or the beginning of bRSV sequence in reporter plasmids.

Construction of the plasmids pbRSVΔ560G, pbRSVΔ560C and pbRSVΔ560U

The mutation for each plasmid was created using two-step PCR. A detailed description of the methodology of the two-step PCR can be seen in Section 5.3.1.1. The method for cloning and sequencing of the plasmids was the same as used in Section 5.3.1.1. For the first PCR step for each plasmid, the primer pairs used for the first PCR step for each plasmid are shown in Table 5.2 and were amplified from the plasmid pbRSVWC. For each plasmid, the two amplicons generated in the first PCR step were joined in the second PCR step using the primer pair bRSVWTF/CATR.

Plasmid	Primer pair used for bRSV amplicon	Primer pair used for CAT amplicon
pbRSVΔ560G	bRSVWTF/Δ560GR	Δ560GF/CATR
pbRSVΔ560C	bRSVWTF/Δ560CR	Δ560CF/CATR
pbRSVΔ560U	bRSVWTF/Δ560UR	Δ560UF/CATR

Table 5.2: Primer pairs for each amplicon used for the first step in two-step PCR for the plasmids pbRSVΔ560G, pbRSVΔ560C and pbRSVΔ560U
Primer pairs for the bRSV amplicon and CAT amplicon in the first step of two-step PCR for plasmids pbRSVΔ560G, pbRSVΔ560C and pbRSVΔ560U.

5.3.2.2. Investigation of the role of Kozak sequence 'strength' for translation initiation at the AUG codon located at position 563

The three mutant plasmids pbRSVΔ560G, pbRSVΔ560C and pbRSVΔ560U and two control plasmids, pbRSVWC and pBluescribe, were individually transfected into HEp-2 cells as described in Section 2.4.1. Levels of CAT protein expressed were quantified using a CAT ELISA assay as described in Section 2.4.3 and all data were normalised against the assay results from the transfected empty vector pBluescribe plasmid as described previously. Levels of CAT protein were expressed as a percentage of CAT protein levels obtained from the plasmid pbRSVWC.

As seen in Fig. 5.6 with plasmid pbRSVΔ560G, expression of the CAT protein from the second ORF was only slightly reduced to 89.9% of the control which was not statistically significantly different to the control (p=0.39). With plasmid pbRSVΔ560C expression of the CAT protein was reduced by a significant level to 75% of the control (p=0.02) and with plasmid pbRSVΔ560U, expression of the CAT protein was further reduced to 52% of the control (p=0.0028). These data are entirely consistent with the order of nucleotide preference at position -3 in the Kozak sequence of A>G>C>U. This strongly suggests that for the AUG codon at position 563 the Kozak sequence is a significant factor in determining the efficiency of translation initiation as is seen in the canonical translation initiation mechanism.

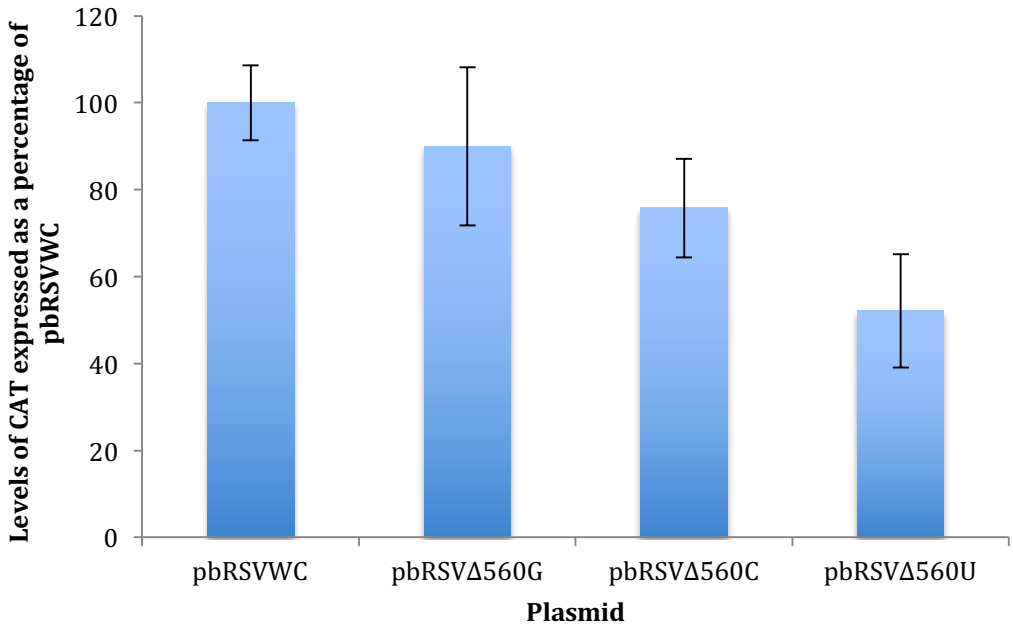


Fig. 5.6: CAT protein expression for the plasmids used to investigate the importance of the Koazk -3 residue at the AUG codon at position 563 for initiation of translation M2 ORF-2
Levels of expression of CAT protein expressed from plasmids individually transfected into HEp-2 cells. Mutations made to these plasmids are described in Section 5.3.2.1. Results were normalised against the empty vector pBluescribe and are expressed as a percentage of the positive control plasmid pbRSVWC. Error bars indicate standard deviation.

5.3.3.1. Investigation of optimal efficiency for translation initiation using the AUG codon at position 563

While the data described above demonstrates that the nucleotide preference at position -3 upstream of the AUG codon at position 563 is consistent with the Kozak sequence preference, the remainder of the sequence in this region is suboptimal (Fig. 4.15). It therefore remains a possibility that the remainder of this sequence plays an additional role in the initiation of translation of M2 ORF-2. To further investigate whether the Kozak sequence-like function is the primary or sole function for the upstream sequence of the AUG codon at position 563, further mutations were made to introduce an optimal Kozak sequence surrounding the AUG codon at position 563. This was then studied to see if the optimisation led to an increase in ORF-2 expression which would strengthen the likelihood that the Kozak sequence is a primary factor in ORF-2 translation. As shown in Fig. 5.7, the optimal Kozak sequence (Kozak, 1987) (GCCGCCACCAUGG) was introduced by mutation into both reporter plasmids pbRSVWC and pbRSVStop to create the plasmids

pbRSVWCKozak and pbRSVStopKozak, respectively. Plasmid pbRSVStopKozak contains an ORF-1/ORF-2 overlap region extended by 42nts compared to the wildtype.

Optimal Kozak Sequence	GCCGCCACCAUGG
bRSV M2	UAAUUGUUAAUGAACAAAAUGAA**UAA**
pbSVWCKozak	UAAUUGUUAGCCGCCACCAUGGA**UAA**
pbRSVStopKozak	UAAUUGUUAGCCGCCACCAUGGACAA

Fig. 5.7: Mutations made to introduce optimal Kozak sequence for the AUG codon at position 563
Mutations in the ORF-1/ORF-2 overlap region for the plasmid pbRSVWCKozak and pbRSVStopKozak. Sites of mutations are highlighted in red. AUG sites in frame for ORF-2 initiation inside the ORF-1/ORF-2 overlap region are underlined. The stop codon for ORF-1 (M2-1 ORF) is highlighted in bold. The sequences displayed, begin at position 545 and end at position 570 from the start of the bRSV mRNA sequence or the beginning of bRSV sequence in reporter plasmids.

Construction of the plasmid pbRSVWCKozak and pbRSVStopKozak

The mutations were created for the plasmids pbRSVWCKozak and pbRSVStopKozak using two-step PCR. A detailed description of the methodology of two-step PCR can be seen in Section 5.3.1.1. The method for cloning and sequencing of the plasmids was the same as used in Section 5.3.1.1. For the first PCR step for each plasmid, the primer pairs used to generate the two products for the first PCR step for the plasmid pbRSVWCKozak were bRSVWTF/KozakR (bRSV amplicon) and KozakF/CATR (CAT amplicon) using plasmid pbRSVWC as a template. The primer pairs used for the first PCR step for the plasmid pbRSVStopKozak were bRSVWTF/KozakStopR (bRSV amplicon) and KozakStopF/CATR (CAT amplicon) using the plasmid pbRSVStop as a template. The amplicon pairs generated in the first PCR step were joined in the second PCR step using the primer pair bRSVWTF/CATR.

5.3.3.2. Expression levels directed by the optimised Kozak sequence adjacent to the AUG codon at position 563

The two plasmids, pbRSVWCKozak and pbRSVStopKozak and control plasmids pbRSVWC, pbRSVStop and pBluescribe were individually transfected into HEp-2 cells as described previously (Section 2.4.1). Levels of CAT protein were determined, normalized and expressed as a percentage of CAT protein expression from the plasmid pbRSVWC, as before.

As seen in Fig. 5.8, introducing an optimal Kozak sequence for the start codon located at 563 in the plasmid pbRSVWCKozak significantly reduced levels of CAT protein expression to 66% of the control pbRSVWC ($p=0.005$). Introducing an optimal Kozak sequence upstream of position 563 and extending the ORF-1 by 42nts, by mutating the wild type ORF-1 termination codon in the plasmid pbRSVStopKozak had little effect on CAT protein expression; giving 94.5% expression of the control pbRSVWC ($p=0.55$) and only a 0.2% increase compared to the control pbRSVStop ($p=0.97$), which also had the ORF-1 stop codon removed but did not have an optimal Kozak sequence for the AUG codon at position 563.

These data strongly suggest that the sequence surrounding the AUG at position 563, is already optimal for the ribosome to initiate translation despite differences with the optimal Kozak sequence. Changing the sequence to resemble the optimum Kozak sequence made no significant difference or reduced translation efficiency. As before, removing the ORF-1 termination codon had no effect on translation of ORF-2 even when an optimal Kozak sequence was introduced. The data therefore indicates that the mechanism used for initiation relies on the surrounding sequence and although the nucleotide at position -3 appears to follow the Kozak 'rules', does not conform completely to the pattern seen with a canonical Kozak sequence.

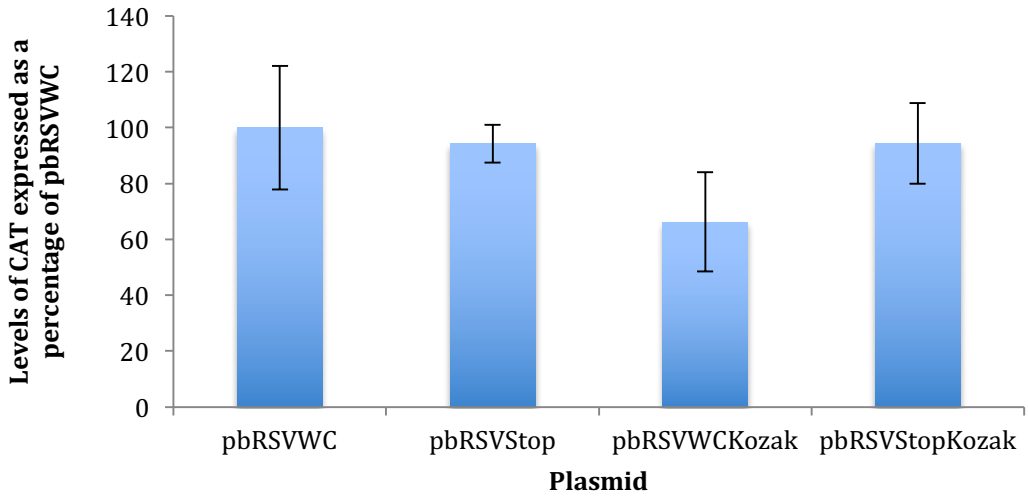


Fig. 5.8: CAT protein expression for plasmids that are used to investigate whether the translation of M2 ORF-2 is at optimal efficiency for initiation at the AUG codon at position 563
Levels of expression of CAT protein expressed from plasmids individually transfected into HEp-2 cells. Mutations made to these plasmids are described in Section 5.3.3.1. Results were normalised against the empty vector pBluescribe and are expressed as a percentage of the positive control plasmid pbRSVWC. Error bars indicate standard deviation.

5.4.1. Potential alternate start sites for initiation of translation of the bRSV M2 ORF-2

It is apparent that the preferred initiation site for M2 ORF-2 (M2-2 ORF) is located at the AUG codon beginning at position 563, with the alternate AUG in the ORF-1/ORF-2 overlap region at position 554 only allowing initiation of translation of ORF-2 to occur at significantly lower levels (5.4% of normal ORF-2 translation). The sequence surrounding the preferred AUG codon is of importance for initiation in regards to the Kozak sequence but also appears to play another role as well as acting as the Kozak sequence. It is also apparent that initiation is not reliant on the translating ribosome of M2 ORF-1 (M2-1 ORF) terminating in the vicinity of the start codon for ORF-2. However it is unclear if the position of the AUG at position 563 is vital for successful initiation of translation of ORF-2.

As only one AUG codon has been tested as an alternative initiation site for translation of ORF-2, it may be possible that start codons at alternate sites may be able to act as a suitable alternate initiation site for M2 ORF-2 when the preferred initiation site is removed. If initiation was able to take place at an alternate site, it would provide evidence of a scanning mechanism used by the ribosome to initiate translation of M2 ORF-2. Alternatively, failure to initiate translation at an alternate site would indicate that the location of the AUG codon was essential for successful translation of M2 ORF-2.

Two plasmids were therefore created to test if translation of ORF-2 could be initiated at other sites if the preferred initiation codon was mutated. Both plasmids used the same reporter system to measure ORF-2 expression as used in Chapter 4 and were in the same format as the reporter plasmid pbRSVWC. The two plasmids, pbRSVAUG557 and pbRSVAUG525, introduced two separate alternate AUGs in the same frame as ORF-2 in and upstream of the ORF-1/ORF-2 overlap region. Both naturally occurring AUGs used for initiation of translation of ORF-2 in the ORF-1/ORF-2 overlap region were mutated in both plasmids by mutating AUG to CUG. As seen in Fig. 5.9, the plasmid pbRSVAUG557 had an artificial AUG codon introduced beginning at position 557 by mutating residues AC to UG at residues 558-559. This AUG codon was introduced between the two mutated AUGs in the ORF-1/ORF-2 overlap region. This was created to investigate if AUGs in close

proximity to the preferred AUG start site of ORF-2 could act as an alternate start codon. The second plasmid, pbRSVAUG524, focused on a native AUG codon beginning at position 524 in the same frame as ORF-2, there was however a stop codon beginning at position 545 in frame further downstream, which if the AUG at position 524 was used for initiation of ORF-2, would lead to premature termination of ORF-2 translation. The stop codon beginning at position 545 was therefore mutated along with both AUGs in the ORF-1/ORF-2 overlap region by mutating UAA to CAA. This was to investigate if AUGs at distance from the preferred start codon of ORF-2 would be able to act as an alternate start codon.

bRSV M2 Sequence

GAAAUAAAUGGUAAUAACCAAGGUGACA**UAA**UUGUUAAUGAACAAAAUGAA**UAA**

pbRSVAUG557

GAAAUAAAUGGUAAUAACCAAGGUGACA**UAA**UUGUUACUGAUGAAACUGAA**UAA**

pBRSVAUG524

GAAAUAAAUGGUAAUAACCAAGGUGACACAAUUGUUACUGAACAAACUGAA**UAA**

Fig. 5.9: Mutations made to alter the potential start site for initiation of translation of ORF-2
Mutations made to alter the potential start site for initiation of translation of ORF-2 for the plasmid pbRSVAUG557 and pbRSVAUG524. Sites of mutations are highlighted in red. AUG sites in frame for ORF-2 initiation are underlined. The stop codon for ORF-1 and stop codon preventing initiation of ORF-2 upstream of the ORF-1/ORF-2 overlap region beginning at position 545 are highlighted in bold. The sequences displayed, begin at position 517 and end at position 570 from the start of the bRSV mRNA sequence or the beginning of bRSV sequence in reporter plasmids.

Construction of the plasmids pbRSVAUG557 and pbRSVAUG524

The plasmids pbRSVAUG557 and pbRSVAUG524 had the relevant mutations incorporated by two-step PCR. A detailed description of the method can be seen in Section 5.3.1.1. The method for cloning and sequencing of the plasmids was the same as used in Section 5.3.1.1. For the first PCR step for each plasmid, the two primer pairs used to generate the first two amplicons for each plasmid were bRSVWTF/AUG557R (bRSV amplicon) and AUG557F/CATR (CAT amplicon) for the plasmid pbRSVAUG557 and bRSVWTF/AUG524R (bRSV amplicon) and AUG524F/CATR (CAT amplicon) for the plasmid pbRSVAUG524. The amplicons were amplified from the plasmid pbRSVDouble, which already had the two AUG codons used for initiation of ORF-2 in the ORF-1/ORF-2 overlap region mutated. For both plasmids, the two amplicons generated in the first PCR step were joined in the second PCR step using the primers bRSVWTF/CATR.

5.4.2. Expression from alternate AUG start sites for initiation of translation of M2 ORF-2

The two plasmids pbRSVAUG557 and pbRSVAUG524 were individually transfected into HEp-2 cells. The control plasmids pbRSVWC and pBluescribe were also individually transfected into HEp-2 cells. pbRSVWC acted as the positive control. Plasmids were transfected into HEp-2 cells as described in Section 2.4.1. Levels of CAT protein expressed were quantified using a CAT ELISA assay as described in Section 2.4.2. All samples were normalised against the assay results from the transfected empty vector pBluescribe plasmid. This was to account for any background signal generated in the ELISA, as the plasmid contained no CAT gene. For each sample, levels of CAT protein expressed were displayed as a percentage of CAT protein expression from the plasmid pbRSVWC.

As seen in Fig. 5.10, mutating both initiation codons for translation of ORF-2 in the ORF-1/ORF-2 overlap region and introducing a start codon beginning at position 557 in the plasmid pbRSVAUG557, reduced levels of CAT protein expression to 1.2 % of the control plasmid pbRSVWC. This significant reduction in the expression of CAT protein compared to the control ($p=0.000096$) indicates that translation of the second ORF has been reduced to nominal levels.

Mutating both initiation codons for translation of ORF-2 in the ORF-1/ORF-2 overlap region and removing a stop codon at position 545 allowing translation for the second ORF to potentially initiate at the start codon beginning at position 524 in the plasmid pbRSVAUG524 reduced levels of CAT protein expression to 0.3% of the control. This significant reduction in the expression of CAT protein in comparison to the control ($p=0.000095$) indicates that translation of the second ORF has been reduced to nominal levels.

These data indicates that the position of the AUG located at 563 is crucial for initiation of translation for the second ORF. The two artificial start codons tested at different locations failed to initiate translation of the second ORF above nominal levels. The naturally occurring AUG at position 554 had higher levels of CAT protein expression (5.4%) compared to the two artificial AUG codons (Section 4.4). However this was still an extremely low level of expression compared to the control,

which indicates the importance of initiation of ORF-2 at the AUG codon at position 563.

It could be argued that as these different AUGs trialed had different Kozak sequences, the strength of the Kozak sequence could be a mitigating factor for the decrease in translation of ORF-2. However, from Section 5.3.3, it was shown that optimal Kozak sequences have little effect on the efficiency of translation initiation for the second ORF. Both of the AUGs trialed in the sequences had similar strength Kozak sequences compared to the preferred AUG, with at least one optimal nucleotide present in the key positions either at position -3 or +4 of the start codon. This means it is unlikely that the Kozak sequence is the primary factor for the failure to initiate translation. Instead it is more likely that the AUG is in a highly specific position in relationship to an unknown factor and shifting the AUG even by several nucleotides inhibits the ribosome from initiating translation of the second ORF.

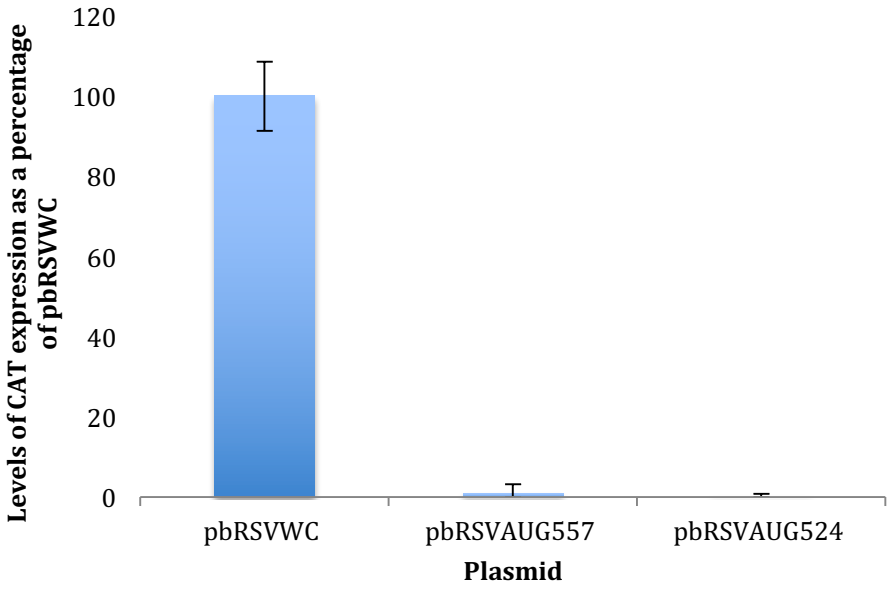


Fig. 5.10: CAT protein expression for the plasmids used to investigate alternate initiation for translation of M2 ORF-2
Levels of expression of CAT protein expressed from plasmids transfected into HEp-2 cells. Mutations made to these plasmids are described in Section 5.4.1. Results were normalised against the empty vector pBluescribe and are expressed as a percentage of the positive control plasmid pbRSVWC. Error bars indicate standard deviation.

5.5.1.1. Identification of key structured regions surrounding the start codon for ORF-2 located at position 563

Data from previous subsections strongly suggests that the position of the AUG codon used for initiation of translation of M2 ORF-2 is important. Initiation of translation of the bRSV M2 ORF-2 does not appear to be improved by improving the Kozak sequence surrounding the preferred AUG and this suggests there may be some other factor which plays a role in initiation. One possibility is the presence of secondary structure in the mRNA adjacent to the M2-ORF-2 initiation codon. As a first step, Mfold modeling software was used to predict the presence of potentially structured regions in the M2 mRNA. This modeling software has been previously used to successfully predict and identify secondary structures located in the M2 gene transcript of other members of the *Pneumoviridae* family (Gould and Easton, 2005).

Using Mfold analysis, the M2 ORF-1 sequence in the vicinity of the overlap region was assessed to identify if there was any potential RNA secondary structure present in the vicinity of the overlap region. If a structure is present in the region upstream of the AUG at position 563, it is likely to be formed by sequences from the local vicinity, as there is no evidence that sequences further away are required for translation of ORF-2. For this reason a region of 60 nts from position 511 to the M2 ORF-1 termination codon ending at position 570 was studied. A full description of settings for the software can be viewed in Section 2.9.

As seen in Fig. 5.11, the analysis of the 60nt region between residues 511-570 which includes the entire M2 ORF-1/ORF-2 overlap region indicated the likely presence of two stem loop structures with a good level of stability (minimum free energy ΔG=-5.10). One of these stem loop structures encompasses the M2 ORF-1/ORF-2 overlap region beginning at position 543 and ending at position 565. This region contains the preferred AUG codon for ORF-2 and the majority of the surrounding upstream sequence which forms base pairs within the stem of the proposed structure.

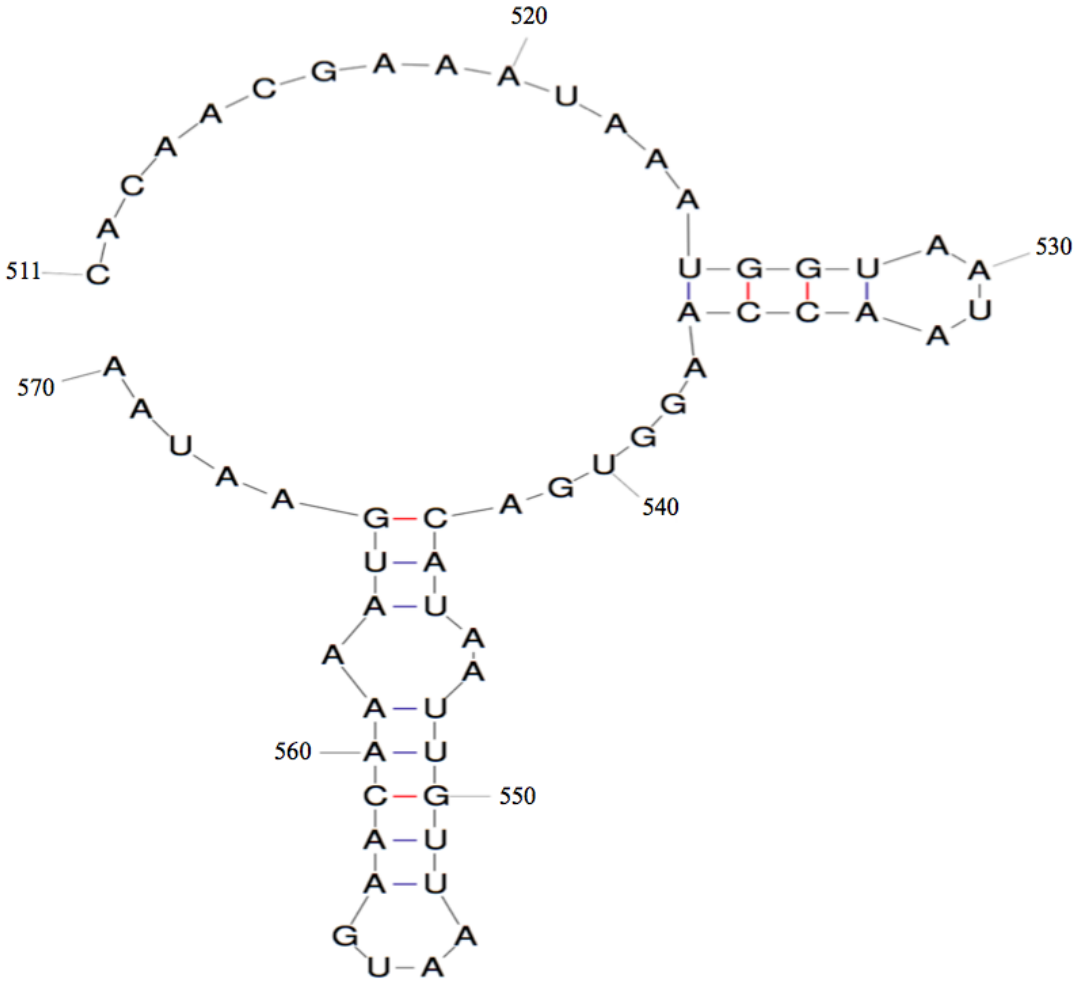


Fig. 5.11 Predicted mRNA structure generated by Mfold sofware
Predicted of mRNA structure generated by Mfold software using sequence from position 511 to 570 of the bRSV M2 mRNA. Minimum free energy ΔG=-5.10

5.5.1.2 Is secondary structure encompassing the ORF-1/ORF-2 overlap region between residues 543 and 565 involved in ORF-2 translation?

As described in Section 5.5.1.1, the RNA folding software Mfold, predicted a stem loop structure encompassing the preferred M2 ORF-2 AUG site at position 563 and the immediate upstream sequence. In order to investigate the significance of the proposed stem loop structure, the plasmid pbRSVΔ550 was generated into which a mutation was inserted to disrupt the predicted structure by mutating the G at position 550 to an A (Fig. 5.12A). Plasmid pbRSVΔ550 followed the same organization as the reporter plasmid pbRSVWC. The mutated sequence (residues 511-570) was predicted to lack the stem loop (Fig. 5.12B and Fig. 5.12C). The effect of this on translation of ORF-2 was investigated.

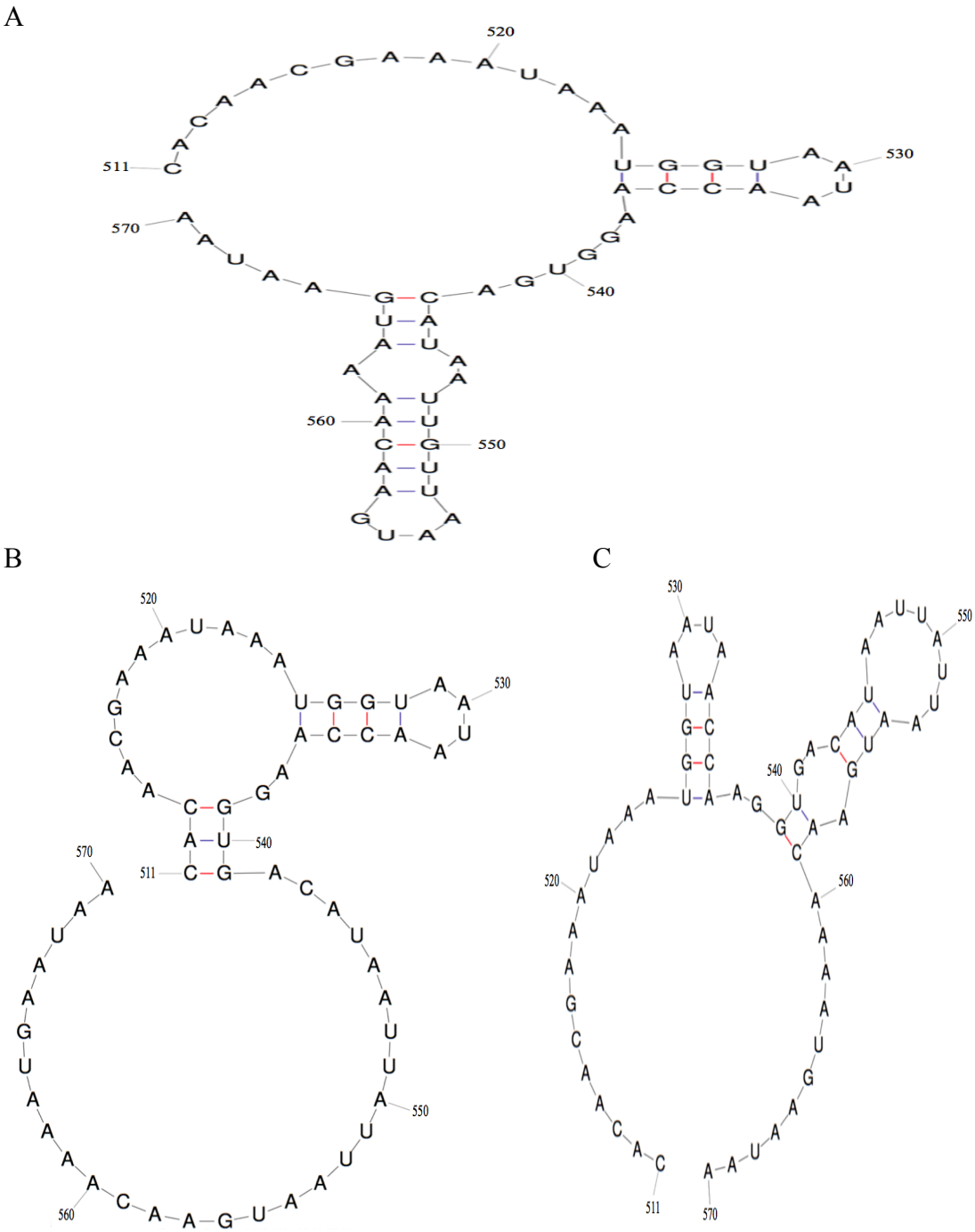


Fig. 5.12: Location of point mutation to disrupt predicted secondary structure between positions 553 and 565 in the bRSV M2 mRNA.

For all figures, the sequence analysed is from position 511 to 570 of the M2 mRNA. The Mfold predicted software predicted two alternate structures when mutated sequence was inputted. (A) Predicted structure of wild type M2 mRNA RNA (B) Predicted structure of mRNA in which residue 550 G was mutated to A (C) Alternate predicted structure of mRNA in which residue 550 G was mutated to A. All predicted structures were generated using the Mfold program.

Construction of the plasmid pbRSVΔ550

The point mutation for the plasmid pbRSVΔ550 was incorporated using two-step PCR and a detailed description of the methodology of the two-step PCR is provided in Section 5.3.1.1. The method for cloning and sequencing of the plasmids was the same as used in Section 5.3.1.1. For the first PCR step, the primer pairs used for each amplicon were bRSVWTF/Δ550R (bRSV amplicon) and Δ550F/CATR (CAT amplicon) using the plasmid pbRSVWC as a template. The two amplicons generated in the first PCR step were joined in the second PCR step using the primer pair bRSVWTF/CATR.

5.5.1.3. The effect of disruption of predicted secondary structure on translation of M2 ORF-2

The plasmid pbRSVΔ550 and control plasmids pbRSVWC and pBluescribe were transfected individually into HEp-2 cells as described previously (Section 2.4.1). Levels of CAT protein expressed were determined, normalized and expressed as a percentage of CAT protein expression from the plasmid pbRSVWC as before. Expression of CAT protein from plasmid pbRSVΔ550 was significantly increased to 223% of the control plasmid pbRSVWC ($p=0.000001$). The increase in CAT expression indicates that the sequence containing the predicted stem loop structure significantly inhibits initiation of translation of M2 ORF-2. As shown in Fig. 5.12A, the initiating AUG codon is predicted to form part of the base paired structure and disruption of the structure may make the codon more accessible thereby improving translation. However, it is clear that the stem loop region does not promote ribosome access to the initiation codon for expression from ORF-2.

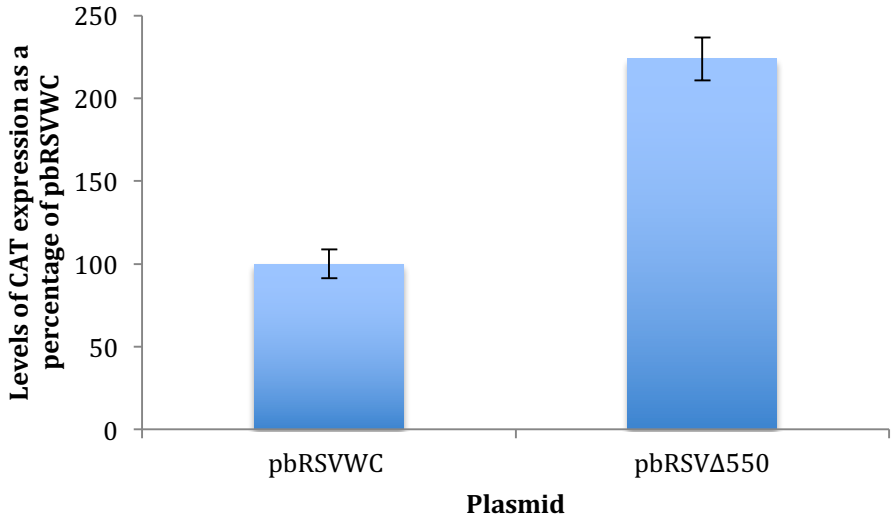


Fig. 5.13: CAT protein expression for the plasmid to investigate if disruption of predicted secondary structure between position 553 and 565 impacts ORF-2 translation
Levels of expression of CAT protein expressed from plasmids individually transfected into HEp-2 cells. Mutations made to this plasmid are described in Section 5.5.1.2. Results were normalised against the empty vector pBluescribe and are expressed as a percentage of the positive control plasmid pbRSVWC. Error bars indicate standard deviation.

5.5.2.1. Identification of highly structured regions in ORF-1 in the bRSV M2 mRNA

From the data in Section 5.5.1.3, it is unlikely expression from the initiation site at position 563 requires the upstream predicted structured region identified in the previous section. However there is the possibility that a regulatory structured region similar to that seen in the hRSV M2 mRNA is present elsewhere. Modelling secondary structures using Mfold over such a large sequence is problematic and it is preferable to refine the location of such structures using conventional biochemical analyses where possible. A feature of the highly structured region in the hRSV M2 mRNA (Gould and Easton, 2005) was that *in vitro* transcribed mRNA was a poor template for reverse transcription and that significant levels of premature termination was seen in hRSV M2 ORF-1. The site of the premature termination coincided with the highly structured region essential for translation initiation of the hRSV M2 ORF-2 (P. Gould and A. Easton, personal communication).

In order to search for a potential structured region in the bRSV M2 ORF-1, an *in vitro* transcription reaction was performed using a template generated by PCR from the plasmid pbRSVWC. This plasmid contained the entire bRSV M2 ORF-1 sequence. An amplicon generated by PCR from the hRSV plasmid pWildCAT

(containing the hRSV M2 ORF-1 sequence) (Appendix B4) was used as a positive control (Gould and Easton, 2005) (Section 2.3.1). The primer pair used for both plasmids were T7/CATInt (Appendix C). For each plasmid, a product (Fig. 5.14A) was generated containing the entirety of the appropriate virus M2 ORF-1 (M2 ORF-1) plus the M2 ORF-1/ORF-2 overlap region and a portion of the CAT gene. The templates also contained a T7 RNA promoter at the 5′ end which was used to generate the RNA. The full size RNA products from each reaction were of different sizes due to the different sizes of the virus ORF-1 (M2 ORF-1). A detailed description of the methodology for T7 *in vitro* transcription reactions is given in Section 2.7. The products of the *in vitro* transcription reactions were separated using denaturing polyacrylamide gel electrophoresis (PAGE) as described in Section 2.7.

Fig. 5.14B shows the products of both *in vitro* transcription reactions. As seen in lane 2, the hRSV template generated three RNAs. The largest RNA (1000 nt) represents the full length RNA and the band of approximately 900 nt is a premature termination product that terminates within the CAT gene sequence. The strong RNA band of approximately 400 nt results from premature termination within ORF-1 (M2 ORF-1) as described by Gould and Easton, 2005. Lane 3 shows the products of the *in vitro* transcription from the bRSV derived template. The band seen at approximately 950nt is the full length RNA product. The absence of other bands in this lane indicates that there was no premature termination of transcription. In turn, this suggests that it is unlikely that there are any highly structured regions located in the bRSV M2 ORF-1. However the possibility of the formation of less stable structures remains.

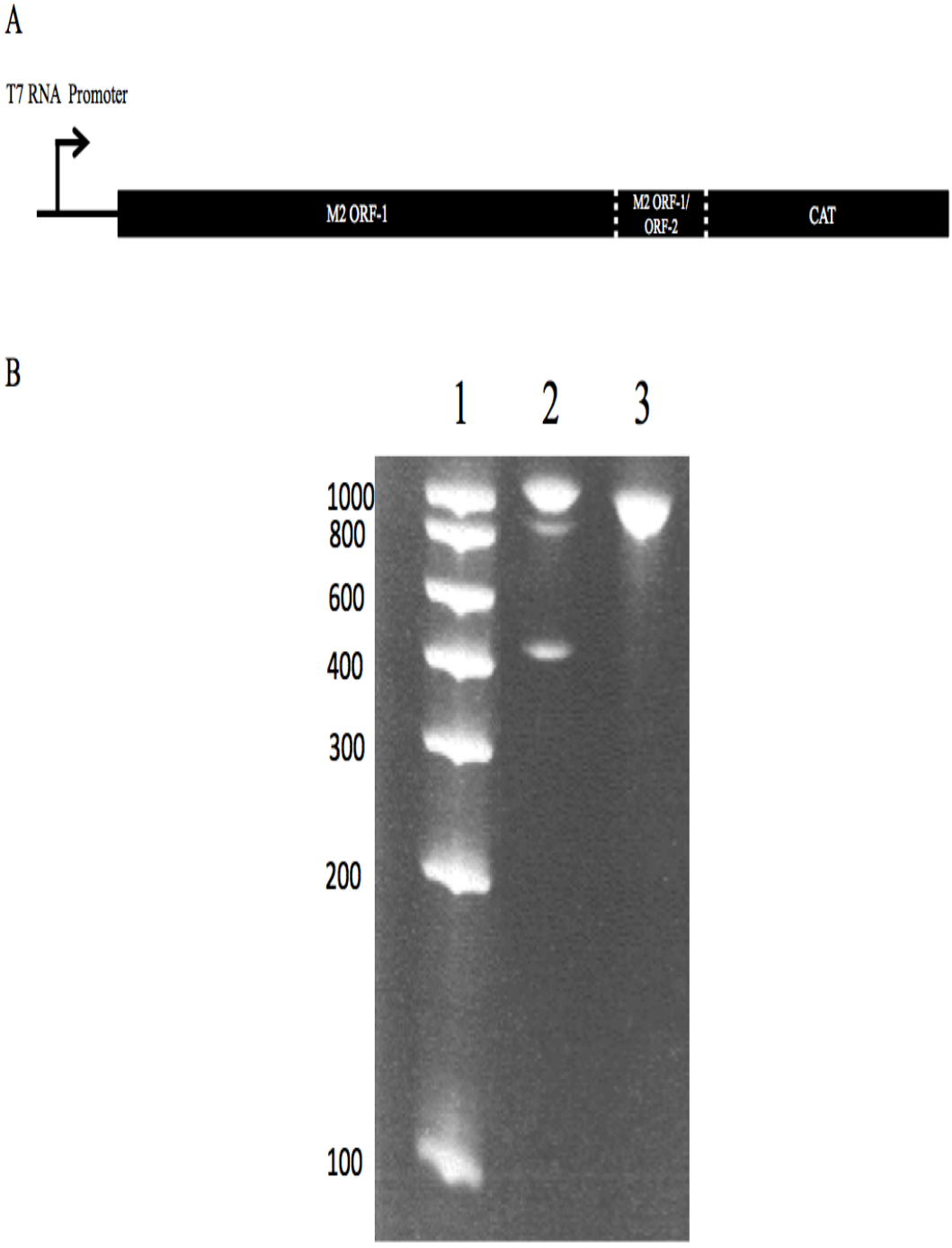


Fig. 5.14: PAGE gel of products of *in vitro* transcribed RNA from the hRSV and bRSV M2 ORF-1

A) Schematic of the hRSV and bRSV templates used in *in vitro* transcription reactions. Both amplicons include a T7 RNA promoter, the entire M2 ORF-1, the M2 ORF-1 / ORF-2 overlap region and a partial CAT gene sequence. (B) Products of *in vitro* transcription were separated on a PAGE gel. Lane 1: Geneflow 100 bp ladder. Lane 2: Transcription product from the hRSV template. Lane 3: Transcription product from the bRSV template. The sizes of the markers are shown.

5.6.1. Identification of critical regions upstream of the ORF-1/ORF-2 overlap region for the initiation of translation of the bRSV M2 ORF-2

The mechanism used to initiate translation of the bRSV M2 ORF-2 remains unclear. While analyses described above failed to identify structured regions upstream of the M2 ORF-1/ORF-2 overlap region that may play a role in initiation of translation of bRSV M2 ORF-2, the possibility remains that other elements are required. In order to identify potential critical sequences in the bRSV M2 ORF-1 for translation of bRSV M2 ORF-2, an approach based on that described for analysis of hRSV was taken (Gould and Easton, 2005). A series of reporter plasmids were constructed using the same reporter system to measure ORF-2 expression as used in Chapters 4 and 5, but in this case the gene encoding eGFP lacking the termination codon was inserted in-frame at the 5′ end of the M2-1 ORF-1, in which the M2 ORF-1 initiation codon had been deleted. In this way, the 5′ proximal ORF was extended by 722nts and expression of the plasmid's ORF-1 sequence required translation initiation from the long eGFP gene. In the plasmid series, increasing proportions of bRSV M2 ORF-1 sequence were deleted (shown schematically in Fig. 5.15A). Each deletion began at the start of the M2-1 coding sequence with the 3′ break point being progressively closer to the ORF-1/ORF-2 overlap region. The presence of the entire eGFP coding sequence ensured that the ribosomes initiating expression of ORF-1 were significant distance from the ORF-1/ORF-2 overlap region. As before, replacement of ORF-2 with the CAT gene allowed expression to be measured. In total six plasmids were created (Fig. 5.15B). For plasmids pbRSV14, pbRSV89, pbRSV209, pbRSV293, pbRSV380 and pbRSV527 14 nts, 89 nts, 209 nts, 293 nts, 380 nts and 527 nts were removed from ORF-1, respectively.

The plasmid pbRSV14Stop was also generated in which 14 nucleotides were removed from the 5′ end of the M2 mRNA in the same format as pbRSV14 and in addition, the termination codon for ORF-1 (M2-1 ORF) was mutated from UAA to CAA extending the ORF-1/ORF-2 overlap region by 42nts. This was used to explore if there was any difference in translation of ORF-2 when the stop codon of ORF-1 was removed with an additional 722 nts of eGFP sequence added to the 5′ of the ORF-1 sequence.

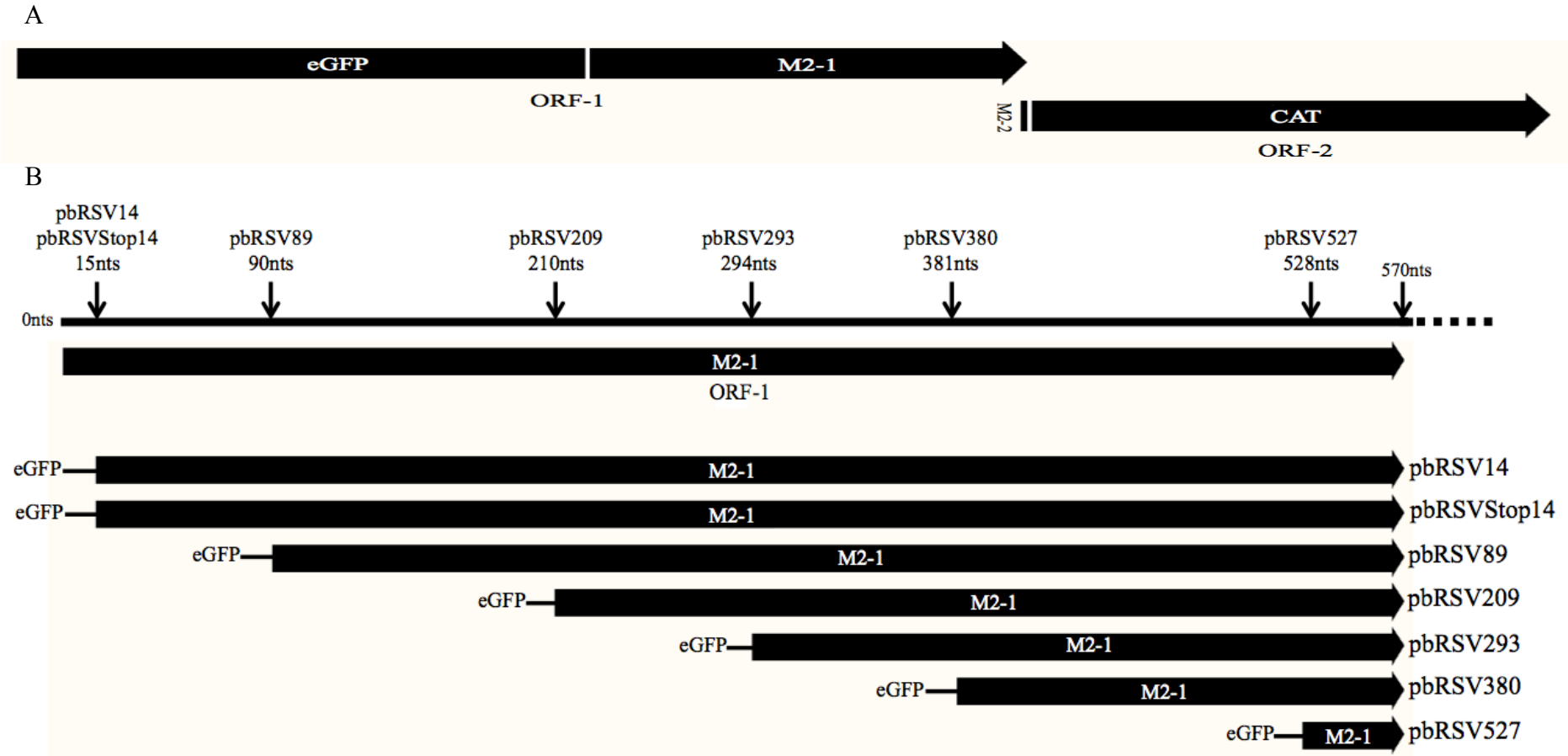


Fig. 5.15 Schematic representation of the arrangement of chimeric reporter plasmids used to investigate critical upstream regulatory sequences in ORF-1 for M2 ORF-2 translation

(A) Schematic representation of the mRNA transcribed from chimeric reporter plasmids used to investigate critical sequences in the bRSV M2 ORF-1 (M2-1 ORF) for the initiation of translation of ORF-2. The diagram shows both ORF-1 (eGFP/M2 ORF-1) and ORF-2 (M2 ORF-2/CAT ORF). (B) Schematic showing the lengths of the M2 ORF-1 coding sequence in ORF-1 for chimeric plasmids pbRSV14, pbRSVStop14, pbRSV89, pbRSV209, pbRSV293, pbRSV380 and pbRSV526 in comparison to the full M2 ORF-1 in the bRSV M2 mRNA (black line). Arrows identify the beginning of the M2 ORF-1 coding sequence for each plasmid.

Construction of the plasmids pbRSV14, pbRSV89, pbRSV209, pbRSV293, pbRSV380 and pbRSV527

The method for construction for all five plasmids was identical except for the use of the forward primer for the amplification of the truncated ORF-1/ORF-2/CAT fragments (bRSV fragment). For construction of all plasmids, two amplicons were generated which were subsequently ligated together into the vector plasmid. The first amplicon (bRSV fragment) consisted of the truncated M2 ORF-1 sequence, the M2 ORF-1/ORF-2 overlap region and the entire of the CAT coding sequence. This amplicon was amplified by PCR using the template plasmid pbRSVWC (Section 2.3.1). For each PCR reaction, the relevant forward primer used can be seen in Table 5.3. The reverse primer CAT Hind III, which bound to the 3′ end of the CAT gene, was used for every PCR reaction. Each forward primer also contained an Sph I site at the 5′ end of the sequence and the reverse primer contained a Hind III site at the 3′end of the sequence. The predicted lengths of the amplified bRSV fragments can be seen in Table 5.3. For all plasmids, the second amplicon was the eGFP fragment. This contained the eGFP coding sequence without the termination codon and was generated by PCR from the plasmid peGFP using the forward primer T7-91 and the reverse primer eGFP Sph I to generate an product of 880bp in size (Section 2.3.1). The product contained a Kpn I site at the 5′ end and an Sph I site at the 3′ end of the eGFP coding sequence. The bRSV and eGFP fragments for each plasmid were digested with the two endonuclease pairs Sph I/Hind III and Kpn I/ Sph I, respectively (Section 2.3.4 and Section 2.3.5) and were ligated together into the digested empty vector pBluescribe in a single ligation reaction (Section 2.3.7), which had been digested with endonuclease pair Kpn I/Hind III (Section 2.3.4 to Section 2.3.6). Plasmids were then transformed into DH5α bacteria and plated onto AMP/IPTG/X-Gal LB agar plates (Section 2.3.8). The bacteria were grown overnight at 37^{o}C and colonies were selected through blue white selection (Section 2.3.9). Small-scale plasmid purification was performed as described in Section 2.3.13. Purified plasmids were digested with endonucleases Kpn I and Hind III to confirm the presence of the expected fragment and the sequence was confirmed as before using the primers T7-91 and T3 (Section 2.3.10).

Construction of the plasmid pbRSV14Stop

The plasmid pbRSV14Stop was created using the same approach as above, except for the amplification of the bRSV fragment, where the primer pair was bRSV14F/CAT Hind III and the template was the plasmid pbRSVStop, which contained the mutated ORF-1 termination codon. The size of the final insert was 1974nts and the sequence was confirmed as before.

Plasmid	Forward Primer	Reverse Primer	Size of bRSV PCR amplicon generated (bp)	Size of correctly ligated insert (bp)
pbRSV14	pbRSV14F	CAT Hind III	1241	1974
pbRSV89	pbRSV89F	CAT Hind III	1165	1898
pbRSV209	pbRSV209F	CAT Hind III	1045	1778
pbRSV293	pbRSV293F	CAT Hind III	961	1694
pbRSV380	pbRSV380F	CAT Hind III	874	1607
pbRSV527	pbRSV527F	CAT Hind III	727	1460

Table 5.3: Primer pairs for each amplicon used for the first step in two-step PCR for generation of chimeric plasmids
Primer pairs used for the bRSV amplicon in the first step of two-step PCR for generation of chimeric plasmids. The size of the PCR fragment and the size of the digested insert for each plasmid once ligated into the plasmid are also included.

5.6.2. Investigation of the potential role of upstream sequences in expression of the bRSV M2 ORF-2

The plasmids created in Section 5.6.1 were individually transfected into HEp-2 cells as described in Section 2.4.1. Cells were also transfected with control plasmids pbRSVWC, pbRSVStop and pBluescribe. CAT protein expression was quantified using a CAT ELISA assay as described in Section 2.4.2 and data were normalised against the assay results from the transfected empty vector pBluescribe plasmid. As previously stated, levels of CAT protein expressed were expressed as a percentage of CAT protein expression from the plasmid pbRSVWC.

As shown in Fig. 5.16, the addition of the eGFP coding sequence to the 5′ end of the truncated M2 ORF-1 lacking the first 14 nucleotides of the M2 mRNA sequence in the plasmid pbRSV14 reduced CAT protein expression to 72.1% of the control. While this reduction was not great it was statistically significant (p=0.0016). Rather than reflect a true reduction, this is likely to be due to a general reduction in expression caused by the addition of the eGFP ORF in these new chimeric plasmids, as it is unlikely that the first 14 nucleotides have such a great effect on the initiation of translation of the second ORF.

Mutation of the ORF-1 (M2-1 ORF) termination codon the plasmid pbRSV14Stop reduced CAT protein expression to 92.2% of the control. There was no significant difference between the plasmid and the control pbRSVWC (p=0.36) and the control pbRSVStop (p=0.21). This indicates that the removal of the ORF-1 (M2-1 ORF) termination codon and increasing the length of ORF-1/ORF-2 overlap region by 42nts, had no effect on initiation of translation of the bRSV M2 ORF-2 in the M2 mRNA, even if the ORF-1 sequence is increased by 717nts. This is consistent with previous data suggesting that the ORF-1 termination codon is not a significant factor in the initiation of the bRSV M2 ORF-2 (Section 4.5.1.2). Removal of 89nts from the 5′ end of M2 mRNA in the plasmid pbRSV89, reduced CAT protein expression to 89.9% of the control (p=0.19) indicating that no essential sequence for initiation of translation of ORF-2 lies in that region. Extending the deletion of ORF-1 sequences to 209 nts from the 5′ end of M2 mRNA in the plasmid pbRSV209, reduced CAT protein expression to 78% of the control. This was significantly different to the control (p=0.0076). However, there was no significant difference between this plasmid pbRSV209 and the plasmids pbRSV14 (p=0.221) and pbRSV89 (p=0.097) which strongly suggests that no essential sequence for the initiation of translation for ORF-2 lies in the first 209 nts of the M2 mRNA.

Removal of 293nts from the 5′ end of M2 mRNA in the plasmid pbRSV293 reduced levels of CAT protein expression to 57.1% of the control (p=0.000063). In comparison to other plasmids in the series there was a significant difference between the plasmids pbRSV293 and pbRSV14 (p=0.014) and between the plasmids pbRSV293 and pbRSV89 (p=0.00013). The difference between the data from plasmids pbRSV293 and pbRSV89 suggests that the sequence between position 210

and 293 in the M2 mRNA is of importance to the initiation mechanism M2 ORF-2 but that loss of this region is not sufficient to completely inhibit initiation of translation of M2 ORF-2.

Extending the length of deletion in plasmid pbRSV380 further reduced levels of CAT protein expression to 45.7% of the control (p=0.000016). Compared to the other plasmids in the series the reduction in expression was significantly different to pbRSV14 (p=0.00010), pbRSV89 (p=0.0000062) and pbRSV293 (0.016). Therefore since a significant difference was observed between pbRSV380 and both plasmids that express CAT protein at basal levels for these new chimeric plasmids and between pbRSV380 and pbRSV293, it can be concluded that the sequence in this region is of importance to the mechanism of initiation of translation of M2 ORF-2. However, this region of sequence is still not critical for the initiation of translation for the second ORF.

Removal of 527nts from the 5´ end of M2 mRNA in the plasmid pbRSV527, reduced levels of CAT protein expression for the second ORF to 1% of the control (p=0.0000067). There was also a significant difference between the data from plasmid pbRSV527 and that from plasmids pbRSV14 (p=0.00000012), pbRSV89 (p=0.00000013), pbRSV293 (p=0.00000017) and pbRSV372 (0.00000073). This indicates that between position 381 and position 527, there is a region of sequence which is essential for the initiation of translation of bRSV M2 ORF-2. Whether the role of this region is linked to other regions further upstream, which also affected ORF-2 expression is not known.

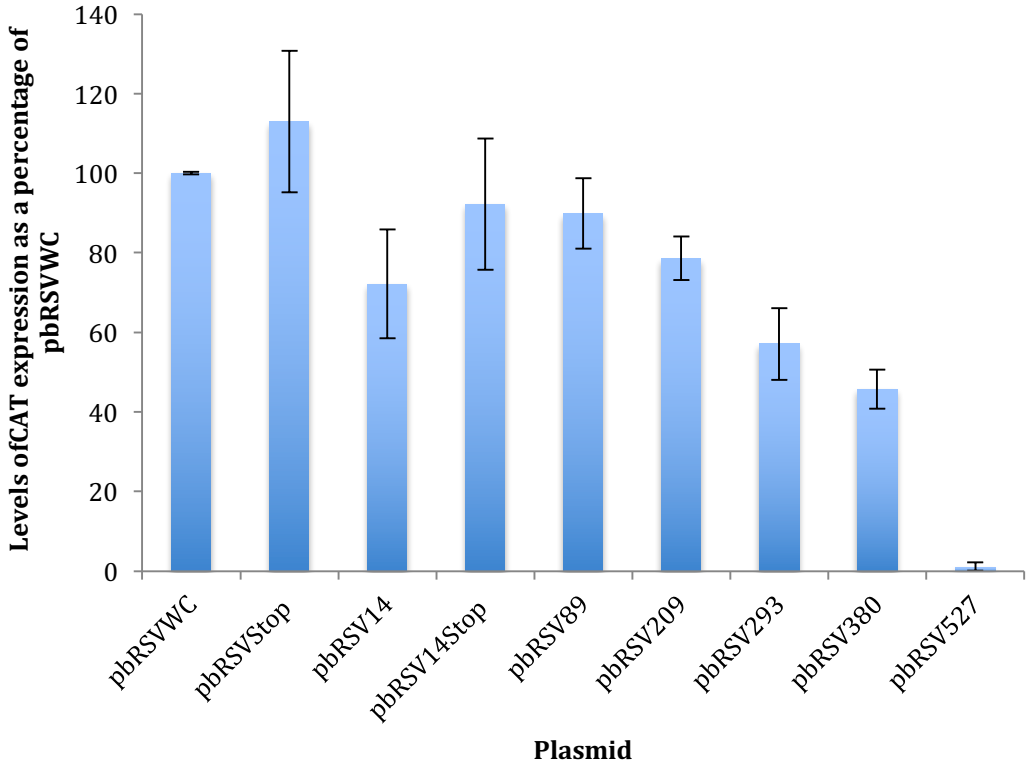


Fig. 5.16: CAT protein expression from ORF-2 from plasmids containing increasing lengths of deletion in M2 ORF-1

Levels of expression of CAT protein expressed from plasmids used to investigate critical sequences in ORF-1 for ORF-2 translation that were individually transfected into HEp-2 cells. As described in Section 5.6.1, each plasmid had increasing proportions of sequence deleted from the 5′ end of ORF-1 and replaced with eGFP. The number in the plasmid name describes the position of the break point between M2 ORF-1 and eGFP coding sequence used to replace the deleted upstream M2 ORF-1 sequence up to this point. Results were normalised against the empty vector pBluescribe and are expressed as a percentage of the positive control plasmid pbRSVWC. Error bars indicate standard deviation.

5.7.1. Investigation of ribosomal shunting as a mechanism for initiation of translation for the bRSV M2 ORF-2 (M2-2 ORF)

Ribosomal shunting is a cap-dependent mechanism used by many viruses to bypass stem loop structures that are common in 5′ UTRs to reach a desired AUG codon. This mechanism which is discussed in more detail in Section 1.6.4, can be also used in conjunction with cap-dependent scanning to initiate translation at a desired AUG codon. In some cases the shunting mechanism can bypass large sections of sequence, which may include entire ORFs. It has been reported that adenoviruses use shunting in conjunction with scanning in late mRNAs in which the shunting process is directed by the common tripartite leader sequence (reviewed in Section 1.6.4) (Yueh and Schneider, 2000). Inside the tripartite leader region are three sequences which are complementary to the 3′end of the 18S rRNA. Removal of these three sequences significantly reduces the efficiency of the shunting mechanism although it has been speculated that this is only one of the factors that enables the shunting mechanism to function (Yueh and Schneider, 2000).

It is unlikely that bRSV uses ribosomal shunting in the classical sense of bypassing and translating a short ORF in the 5′ UTR, as the M2 5′ UTR is only 9 nts long and does not contain an upstream short ORF. However it is possible that the ribosome could use ribosomal shunting to bypass M2 ORF-1 to directly access ORF-2. As seen in Fig. 5.17A, significant regions in the vicinity of the bRSV M2 ORF-1/ORF-2 overlap region aligned with the 3′end of the 18S rRNA. This alignment was in a similar location to where the adenovirus tripartite leader complementary sequence (referred to as C1) bound to the 3′end of the 18S rRNA (residues 1854-1862 and 1836-1847 on the 18S rRNA) (Yueh and Schneider, 2000). It was reported that removal of the C1 region had only a small effect on the shunting mechanism. However, removal of all three regions led to inhibition of the shunting mechanism.

Two plasmids were generated to investigate whether the regions of complementarity to the 18S rRNA in the M2 mRNA are involved in the initiation of translation of the bRSV M2 ORF-2. Both plasmids followed the same format as the reporter plasmid pbRSVWC but each contained specific mutations to alter the sequences that may base pair with 18S rRNA. As seen in Fig. 5.17A, the first plasmid pbRSVShuntK, was used to investigate the significance of this complementary sequence to the 3′

end of the 18S rRNA by mutating residues at positions 548-554 and 556-558 to reduce the complementary between the 3′ end of the 18S rRNA sequence and the M2 ORF-1 sequence. The second plasmid pbRSVShuntP, was created to investigate if increasing the potential interaction with the ribosome would increase the levels of translation of ORF-2. This was achieved by mutating residues at positions 554, 557, 559, 561 and 562 to increase the complementarity between the 3′ end of the 18S rRNA and the ORF-1 sequence (Fig. 5.17A).

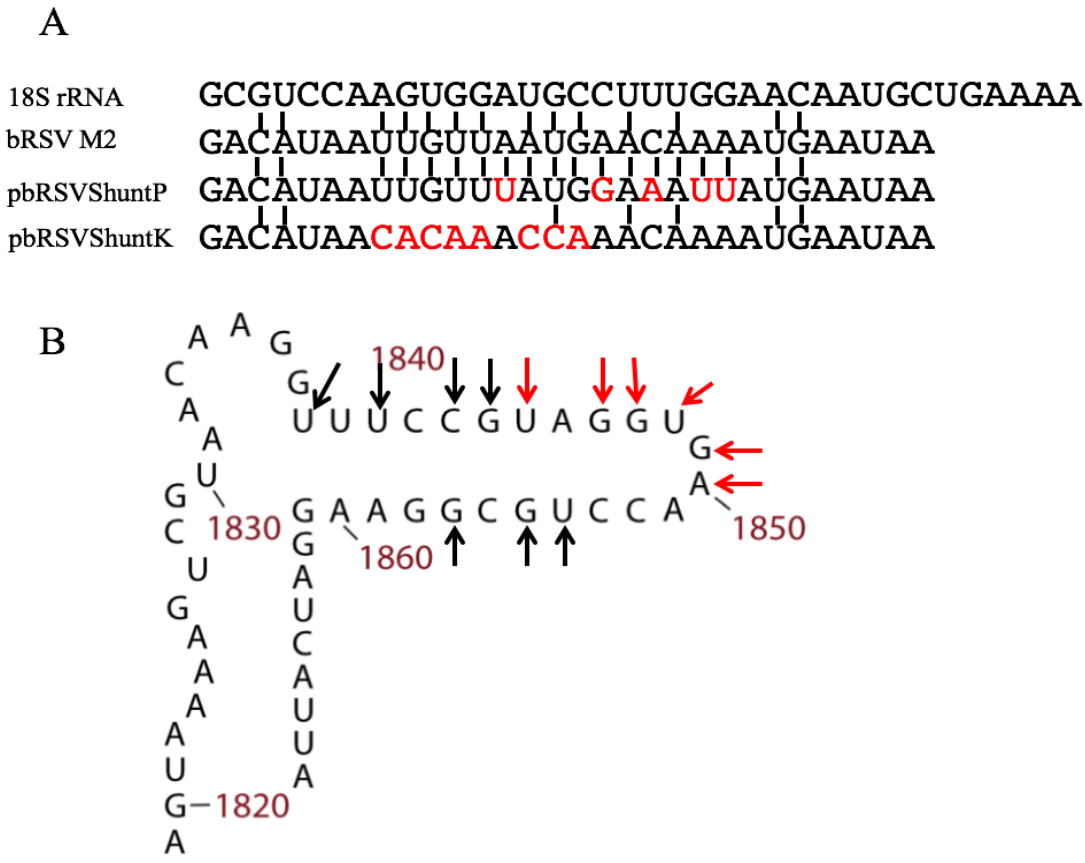


Fig. 5.17: Alignment of 18S rRNA against M2 mRNA and against sites of mutations to either disrupt or improve potential binding of the M2 mRNA to 18S rRNA
(A) Alignment of 18S rRNA to potential binding sites for M2 mRNA and plasmids pbRSVShuntP and pbRSVShuntK. Mutations for both plasmids are highlighted in red. Sequence for 18S rRNA runs from the 3′ to 5′ from position 1857-1822. bRSV M2 mRNA sequence runs from 5′ to 3′ from position 541-570 (B) Site of bRSV M2 potential binding sites on 18S rRNA. Arrows indicate potential binding sites. Red arrows indicate sites mutated to prevent binding in plasmid pbRSVShuntK

Construction of the plasmids pbRSVShuntK and pbRSVShuntP

Mutations were created using two-step PCR as described in Section 5.3.1.1. The method for cloning and sequencing of the plasmids was the same as used in Section 5.3.1.1. For the first PCR step for each plasmid, the primer pairs used for the first

PCR step for the plasmid pbRSVShuntK were bRSVWTF/ShuntKR (bRSV amplicon) and ShuntKF/CATR (CAT amplicon). For the plasmid pbRSVShuntP, the two primer pairs used for the first PCR step were bRSVWTF/ShuntPR (bRSV amplicon) and ShuntPF/CATR (CAT amplicon). For both plasmids, the products were amplified from the plasmid pbRSVWC. The two fragments generated in the first PCR step for each plasmid, were joined in the second PCR step using the primer pair bRSVWTF/CATR.

5.7.2. Expression of the plasmids used to investigate if ribosomal shunting is used for initiation of M2 ORF-2

The two plasmids pbRSVShuntK and pbRSVShuntP and control plasmids pbRSVWC and the empty vector pBluescribe were individually transfected into HEp-2 cells (Section 2.4.1) and CAT protein expression was measured with a CAT ELISA as described in Section 2.4.2. As in previous sections, protein levels were expressed as a percentage of CAT protein expression from the control plasmid pbRSVWC.

As shown in Fig. 5.18, mutation of the sequence in ORF-1 complementary to the 3′ end of the 18S rRNA in the plasmid pbRSVShuntK increased CAT protein expression to 133.3% of the control ($p=0.0053$). This shows clearly that there is no requirement for homology to the 18S rRNA for ORF-2. Mutating the sequence in ORF-1 to increase the degree of complementarity to the 3′ end of the 18S rRNA in the plasmid pbRSVShuntP also increased CAT protein expression from the second ORF to 125.3% of the control ($p=0.021$). There was no significant difference in CAT protein expression between the plasmids pbRSVShuntK and pbRSVShuntP ($p=0.20$). Taken together these data show that there is no requirement for base pairing between 18S rRNA and bRSV M2 mRNA. This strongly suggests that ribosomal shunting is not used for expression of the bRSV M2 ORF-2.

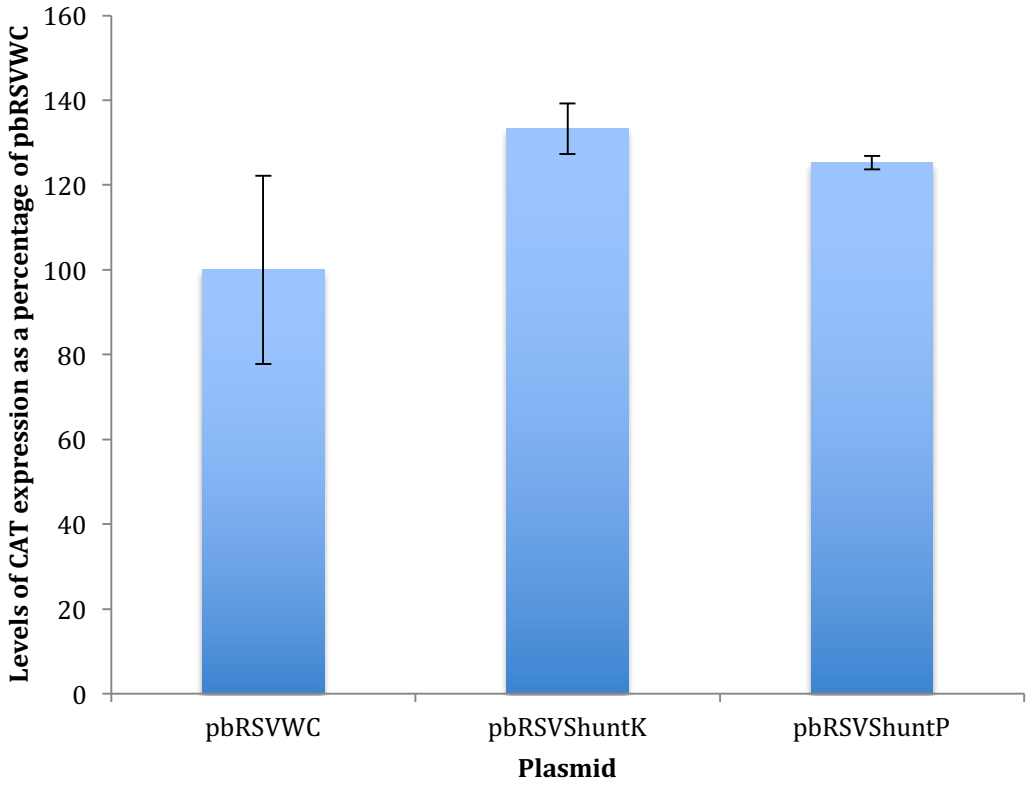


Fig. 5.18: CAT protein expression for plasmids used to investigate if ribosomal shunting is used for initiation of translation of M2 ORF-2

Levels of expression of CAT protein expressed from plasmids individually transfected into HEp-2 cells. Mutations made to these plasmids are described in Section 5.7.1. Results were normalised against the empty vector pBluescribe and are expressed as a percentage of the positive control plasmid pbRSVWC. Error bars indicate standard deviation.

5.8.1. Investigation of internal initiation as a mechanism to initiate translation of the bRSV M2 ORF-2 (M2-2 ORF)

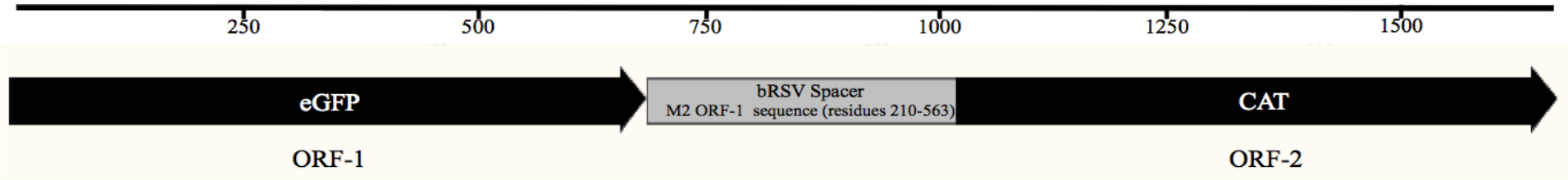
The role of internal ribosome entry sites (IRES) is discussed in Section 1.6.5. IRES regions are used in both viral and eukaryotic mRNAs to initiate translation in a cap-independent manner. The IRES is comprised of a sequence in mRNA that can bind the ribosomal 43S complex in close proximity to the desired AUG codon. For some IRES regions, this mechanism can occur without the involvement of many of the initiation factors required for cap-dependent translation. IRESs tend to be located in the 5′ UTR such as in poliovirus (Pelletier and Sonenberg, 1988) and hepatitis C virus (Tsukiyama-Kohara et al., 1992). They have also have been located inside translated ORFs of polycistronic mRNAs, in which only one of the ORFs uses the IRES for initiation (Kang et al., 2009).

A standard method frequently used to test for internal initiation sequences, including IRES, is through the use of bicistronic reporter plasmids (Pelletier and Sonenberg, 1988). This method works on the principle that these plasmids transcribe a bicistronic mRNA, which contains two separate ORFs each expressing a different reporter protein. The ORFs do not overlap each other and are separated by a spacer sequence containing the putative internal initiation sequence that will not be translated. A control plasmid is also generated where random sequence is used as the spacer sequence. The principle is that the first ORF will initiate translation through a cap-dependent process but terminating ribosomes for the first ORF will never initiate translation of the second ORF due to the presence of the spacer sequence. Expression from the second ORF will only be detected if the test sequence is able to promote internal initiation of translation. This approach was used to investigate the bRSV M2 mRNA. However, as the two ORFs overlap with each other, the sequence potentially directing internal initiation would be located within ORF-1 upstream of the initiation codon for ORF-2.

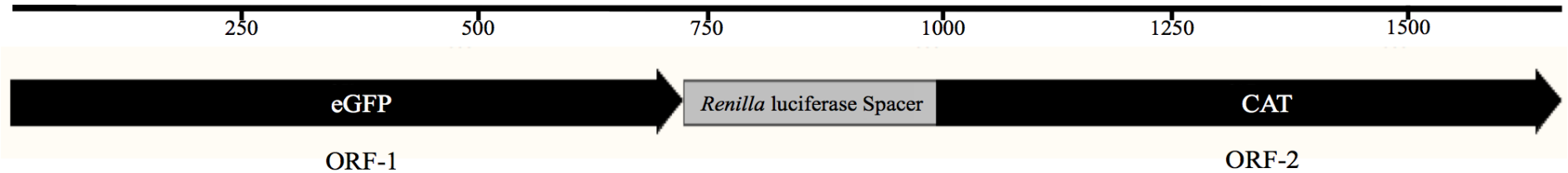
A reporter plasmid, pbRSVIRES, was created to test for internal initiation. As seen in Fig. 5.19A, the plasmid contained the full eGFP ORF as the first ORF. The spacer region and second ORF sequence were taken from the plasmid pbRSVWC. The untranslated spacer sequence consisted of the bRSV M2 ORF-1 sequence beginning at position 210 and continuing until position 562. The second ORF contained the

CAT ORF from the plasmid pbRSVWC (Fig. 5.19C), where the CAT ORF initiation codon at 5′ end of the ORF was replaced with 8nts of bRSV M2 mRNA sequence between position 563 and 570. The AUG codon from position 563 would be used as the initiation codon for the translation of the CAT ORF in this plasmid. The length of the spacer sequence was chosen based on the analysis of the role of upstream sequences described in Section 5.6.3. A control plasmid pRenLucIRES (Fig. 5.19B) was also created. The second ORF in pRenLucIRES consisted of the same CAT ORF used in the plasmid pbRSVIRES, but here the CAT ORF AUG start codon was replaced with the bRSV M2 ORF-2 AUG codon (residues 554-570). This also included the upstream Kozak sequence used for this AUG codon. The *Renilla* luciferase coding sequence (254nts: position 317-570 of the *Renilla* luciferase coding sequence) was used as the control untranslated spacer sequence.

A



B



C

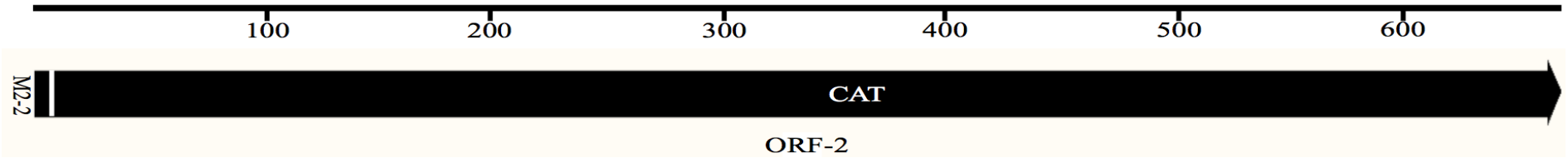


Fig. 5.19: Schematic of the mRNAs to investigate potential internal initiation of translation of M2 ORF-2

(A) Schematic of the mRNA transcribed from the reporter plasmid pbRSVIRES. Figure includes both eGFP and CAT ORFs (black arrows) and untranslated bRSV spacer sequence (grey box) derived from residues 210-563 on the bRSV M2 mRNA. (B) Schematic of the mRNA transcribed from the reporter plasmid pRenLucIRES. Figure includes both eGFP and CAT ORFs (black arrows) and *Renilla* luciferase spacer (grey box) derived from residues 317-570-of the *Renilla* luciferase coding sequence. (C) Schematic of ORF-2 of reporter plasmids pbRSVIRES and pRenLucIRES. CAT ORF initiation codon has been replaced with the bRSV M2 ORF-2 5′ codon residues (563-570 pbRSVIRES) and (554-570 for pRenLucIRES).

Construction of the plasmid pbRSVIRES

For construction of the plasmid pbRSVIRES, two amplicons were generated by PCR. The first was the eGFP fragment containing the eGFP ORF which was amplified by PCR using the primer pairs T7-91/eGFPIRESR from the plasmid pbRSV89 (Section 2.3.1). The primer eGFPIRESR also reintroduced the eGFP termination codon into the fragment. The eGFP fragment contained a Kpn I and Sph I site at the 5′ and 3′ ends, respectively. The second amplicon was the bRSV fragment, containing the bRSV spacer sequence and the CAT reporter coding sequence. The fragment was generated by PCR using the primer pairs bRSVIRESF/CATR from the plasmid pBRSV89 (Section 2.3.1). The fragment contained a Sph I site and a Hind III site at the 5′ and 3′ ends, respectively. The eGFP fragment was digested with the endonuclease pair Kpn I/Sph I and the bRSV fragment with the endonuclease pair Sph I/ Hind III (Section 2.3.4 and Section 2.3.5). The two fragments were ligated simultaneously into the plasmid pBluescribe (Section 2.3.7) previously digested with endonuclease pair Kpn I/Hind III (Section 2.3.4 to Section 2.3.6). The ligated plasmid was then transformed into *E.coli* DH5α bacteria and recombinant plasmids were identified as before. Purified plasmids were digested with endonucleases Kpn I and Sph I to confirm the presence of the desired insert (1785nts) and the sequence was confirmed.

Construction of the plasmid pRenLucIRES

A DNA fragment containing the eGFP ORF was amplified by PCR using the primer pairs T7-91/eGFPIRESR from the plasmid pbRSV89 (Section 2.3.1). The primer eGFPIRESR also reintroduced the eGFP stop codon into the fragment. The eGFP fragment contained a Kpn I and Sph I site at the 5′ and 3′ ends, respectively. A second fragment containing the combined *Renilla* Luciferase spacer sequence and the CAT reporter ORF was created using a two-step PCR. A detailed description of the methodology of two-step PCR is given in in Section 5.3.1.1. For the first PCR step the spacer amplicon, containing the *Renilla* luciferase sequence, was generated using the primer pair RenLucF/RenLucR and was amplified from the plasmid pbRSVRen. A fragment containing the CAT ORF was amplified using the primer pair bRSVCATF/CATR. The two fragments were joined using the primer pair RenLucF/CATR to produce the final reporter gene fragment which contained an Sph I and Hind III site at the 5′ and 3′ ends, respectively. The eGFP and CAT fragments

were digested with the endonuclease pairs Kpn I/Sph I and Sph I/Hind III as appropriate (Section 2.3.4 and Section 2.3.5). The resulting fragments were simultaneously ligated into the plasmid pBluescribe (Section 2.3.7) previously digested with endonuclease pair Kpn I/Hind III (Section 2.3.4 to Section 2.3.6) and the product was transformed into *E.coli* DH5α bacteria. Recombinant plasmids were identified by colour selection, the inserts confirmed by restriction digestion and the sequence determined.

5.8.2. Expression of plasmids to investigate a potential internal initiation mechanism to initiate translation of M2 ORF-2

HEp-2 cells were transfected individually with pbRSVIRES, pRenLucIRES or the control plasmids pbRSVWC and pBluescribe as described in Section 2.4.1. CAT protein expression was quantified using a CAT ELISA assay (as described in Section 2.4.2) and data were normalised against the assay results from the transfected empty vector pBluescribe plasmid. As stated previously, levels of CAT protein expressed were expressed as a percentage of CAT protein expression from the plasmid pbRSVWC. To control for potential variation in CAT protein expression due to differences in transfection efficiencies between individual samples, the levels of eGFP expressed from ORF-1 in the plasmids pbRSVIRES and pRENLucIRES were determined using a GFP ELISA as described in Section 2.4.3. This was used to normalise the levels of eGFP protein between different samples, thereby allowing the levels of CAT protein detected to be also adjusted by the same values, which allowed comparison of ORF-2 expression from the various samples. The levels of eGFP protein expression from each set of transfections for each plasmid are found in Appendix (B3)

As seen in Fig. 5.20, expression of the CAT protein in the second ORF with the *Renilla* luciferase spacer sequence in the plasmid pRenLucIRES was only 5.6% of the control ($p=0.00018$). Expression of the CAT protein in the second ORF with the bRSV spacer in the plasmid pbRSVIRES was 62% of the control. While this was a significant drop in comparison to the control ($p=0.003$) a significant difference was seen between expression from pRenLucIRES and pbRSVIRES ($p=0.0000059$). This increase in expression from pRenLucIRES to pbRSVIRES suggests that an IRES or

some form of internal initiation site is present between position 210 and position 570 that is used for initiation of M2 ORF-2.

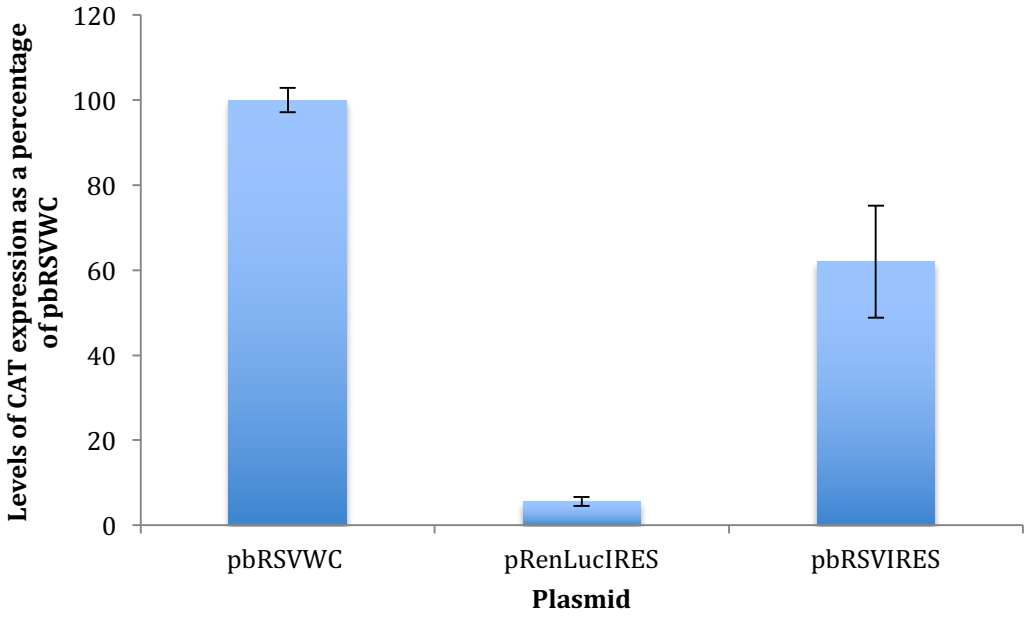


Fig. 5.20: CAT protein expression for plasmids used to investigate if an internal initiation mechanism is used for initiation of translation of M2 ORF-2
Levels of expression of CAT protein expressed from plasmids individually transfected into HEp-2 cells. Mutations made to these plasmids are described in Section 5.8.1. Results were normalised against the empty vector pBluescribe and are expressed as a percentage of the positive control plasmid pbRSVWC. Error bars indicate standard deviation.

5.9. Translational profile of the M2 mRNA to identify sites of internal initiation for translation of M2 ORF-2

Ribosomal profiling was used to investigate if there were regions in the bRSV M2 mRNA with a low distribution of elongating ribosomes, which may indicate the presence of an IRES or another form of internal initiation site. A detailed description of the technique of ribosomal profiling is given in Section 4.6.1. Analysis of the bRSV M2 mRNA would be more complex as any putative internal entry site would lie within the M2 ORF-1 which is being actively translated. Two ribosomal profile sequence libraries were generated using RNA extracted from MDBK cells that had been infected with bRSV Snook strain. One library was generated using RNA extracted at 4 hpi and the other at 8 hpi (sequencing libraries bRSV4P and bRSV8P, respectively). These sequence libraries contained data only originating from the protected footprints of elongating ribosomes. A description of the methodology of ribosomal profiling and the sequencing of these libraries is given in Section 2.10. The processing pathway to analyse the raw sequencing data is in Section 3.2. All profiles were prepared using the IGV software package, and the spectra represent the distribution of elongating ribosomes on the M2 mRNA. Each point represents the first nucleotide of each sequence read aligned to the M2 mRNA.

For the RNA harvested at 4 hpi (bRSV4P), a total of 6,774,318 reads were obtained of which 13757 sequence reads (0.2%) aligned to the bRSV genome. Of the virus-specific sequences 407 were from the M2 mRNA. For the RNA harvested at 8 hpi (bRSV8P), from a total of 6,648,581 reads, 67977 (1.0%) aligned to the bRSV genome and of these, 1871 reads were from the M2 mRNA. The bRSV8P sequencing library was used to generate a translational profile of the M2 mRNA due to the higher level of reads aligned to the gene. All profiles presented in this section were created using data from the bRSV8P sequencing library

The data presented in Fig. 5.21 shows the distribution of elongating ribosomes in the M2 ORF-1 (M2-1 ORF). The profile has also been separated into segments, which represent the locations of the junction points between the eGFP coding sequence and M2-1 ORF sequences for the chimeric plasmids discussed in Section 5.6. For each of these plasmids the sequence upstream of the appropriate line was deleted and replaced by eGFP coding sequence. Fig. 5.22A shows the distribution frequency of

elongating ribosomes in the region of the M2 ORF-1 that was used as the spacer sequence for the plasmid pbRVIRES. This region, located between nucleotides 210 and 570 of the bRSV M2 ORF-1 was identified as containing an internal initiation site or IRES. Fig. 5.22B shows the frequency distribution of elongating ribosomes in the region of the M2 ORF-1 from position 380 to 527. From the results of the transfection of the plasmid pbRSV527 in Section 5.6, it was apparent that between positions 380 and 527, there was sequence critical for the initiation of translation of the bRSV M2 ORF-2 (M2-2 ORF).

As seen in both Fig. 5.22A and Fig. 5.22B between residues 380-428 (area of interest highlighted in black box), there are two areas in which very few reads were found (residues 380-396 and 419-428). In comparison to the entire M2 ORF-1, there were no similar gaps where such low frequencies of reads were aligned to M2 ORF-1. Between these two regions of low frequency (residues 409-414), there were several nucleotides where a high frequency of reads aligned. For these peaks observed, the high frequency of reads aligned in this region were continuous over a 5 nucleotide stretch. This was not observed in any other high frequency peaks within the M2 gene transcript.

This region of interest is located in the spacer sequence used in pbRSVIRES, the majority of this sequence is also in the critical region identified in pbRSV527 (position 381-528). The sizes of these low frequency regions were much smaller (below 20 nts) than observed in translational profiles of cricket paralysis virus IRESs (Khong et al., 2016). However, if an IRES was used in bRSV, then elongating ribosomes translating the M2 ORF-1 were likely present throughout this sequence. Therefore footprints could be aligned across sections of this entire critical structure, which would mask a much larger region compared to if the IRES was outside the ORF. These two low frequency regions may be the only locations that directly affect elongating ribosomes translating the M2 ORF-1 resulting in stacking of ribosomes either side of that element. Although this only speculative it could be plausible evidence for a site where this internal initiation mechanism is located.

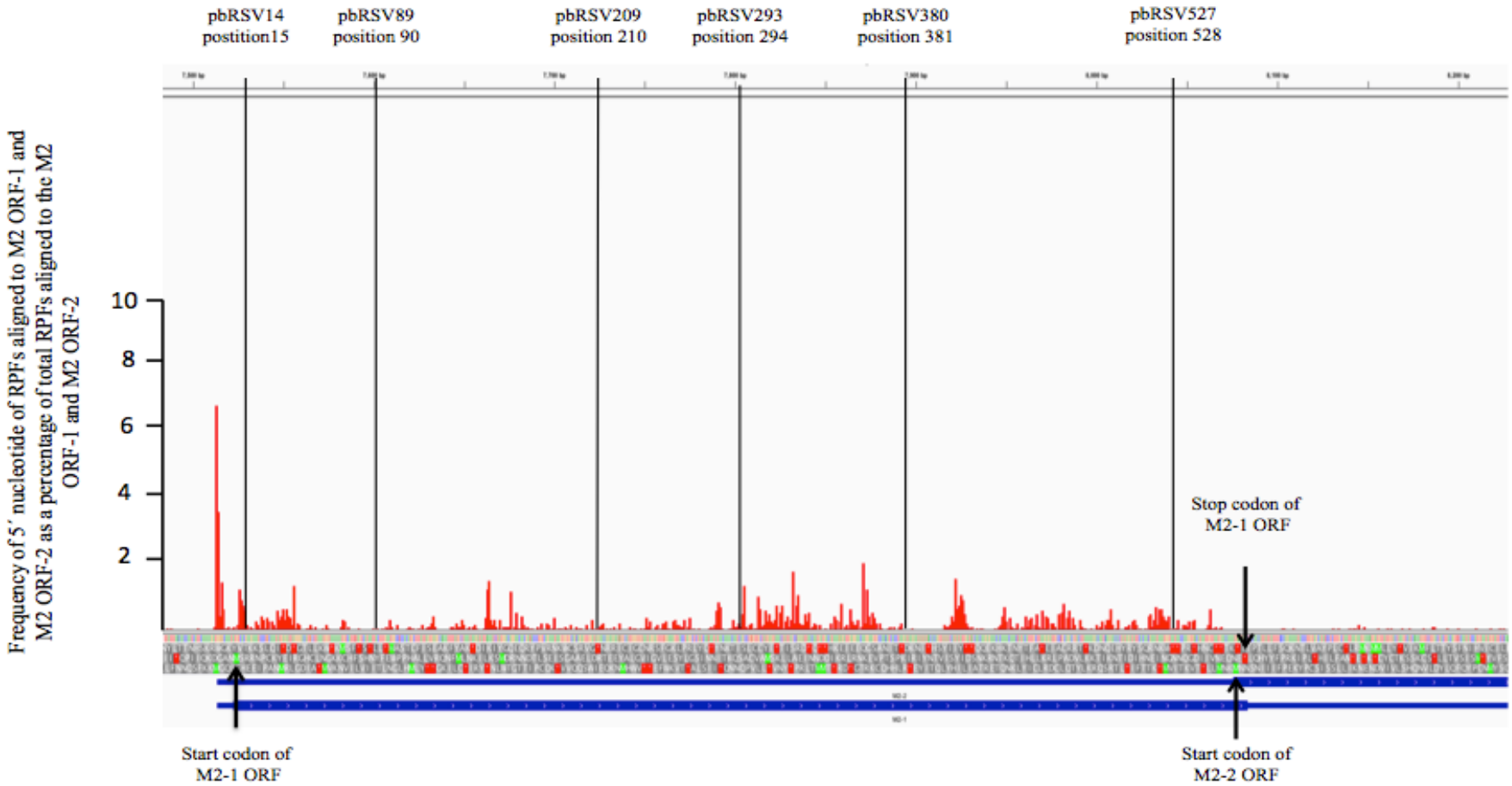
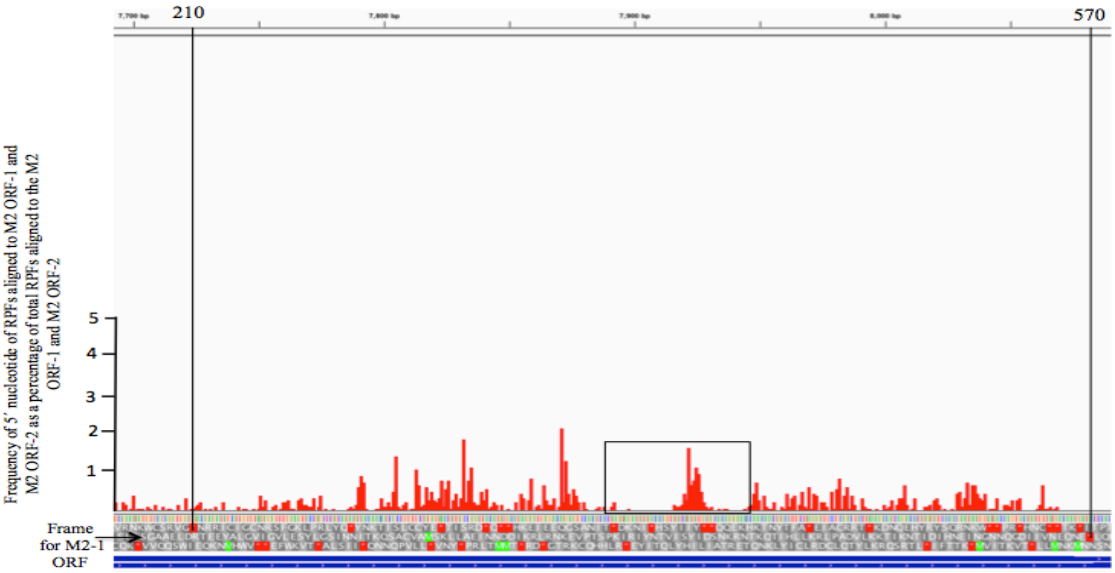


Fig. 5.21: Translational profile of the bRSV M2 mRNA showing sites of deletions in plasmids described in Section 5.6
Translational profile of the bRSV M2-1 ORF. Location of all putative start (green) and stop (red) codons in all three frames in the M2 ORF-1 are found below the spectra. Locations of preferred start and stop codons for M2 ORF-1 and M2 ORF-2 are indicated. Lines indicate locations of where eGFP was fused to the M2-1 ORF for the chimeric plasmids described in Section 5.6. In these plasmids sequences upstream of the line was deleted for that plasmid and replaced with eGFP.

A



B

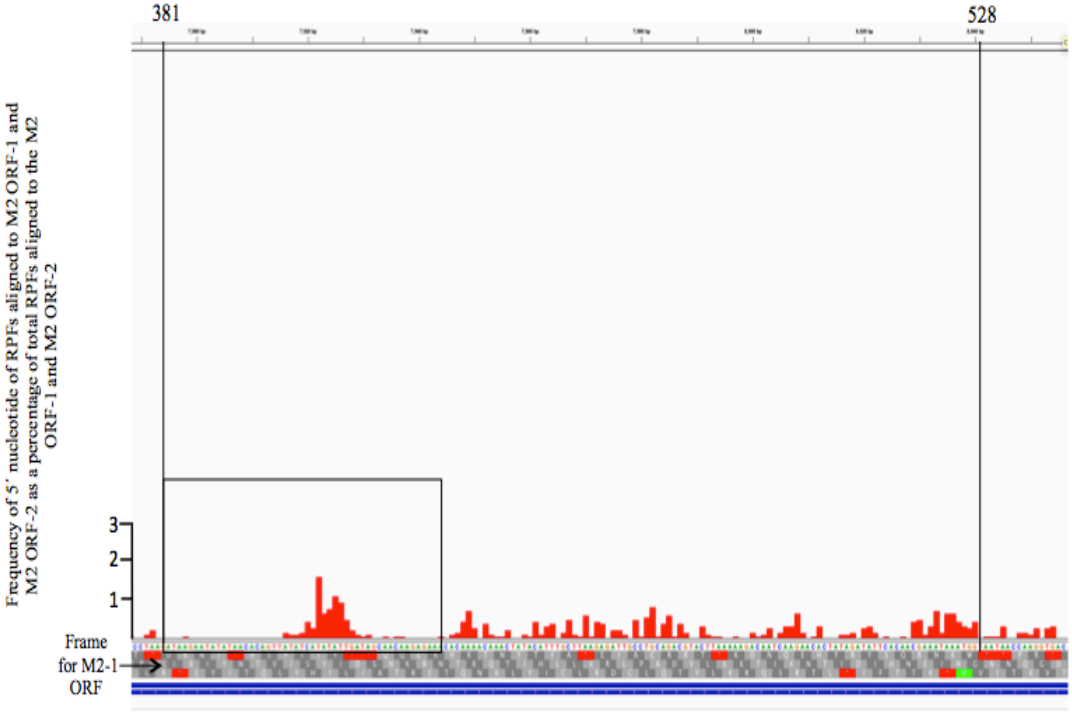


Fig. 5.22: Translational profile of region inside the bRSV M2 ORF-1 that was critical for translation of bRSV M2 ORF-2

(A) Translational profile of the bRSV M2 ORF-1. Location of all putative start (green) and stop (red) codons in all three frames in the M2 ORF-1 are found below the spectra. Lines indicate region of sequence that was used as a spacer sequence (residues 210-570) for the plasmid pbRSVIRES in Section 5.9. Box is used to highlight area of interest discussed in this Section. (B) Translational profile of the bRSV M2 ORF-1. Location of all putative start (green) and stop (red) codons in all three frames in the M2 ORF-1 are found below spectra. Lines indicate region of sequence that when deleted from chimeric plasmid pbRSV527 in Section 5.6, led to an inhibition of translation for the second ORF (residues 381-528). Box is used to highlight area of interest discussed in this this section.

5.10. Conclusions

Data acquired in this chapter has further characterised the unknown mechanism used by bRSV for initiation of translation of M2-ORF-2. From the data, it can concluded that initiation of translation of M2 ORF-2 must occur at the AUG codon located at position 563 and both the Kozak sequence and the location of the AUG codon were found to influence this preferred site for initiation of M2 ORF-2.

It is clear that the Kozak sequence is a factor in translation of M2 ORF-2. Introducing a Kozak sequence of poor context for the AUG codon at position 563, significantly reduced protein expression of ORF-2. Furthermore mutating the -3 position in the Kozak sequence from the optimal nucleotide to less preferable nucleotides, also decreased expression in the same order as the preference of each nucleotide for that position. However in both experiments, ORF-2 expression was never fully inhibited. This suggests that the Kozak sequence was not the defining factor that allows initiation of translation to occur at this position. This was further supported with the introduction of an optimal Kozak sequence for the preferred AUG codon. This new sequence either did not improve or may have slightly lowered CAT protein expression from ORF-2. This would suggest that the initiation sequence is already optimal in regards to the Kozak sequence and also the Kozak sequence is the not the sole factor for initiation.

A possible explanation for the preference for the initiation site being located at position 563 is due a secondary structure located in the vicinity of the AUG codon which may influence initiation. Inputting the last 60 nts of the M2 ORF-1 into Mfold predictive software predicted a secondary structure in which both the AUG codon and the surrounding sequence was encompassed in the stem of a predicted structure. However it was apparent that this structure was in fact an inhibitory factor on the initiation of M2 ORF-2. A single point mutation created to disrupt the predicted secondary structure lead to an increase in translation of ORF-2. As the AUG codon was base-paired in the stem loop, disruption of the structure may allow an initiating ribosome to access the AUG codon without unwinding the structure. This also could suggest that this mechanism is reliant on helicases required to unwind this region.

As there were limitations to the size of the sequence that could be inputted into the Mfold modeling software, *in vitro* transcription reactions were used to identify the presence of large secondary structures through the production of premature termination products. It was apparent that there was no large stable secondary structure present inside the M2 ORF-1 sequence. This data aided in identifying the mechanism used for translation of M2 ORF-2.

Using chimeric reporter plasmids with increasing regions of M2 ORF-1 deleted, it was apparent that the sequences between positions 210 and 528 in the M2 ORF-1 were used by the ribosome to initiate translation of M2 ORF-2 and were essential for successful translation. Sequences upstream of this region did not appear to be required for ORF-2 translation. The sequences between position 210 and 381 was important but not essential for translation of M2 ORF-2, with the removal of this region only decreasing translation and not inhibiting it. However, the region between 382 and 528 was critical for translation with removal inhibiting translation. This suggests that the mechanism used for translation of M2 ORF-2 is likely to use this region for successful translation.

It is likely that even though bRSV does not use coupled translation for the initiation of translation of M2 ORF-2, a non-canonical initiation mechanism is used due to the presence of the upstream M2 ORF-1. There are four mechanism which may be used (PRF, leaky scanning, shunting and IRES). The data acquired from this chapter has ruled out the possibility of at least two of these mechanisms.

It is highly unlikely that the method of initiation for ORF-2 is that of leaky scanning. The ribosome has to bypass 14 alternate start codons, with three of these codons having similar strength Kozak sequences. It is also unlikely that the mechanism of programmed ribosomal frameshifting is used. The lack of a fusion protein present in lysed samples would suggest that PRF is not used. It is also unusual for PRF to be reliant on a start codon as the site of the frameshift usually occurs at a slippery sequence. However evidence from Chapter 4 suggests that the start codon at position 563 is indeed the location for initiation of ORF-2. There is also the requirement for a secondary structure downstream of the frameshift site, this usually begins within 5-6 nts from the slippery sequence (Namy et al., 2006). However with the replacement of

ORF-2 downstream of the ORF-1/ORF-2 overlap region used for these reporter plasmids, any structured element that would be used for PRF is removed and replaced. It is therefore highly unlikely that a similar structure is formed from the new reporter sequence downstream of the ORF-1/ORF-2 overlap region.

The data acquired in this chapter suggests that bRSV does not use a ribosomal shunting mechanism similar to that used by Adenovirus in late mRNA translation. Regions of sequence surrounding the preferred AUG in bRSV M2 ORF-1 that were complementary to the same 3′ 18S RNA sequence used by Adenoviruses (Yueh and Schneider, 2000) for shunting were not critical for translation of M2 ORF-2. Although this suggests that shunting is not used in this style, it may be possible that a differing shunting mechanism may be used.

There is also further evidence that the ribosome does not use a scanning style mechanism for translation of M2 ORF-2. Mutations that incorporated artificial AUG codons into M2 ORF-1 that could be used for initiation of ORF-2, did not lead to successful translation of ORF-1 when the both native initiation codons for ORF-2 were mutated. This suggests that as translation of ORF-2 was completely inhibited, it suggests that mechanism may not use scanning. It also provides further evidence that the location of the AUG codon at position 563 is also a factor for successful translation.

Creating bicistronic reporter constructs separated by spacer sequences showed that there is evidence that an internal initiation mechanism is used. However, identifying the precise internal initiation mechanism is difficult. This type of test is only an initial test used for internal initiation, with further experiments needing to be performed before coming to a conclusion on the type of internal initiation mechanism used such as an IRES. Ribosomal profiling has identified regions of interest within the sequence where internal initiation may take place, although more evidence is needed to conclusively prove these regions of interest are involved in initiation of ORF-2.

Chapter 6
Characterisation of the Mechanism of Coupled Translation Termination/Initiation used for Translation of PVM M2 ORF-2

6.1. Introduction

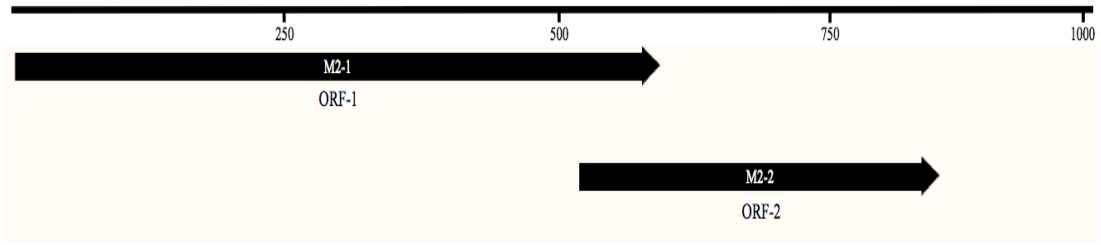
It has been reported that PVM, a member of the *Orthopneumovirus* genus in the *Pneumoviridae* family also uses the mechanism of coupled translation termination/ initiation for the translation of M2 ORF-2 for the M2-2 protein, found on the bicistronic gene transcript of M2, as described for hRSV. For PVM the site of initiation differs significantly to that of hRSV. However, PVM is still reliant on the termination of translation of the first ORF in order for initiation of translation of the second (Gould and Easton, 2007).

As in hRSV, the PVM M2 gene is the ninth gene from the 3′ leader end of the genome and is immediately prior to the gene encoding the L polymerase protein. The M2 gene of PVM is 869 nts in length and similar to hRSV, contains two ORFs that are translated to the M2-1 and M2-2 proteins, respectively. The PVM M2 ORF-1 (535 nts) is significantly larger than that of M2 ORF-2 (297 nts: Fig. 6.1A). The PVM M2 ORF-2 is in reading frame 3, in contrast to the situation for the hRSV M2 ORF-2 that is in reading frame 2. The two ORFs in the PVM M2 mRNA overlap each other by 67 nts, which is significantly longer than seen in the other viruses in the family. In PVM, there are two potential initiation sites for M2 ORF-2 which are separated by 62 nts (Fig. 6.1B). The second start codon is located at the 3′ of the overlap and overlaps the stop codon of M2 ORF-1 by 2 nucleotides.

Gould and Easton, 2007 reported that the initiation of translation of the PVM M2 ORF-2 was not reliant on M2 ORF-2 sequence downstream of the M2 ORF-1/ORF-2 overlap region. It was also suggested that both AUG start sites for translation of M2 ORF-2 in the M2 ORF-1/ORF-2 overlap region are used equally and removal of both start codons reduced levels of translation of the second ORF to negligible levels (Gould and Easton, 2007). For hRSV (Ahmadian et al., 2000), removal of the terminating stop codon of M2 ORF-1 and moving the stop codon downstream from the original site of termination has been used to test for coupled translation termination/initiation. This leads to inhibition of translation of M2 ORF-2 as the mechanism is distance dependent on the site of the stop codon. Gould and Easton, 2007 reported that for PVM, translation of the M2 ORF-2 was inhibited if the stop codon of M2 ORF-2 moved 138 nts downstream (Gould and Easton, 2007). This indicates that PVM uses the mechanism of coupled translation termination/initiation

for initiation of M2 ORF-2, as initiation is reliant on translation and termination of M2 ORF-1 in the vicinity of the M2 ORF-2 initiation site.

A



B

ORF-1/ORF-2 Overlap

AUGCAAUCUGAUCCAAUCUGUCAUCUCCAUCGAGGAGAAGAUAAAUUCUUCUAUGAAAACAGA**AUGA**

Fig. 6.1: Schematic of PVM M2 gene transcript and M2 ORF-1/ORF-2 overlap region
(A) Schematic of the bicistronic PVM M2 gene transcript including the locations of the two ORFs. The line represents the M2 mRNA and the arrows represent the location of the two ORFs. M2 ORF-1 (535nts long) is translated into the M2-1 protein. M2 ORF-2 (297nts long) is translated into the M2-2 protein. (B) Sequence of the M2 ORF-1/ORF-2 overlap region. AUG initiation codons for M2 ORF-2 are underlined. The M2 ORF-1 termination codon is indicated in bold.

It has been shown that in hRSV, a highly structured region 150 nts 5′ of the M2 ORF-1/ORF-2 overlap region and within M2 ORF-1 is essential for successful initiation of translation of M2 ORF-2 (Gould and Easton, 2005). The structured region is thought to aid initiating ribosomes for M2 ORF-2 by pausing elongating ribosomes translating M2 ORF-1 further upstream, allowing ribosomes to reinitiate translation of M2 ORF-2 (P. Gould and A. Easton, personal communication) (Section 1.4.2). Other viruses that use the mechanism of coupled translation termination/initiation such as members of the *Caliciviridae* family, utilize a termination upstream ribosomal binding site (TURBS) sequence upstream of the initiation site for reinitiating ribosomes. The TURBS is thought to bind to the h26 helix in 18S rRNA to position the reinitiating ribosome in an optimal position for reinitiation of translation (Section 1.5.1). For all members of the *Caliciviridae* family that use this mechanism, a short UGGGA sequence is conserved throughout all TURBS sequences and is responsible for binding to the h26 helix in 18S rRNA.

While there is evidence that the PVM M2 mRNA utilises coupled translation to access M2 ORF-2 (Gould and Easton, 2007), there has been no detailed exploration of the mechanism by which the PVM M2-2 protein is translated. In particular there is

no information about the potential presence or possible role of sequences within the M2 ORF-1 on the translational coupling process in PVM.

The aim of this chapter was to further characterise the mechanisms of coupled translation termination/initiation in the PVM M2 mRNA and to search for sequences upstream of the M2 ORF-1/ORF-2 overlap region between M2 ORF-1 and M2 ORF-2 that may be of importance to the process.

6.2. Establishing successful expression of ORF-2 in plasmids used to test for coupled translation in PVM M2 ORF-2

Plasmids used to test for coupled translation in PVM from Gould and Easton, 2007 would be used as templates or as controls for all plasmids used to further investigate this mechanism in PVM in this section. Transfection experiments that were used to report these findings in Gould and Easton, 2007 were repeated in order to reconfirm the vector model for PVM, and that the current plasmids acquired successfully expressed the reporter gene CAT.

All plasmids used as templates or controls from Gould and Easton, 2007 in this study were acquired from Dr Phillip Gould. A schematic of the reporter plasmid pPVMWC which was the template used for all plasmids in this chapter can be seen in Fig. 6.2A. The gene transcript transcribed from the reporter plasmid for PVM contained a 5′-UTR, which was 21 nts in size, of which the final 4 nts originated from the 5′-UTR of the M2 gene transcript sequence in PVM. The plasmid contained two ORFs of which ORF-1 contained the entirety of the PVM M2 ORF-1. For ORF-2, the PVM M2 ORF-2 excluding the overlapping sequence was replaced with the bacterial reporter gene CAT and placed in the same reading frame as the M2 ORF-2. Therefore initiation of translation of ORF-2 via initiation at the AUG codons in the ORF-1/ORF-2 overlapping sequence would lead to the expression of the CAT reporter gene. This was then quantified using an ELISA assay. The plasmid pPVMStop was used to test for coupled translation by removing the ORF-1 (M2 ORF-1) stop codon by mutating UGA to UGG. The plasmids pPVMmut1, pPVMmut2, pPVMmut3, pPVMmut4 and pPVMmut5 removed both start codons (positions 465 and 528 from the start of the M2 gene transcript) for ORF-2 in the ORF-1/ORF-2 overlap region in different combinations. The various point mutations made in each plasmid in the ORF-1/ORF-2 overlap region for the analyses described below can be seen in Fig. 6.2B.

Plasmids were individually transfected into HEp-2 cells as described in Section 2.4.1. The cells were incubated for 48 hours and then lysed and the protein harvested. The level of CAT reporter protein was determined using an ELISA as described in Section 2.4.2. All data were normalized against the values obtained from cells transfected with empty pBluescribe plasmid. This provided a background signal. For

each sample, levels of CAT protein expressed have been visualized as a percentage of CAT protein expression from the plasmid pPVMWC.

Similar levels of CAT reporter gene expression were observed in transfected plasmids as reported in Gould and Easton, 2007 (Fig. 6.2C). Removal of ORF-1 (M2 ORF-1) stop codon in the plasmid pPVMStop reduced CAT protein expression to negligible levels (6.5%), confirming that expression of the PVM M2 ORF-2 requires translation of ORF-1 and therefore uses the process of coupled translation termination/initiation. Removal of both ORF-2 start codons in the ORF-1/ORF-2 overlap region in plasmids pPVMmut3 (10.8%) and pPVMmut5 (9%) reduced levels of CAT protein expression significantly as described in Gould and Easton, 2007. This indicated that both AUG codons in the overlapping region are the preferred initiation sites for translation of M2 ORF-2. However, there was a visibly lower level of expression in the plasmid pPVMmut1 compared to the results in Gould and Easton, 2007 (34% decrease). This may suggest that there is a larger preference for initiation at the AUG codon located at the 5´ end of the M2 ORF-1/ORF-2 overlap than initially reported. It also possible that the quality of the plasmid may have caused a lower level of ORF-2 expression than observed in Gould and Easton, 2007. A repeat with newly purified pPVMmut1 plasmid would need to performed to ascertain if this difference was due to preference or plasmid purity

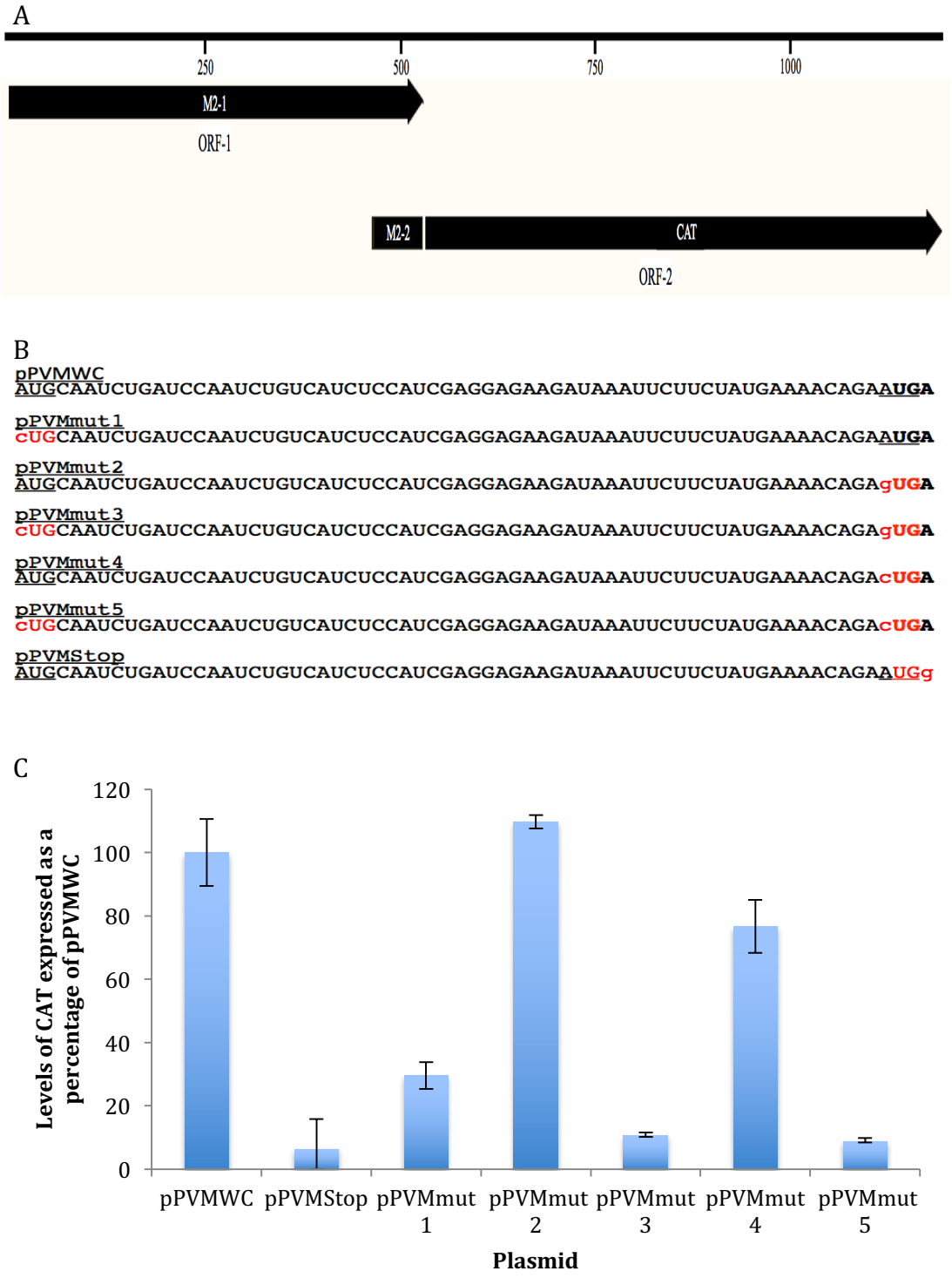


Fig. 6.2: Schematic of plasmids and ELISA results measuring CAT expression for constructs investigating coupled translation of M2 ORF-2 in PVM

(A) Schematic of the two ORFs located on the mRNA transcribed from reporter gene plasmid pPVMWC used to measure the translation of ORF-2 (M2 ORF-2) in the PVM M2 gene transcript. (B) Sequence showing the mutations made to the ORF-1/ORF-2 overlap region for each of the PVM reporter plasmids. Potential AUG initiation codons for the M2 ORF-2 are underlined and mutated codons are coloured red. The M2 ORF-1 translation termination codon is indicated in bold. (C) Levels of CAT protein expressed by each reporter plasmid. Values were normalised against the empty vector pBluescribe and are expressed as a percentage of the expression level of the plasmid pPVMWC. Error bars are standard deviation.

6.3.1. Identification of sequences critical for initiation of PVM M2 ORF-2 translation upstream of the M2 ORF1/ORF2 overlap region

Following confirmation of the results of Gould and Easton, 2007, the mechanism of coupled translation in PVM was further explored. Using the same approach reported for analysis of the hRSV M2 mRNA (Gould and Easton, 2005), a series of reporter plasmids were constructed using the same reporter system to measure ORF-2 expression as used in Section 6.2. But in this case, the gene encoding eGFP lacking the termination codon was inserted in-frame at the 5′ end of the M2 ORF-1 in which the M2 ORF-1 initiation codon had been deleted. In this way the 5′ proximal ORF was extended by 722 nts and expression of the virus ORF-1 sequence required translation initiation from the long eGFP gene. In the plasmid series, increasing proportions of PVM M2 ORF-1 sequence were deleted (shown schematically in Fig. 6.3A). Each deletion began at the start of the M2 ORF-1 coding sequence with the 3′ break point being progressively closer to the ORF-1/ORF-2 overlap region. The presence of the entire eGFP coding sequence ensured that the initiating ribosomes were not close to the ORF-1/ORF-2 overlap region, resembling the native PVM M2 mRNA organisation. As before, replacement of ORF-2 with the CAT gene allowed expression to be measured. In total six plasmids were created (Fig. 6.3B). For plasmids pPVM6, pPVM84, pPVM204, pPVM370 and pPVM424 had 6, 84, 204, 370 and 424 nts removed from ORF-1 for each plasmid respectively.

The plasmid pPVM6Stop had 6nts removed from the 5′ end of the M2 gene transcript sequence in the same format as pPVM6. The plasmid however also had the terminating ORF-1 stop codon mutated (UGA to UGG). This was to ensure that coupled translation initiation/termination was still the mechanism of initiation for translation of ORF-2.

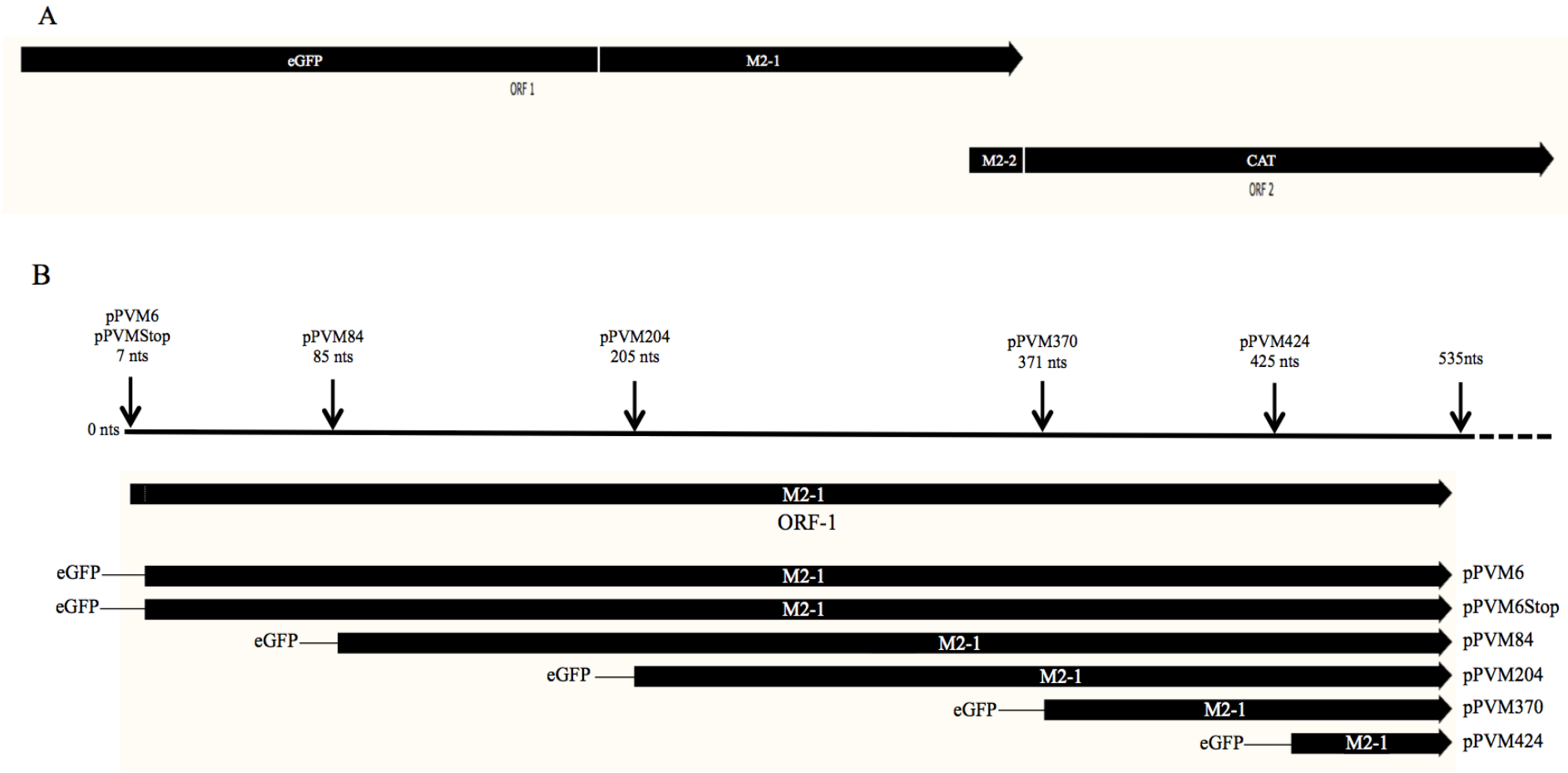


Fig. 6.3: Schematic of chimeric reporter gene plasmids
(A) Schematic representation of chimeric reporter gene plasmids used to investigate critical sequences in ORF-1 (M2 ORF-1) for the initiation of translation of ORF-2 (M2 ORF-2). Schematic identifies both the ORF-1 (eGFP/M2 ORF-1) and the ORF-2 (M2 ORF-2/CAT). (B) Schematic displaying the variable lengths of the M2 ORF-1 coding sequence in ORF-1 for chimeric plasmids pPVM6, pPVM6Stop, pPVM84, pPVM204, pPVM370 and pPVM424 in comparison to the full length M2 ORF-1 in the M2 gene transcript of PVM. The black line represents the M2 gene transcript of PVM. Arrows on this line are used to identify the beginning of the M2 ORF-1 coding sequence for each plasmid in relation to the M2 gene transcript.

Construction of plasmids pPVM6, pPVM6Stop, pPVM84, pPVM204, pPVM370 and pPVM424

For the plasmids pPVM6, pPVM84, pPVM204, pPVM370 and pPVM424, the method for construction of the plasmids was identical except for the use of the forward primer for the PCR of the truncated M2 ORF-1/M2 ORF-2/CAT (PVM fragment). The method for cloning and sequencing was identical to the protocol used in Section 5.6.1.

For these plasmids, the truncated M2 ORF-1/M2 ORF-2/CAT (PVM fragment) amplicon was amplified via PCR as described in Section 2.3.1. The template used was pPVMWC. For each PCR reaction the relevant forward primer used can be seen in Table 6.1. The reverse primer CAT Hind III, was used for every PCR reaction. For all plasmids, an eGFP amplicon was generated via PCR from the plasmid peGFP using the forward primer T7-91 and the reverse primer eGFP Sph I to generate an amplicon of 880 bp in size. The predicted lengths of the resulting PVM inserts can be seen in Table 6.1. The size of fragments generated from digestion for each of the respective cloned plasmids with the newly ligated inserts can be seen in Table 6.1.

For the plasmid pPVM6Stop, the amplicon of the truncated M2 ORF-1/M2 ORF-2/CAT fragment was generated using the primers PVM6 F and CAT Hind III from the plasmid pPVMStop (plasmid used in Section 6.2, where the terminating ORF-1 stop codon was removed) to generate an amplicon of 1202 bp. An eGFP amplicon was generated via PCR from the plasmid peGFP using the forward primer T7-91 and the reverse primer eGFP Sph I to generate an amplicon of 880 bp in size as described in Section 2.3.1. The method for cloning and sequencing was identical to the protocol used in Section 5.6.1. A correctly digested fragment of 1948 bp was observed when the insert was digested from the newly ligated plasmid to confirm successful ligation of the insert.

Plasmid	Forward primer	Reverse Primer	Size of fragment generated (bp)	Size of correct digested insert from plasmid (bp)
pPVM6	PVM6 F	CAT Hind III	1202	1948
pPVM84	PVM84 F	CAT Hind III	1124	1868
pPVM204	PVM204 F	CAT Hind III	1015	1750
pPVM370	PVM370 F	CAT Hind III	839	1584
pPVM424	PVM424 F	CAT Hind III	785	1530

Table 6.1: Table of primers used
Table of primers to used to generate (PVM fragments) amplicons used for each plasmid and size of fragments generated for PCR and digest.

6.3.2 Investigation of the potential role of upstream sequences in expression of the PVM M2 ORF-2

The chimeric plasmids described in Section 6.3.1 were individually transfected into HEp-2 cells as described in Section 2.4.1. Control cells were transfected with either pPVMWC or the empty vector pBluescribe. The levels of CAT protein expressed by each transfected plasmid were quantified using a CAT ELISA assay as described in Section 2.4.2. All samples were normalized against the results from cells transfected with empty pBluescribe plasmid to control for any background signal generated in the ELISA. For each sample, levels of CAT protein were expressed as a percentage of those obtained from the positive control plasmid pPVMWC.

As seen in Fig. 6.4, the addition of the eGFP coding sequence to the 5′ end of the M2 ORF-1 lacking the first 6 nucleotides of the M2 gene transcript sequence in the plasmid pPVM6 significantly reduced levels of CAT protein expression from ORF-2 to 29% of that of the control ($p=0.00022$). A similar reduction of 30% in CAT protein expression ($p=0.0052$) was also seen with plasmid pPVM84 lacking the first 84 nucleotides of the M2 gene transcript sequence. Such a significant reduction in reporter gene expression in a plasmid in which only 6 nucleotides were lacking from the 5′ end of the M2 gene transcript was unexpected and this was not seen in equivalent plasmids generated by the same method used to investigate hRSV (Gould and Easton, 2005). As the expression levels of the two PVM plasmids were

extremely similar (p=0.865), it is likely that the lower level of expression reflects the base level of expression for the chimeric plasmids containing the eGPF ORF fused to the PVM M2 ORF-1.

The plasmid pPVM6Stop was generated to confirm that expression of ORF-2 M2 ORF-2/CAT ORF was achieved using coupled translation termination/initiation. In pPVM6Stop the ORF-1 stop codon was mutated from UGA to UGG, extending the ORF to the next in-frame stop codon located 135nts away. As seen in Fig. 6.4, the level of CAT protein expressed by the plasmid pPVM6Stop was 4.6% of that of the control plasmid pPVMWC. This was similar to the level of expression observed from plasmid pPVMStop (Section 6.2, 6.5%) which contained the entire M2 ORF-1, no eGFP sequence and had the M2 ORF-1 stop codon mutated from UGA to UGG (see Section 6.2). This 6.3-fold reduction in expression arising from extending the first ORF indicates that initiation of translation of the M2 ORF-2 in the plasmids containing the eGFP ORF fused to the M2 ORF-1, is achieved by coupled translation termination/initiation.

Removal of 204nts from the 5′ end of the M2 gene transcript in the plasmid pPVM204 reduced levels of CAT protein expression to 18% of that seen with pPVMWC, a 39% reduction in expression in comparison to pPVM6. The expression values obtained from pPVM6 and pPVM204 were not statistically significant (p=0.21) indicating that no sequences in this region are required for expression of ORF-2.

Removal of 370 nts from the 5′ end of M2 mRNA sequence in the plasmid pPVM370 reduced levels of CAT protein expression to 3.3% of that of the control. This was similar to expression from plasmid pPVM6Stop and there was no significant difference between the two plasmids (p=0.71). Thus, removal of the sequence between residues 205 and 370 of the PVM M2 mRNA reduced ORF-2 translation to basal levels indicating that this 165 nts region contains a sequence critical for coupled translation termination/initiation to occur. This conclusion was further supported by analysis of expression from the plasmid pPVM424 which lacked the first 424nts of the M2 mRNA and where the level of CAT protein expression was reduced to 5.9% of control levels. This again was similar to levels of

expression from pPVM6Stop (p=0.81), indicating that there was no translation of the second ORF.

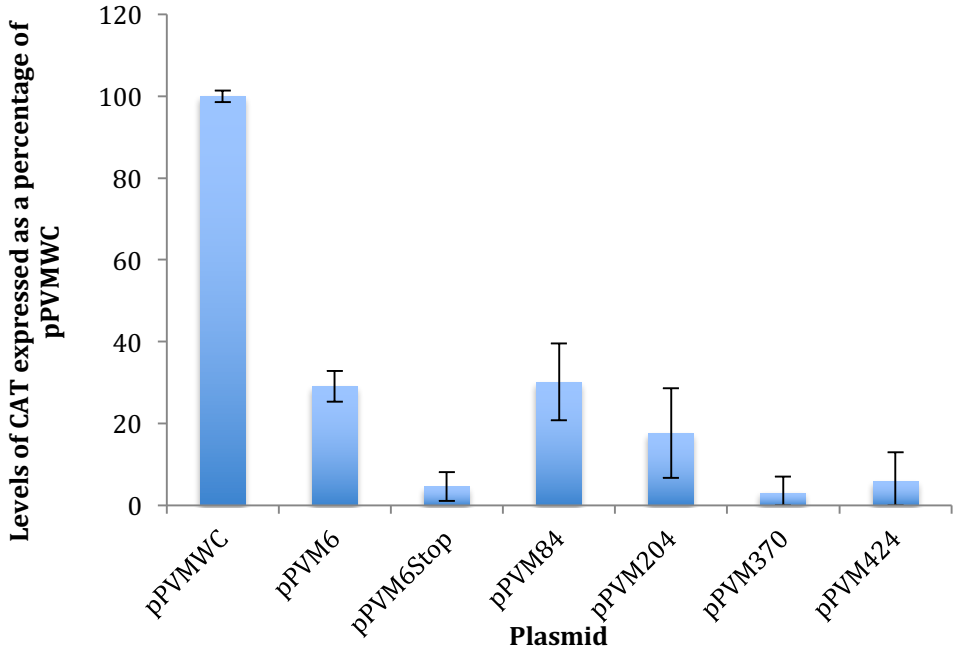


Fig. 6.4: ELISA results measuring CAT expression for chimeric plasmids investigating critical sequences for the mechanism of coupled translation of M2-2 in PVM
The levels of expression of CAT protein from plasmids transfected into HEp-2 cells. Mutations made to these plasmids are described in Section 6.3.1. Levels of CAT expression were normalised against the empty vector pBluescribe and are expressed as a percentage of the control plasmid pPVMWC. Error bars are standard deviation.

6.4.1 Identification of critical sequence for the initiation of PVM M2 ORF-2 between 204 nts and 370 nts from the 5′ end of the M2 gene transcript

Analysis of the region of the PVM M2 mRNA (residues 205-370) that was shown to be required for the mechanism of coupled translation of ORF-2 identified a sequence identical to the conserved TURBS sequence (UGGGA) at residues 222 to 226. In the *Caliciviridae* family, the TURBS sequence is located 70 nts upstream of the start codon of the second ORF (Luttermann and Meyers, 2009; McCormick et al., 2008; Meyers, 2003, 2007; Napthine et al., 2009), while in the PVM M2 mRNA this sequence is located 242 nts from the first start codon of the M2 ORF-2 in PVM. Further analysis was carried out to further define the sequence required for coupled translation in PVM and whether the TURBS sequence was also playing a role.

A second series of plasmids was created to narrow down the boundaries of the region required for coupled translation. Two chimeric plasmids in the same format to those used in Section 6.3.3 were created to clarify the upstream 5′ boundary of this region to within 80 nucleotides and to analyse the potential role of the TURBS-like sequence identified in the PVM M2 ORF-1. eGFP was ligated to the 5′ end of the M2 gene transcript as described in Section 6.3.1.

The names of these two plasmids indicate the extent of sequence deletion as described in Section 6.3.1. As seen in Fig. 6.5, 228 nts and 306 nts were removed from the M2 gene transcript sequence and eGFP coding sequence was ligated to the 5′ end of the newly shortened M2-1 coding sequence for the plasmids pPVM228 and pPVM306. The putative TURBS sequence (residues 222 to 226) was located in the M2 ORF-1 coding sequence in the plasmid pPVM228 but not in the plasmid pPVM306 (Fig. 6.5 black box). A decrease in expression of the CAT reporter gene for the pPVM228 may indicate that putative TURBS sequence may play a role in the initiation of translation of M2-2.

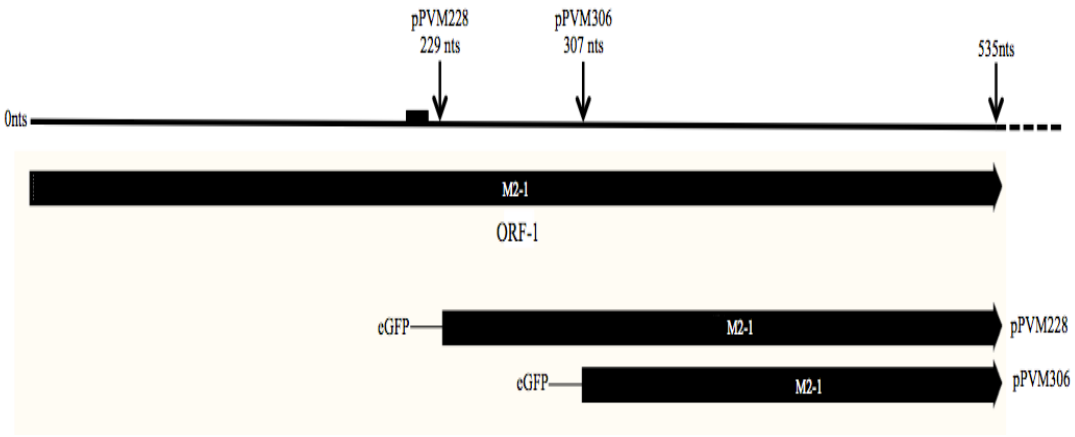


Fig. 6.5: Schematic of chimeric reporter gene plasmids pPVM228 and pPVM306
Schematic displaying the variable lengths of the M2 ORF-1 coding sequence in ORF-1 for chimeric plasmids pPVM228 and pPVM306 in comparison to the full length M2 ORF-1 in the M2 gene transcript of PVM. The black line represents the M2 gene transcript of PVM. Arrows on this line are used to identify the beginning of the M2 ORF-1 coding sequence for each plasmid in relation to the M2 gene transcript. The black box on this line represents the location of the putative TURBS sequence.

Construction of pPVM228 and pPVM306

For all plasmids, truncated versions of M2-1/M2-2/CAT (PVM fragment) amplicon were amplified using PCR as described in Section 2.3.1. The template was the

plasmid pPVM204 with the forward primer PVM228 and PVM306 used for the plasmids pPVM228 and pPVM306 respectively, and the reverse primer CAT Hind III was used for both reactions. This generated fragments of 981bps and 913bps for plasmids pPVM228 and pPVM306 respectively. For both plasmids, an eGFP amplicon was generated via PCR from the plasmid peGFP using the forward primer T7-91 and the reverse primer eGFP Sph I to generate an amplicon of 880bp in size as described in Section 2.3.1. The method for cloning and sequencing was identical to the protocol used in Section 5.6.1.

6.4.2. Further characterisation into the role of critical sequence for M2 ORF-2 translation located between positions 204 and 370

The plasmids pPVM228 and pPVM306 were individually transfected into HEp-2 cells. Control cells were individually transfected into HEp-2 cells with pPVMWC, pPVM6, pPVM84, pPVM370, pPVM424 and pBluescribe as described in 2.4.1. The levels of CAT protein expressed were quantified using a CAT ELISA assay as described in Section 2.4.2. All samples were normalized against cells transfected with the empty vector pBluescribe plasmid. For each sample the levels of CAT protein were expressed as a percentage signal obtained from the positive control plasmid pPVMWC. pPVM6 and pPVM84 were used as controls for baseline expression of these new chimeric plasmids. pPVM370 and pPVM424 were used as controls where negligible levels of CAT were expressed, indicating a inhibition of translation of the mechanism of coupled translation termination/initiation.

As seen in Fig. 6.6, removal of 228nts from the 5′ end of the M2 gene transcript in plasmid pPVM228 reduced levels of reporter expression to 46.7% of that from the control plasmid pPVMWC. This was not significantly different ($p=0.29$) to that seen with pPVM6, indicating that pPVM228 was able to direct synthesis from the ORF-2 normally and that the nucleotide sequence between residues 204 and 228 did not play any role in expression of ORF-2. The region between residues 204 to 228 includes the putative TURBS sequence and these data therefore confirmed that this did not play a role in coupled translation of ORF-2 the PVM M2 mRNA.

Removal of the first 306nts of the M2 gene transcript in plasmid pPVM306 reduced reporter expression levels to 3.7% of that seen with the plasmid pPVMWC. This was

similar to the expression seen with both pPVM370 and pPVM424, with no significant difference between either the control plasmid pPVM370 (p=0.81) and for the control plasmid pPVM424 (p=0.71). Taken together these data indicate that the sequence lying between 229 nts and 306 nts from the 5′ end of the M2 gene transcript plays a critical role in regulating expression of M2 ORF-2 that lies 226 nts downstream of this region. This is very similar to the situation with hRSV, in which a highly structured region lying approximately 150nts upstream of the 5′ proximal ORF-2 initiation codon is essential and plays a key role in the coupled translation termination/initiation process.

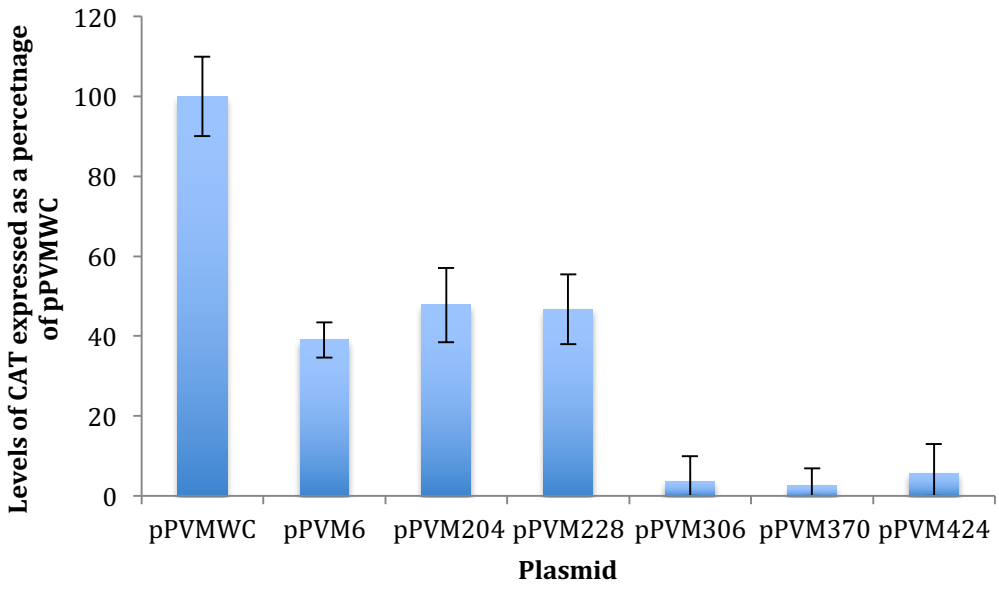


Fig. 6.6: ELISA results measuring CAT protein expression from transfection of plasmids into HEp-2 cells
Levels of expression of CAT from chimeric plasmids transfected into HEp-2 cells. Mutations made to these plasmids are described in Section 6.4.1. Levels of CAT expression have been normalised against the empty vector pBluescribe and are as a percentage of the plasmid pPVMWC. Error bars are standard deviation.

6.5. Identification of structured regions in the PVM M2 gene transcript

The data described above indicates that sequences upstream of the M2 ORF-1/ORF-2 overlap region in PVM are required for initiation of translation of ORF-2. Unlike in hRSV (P. Gould and A. Easton, personal communication), it is not known whether this region constitutes a highly structured region formed in the M2 gene transcript similar to the structure observed in hRSV (Gould and Easton, 2005). In hRSV, this structured region is thought to slow elongating ribosomes, allowing terminating ribosomes downstream to reinitiate translation for the second ORF.

A feature of the highly structured region in the hRSV M2 mRNA reported in Gould and Easton, 2005, was that in *in vitro* transcribed RNA, significant levels of premature termination was seen. The site of the premature termination coincided with the highly structured region essential for translation initiation of the hRSV ORF-2.

In order to search for a potential structured region in the PVM M2 mRNA, an *in vitro* transcription reaction was performed using amplicons generated by PCR from the plasmid pPVMWC to produce RNA (Section 2.7). Amplicons generated by PCR from the hRSV plasmid pWildCAT (Appendix B4) was used in an *in vitro* transcription reaction to transcribe mRNA as a positive control (Gould and Easton, 2005). The products of both reactions were separated using denaturing PAGE (Section 2.7).

The primer pair T7/CATInt was used to generate amplicons for both plasmids (Section 2.3.1). Both amplicons contained the sequence for a T7 RNA polymerase promoter, the entirety of M2 ORF-1, the M2 ORF-1/ORF-2 overlap region and a region of the CAT coding sequence **(**Fig. 6.7A). CATInt was an internal CAT primer used to reduce the size of the DNA, in order to increase the quantity of RNA transcribed. Due to the difference in the size of both M2 ORF-1s, the amplicon generated from the hRSV plasmid pWildCAT was larger than the amplicon generated from the PVM plasmid pPVMWC.

The products of the *in vitro* transcription reactions are shown in Fig. 6.7B. As seen in lane 2, *in vitro* transcription of the amplicon derived from the hRSV pWildCAT

plasmid produced three products. The largest band of 1000 nt represents the full length RNA transcribed by the T7 polymerase. The strong band found at around 400 nts represents premature termination within ORF-1 (M2 ORF-1) reported by Gould and Easton, 2005. The band of approximately 800 nt in size results from premature termination within the CAT ORF.

As seen in Fig. 6.7B lane 3, *in vitro* transcription of the amplicon derived from the PVM pPVMWC plasmid generated two products. The band at 900 nts, represents the full length RNA product. A faint band of approximately 300 nts was produced which indicates the presence of premature termination during transcription. The premature termination product from pPVMWC suggests that a region in the vicinity of 273 nts from the start of the M2 gene transcript, prevents extension of some of the transcription product. This is likely to be a structured region analogous to that found in the hRSV M2 mRNA (Gould and Easton, 2005) and is consistent with the presence of an essential region lying between residues 229 and 306 described in Section 6.4.3.

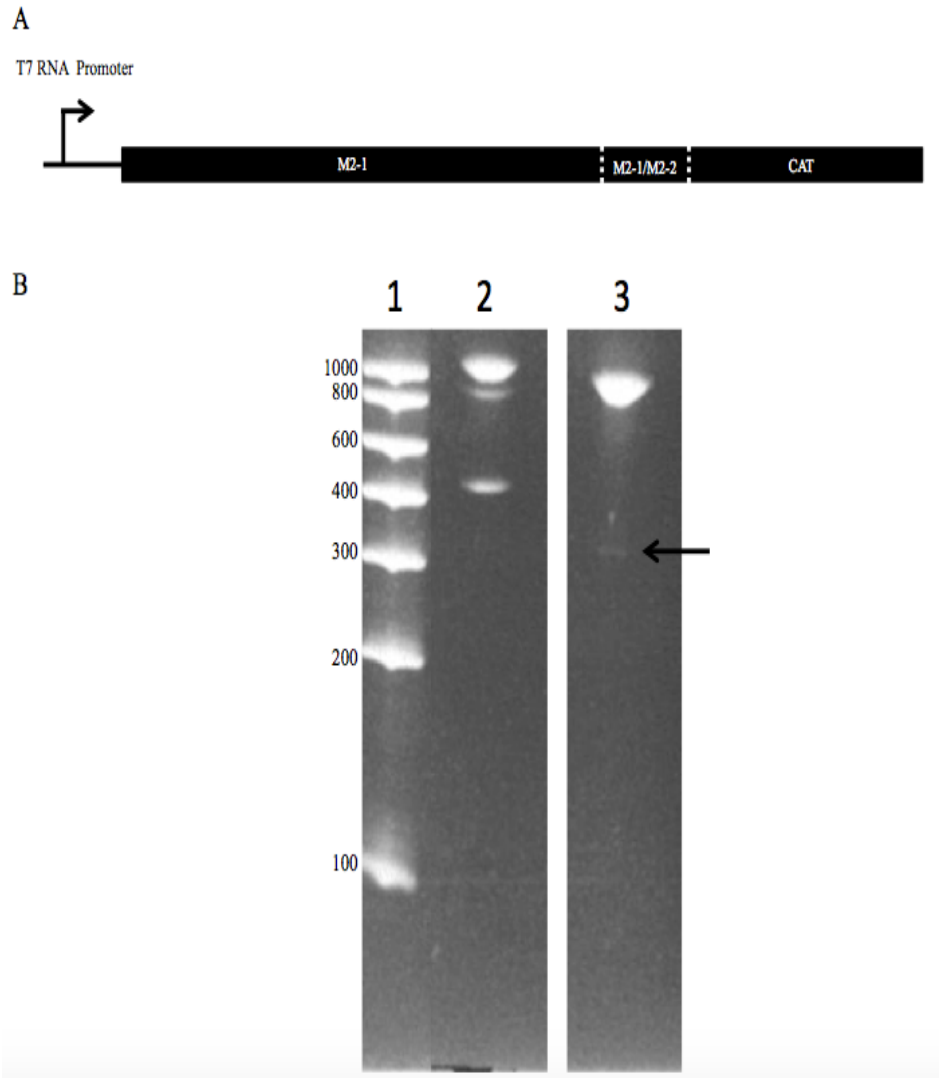


Fig. 6.7: PAGE of products of *in vitro* transcription of plasmids pWildCAT and pPVMWC

(A) Schematic of amplicons generated from the hRSV plasmid pWildCAT and the PVM plasmid pPVMWC. Both amplicons include the T7 RNA promoter, M2 ORF-1, M2 ORF-1/ORF-2 overlap region and a partial CAT coding sequence. (B) Products of *in vitro* transcription were separated on a PAGE gel. Lane 1: Geneflow 100bp ladder. Lane 2: Transcription product of amplicon generated from the hRSV plasmid pWildCAT. Lane 3: Transcription product of the amplicon generated from the PVM plasmid pPVMWC. The sizes of the markers are shown. Black arrow indicates site of faint band in lane 3. All lanes orginated from the same gel, white gaps indicate lanes that were not adjacent to each other.

6.6.1. Predicting structured regions in PVM M2 ORF-1 using Mfold modelling software

From the data presented in Sections 6.4 and 6.5, an upstream sequence was identified that is critical for the initiation of translation for the second ORF in the PVM M2 gene transcript. Analysis of *in vitro* transcription products also suggests that this region contains some degree of secondary structure. A preliminary analysis was carried out using the RNA folding software Mfold to predict potential structures within the RNA. This software predicts the most likely stable secondary structure within a sequence but is limited in that it bases its predictions on the maximum possible base pairing within the target sequence to create the most stable structure. For this reason, several different sized sequences were analysed to try to assess the most likely potential structure. Studies with eGFP chimeric plasmids indicated that no critical sequences required for expression of ORF-2 were located before residue 229 from the start of the M2 gene transcript and this was taken as the 5´ position of the sequences inputted. Three sets of sequences were inputted in the Mfold software each with decreasing sizes of sequence inputted from residue 229. Settings used for the software are described in Section 2.9. Sequence 1 (Fig. 6.8) extended from residue 229 to 535 of the M2 gene transcript representing the essential sequence determined above. Sequence 2 (Fig. 6.9) extended from residue 229 to 422 of the M2 gene transcript and sequence 3 (Fig. 6.10) extended from residue 229 to 322 of the M2 gene transcript.

For Fig. 6.8 (residues 229-535) and 6.9 (residues 229-422), the software predicted two step loop structures located between residues 231-253 and 304-318 surrounding the location identified as containing a putative highly structured region (Section 6.5). The majority of the sequence within this location was base paired with sequence at a significant distance from this location. It is possible that this could be an accurate structure, however it is more likely that this was due to the specific parameters used by Mfold software, which attempt to create the stable secondary structure possible with the sequence provided.

In Fig. 6.10 (residues 229-322), reducing the input sequence from residue 229-322, predicted three stem loops structures at residues 231-253, 254-301 and 303-313. The stem loop structure located between residues 254-301 was in location of the putative

secondary structure identified in Section 6.5. This is also in the area that was identified as being critical for initiation of translation of the M2 ORF-2 from the plasmids pPVM228 and pPVM306 (Section 6.4).

A recent analysis using a combination of Mfold informed by biochemical studies has generated a 3 dimensional prediction of the structured region within the hRSV M2 mRNA (Fig. 6.11) (P. Gould and A. Easton, personal communication). Comparison of this with the data presented here for the PVM M2 mRNA demonstrates some similarities. In particular, stems I, III and IV in the hRSV structure are similar to the three stem loops generated by Mfold in Fig. 6.10.

Two predicted structures located between residues 231-253 and 254-301 also contain loops that are A-rich, one found on the loop located between residues 240-245 and the second loop located between residues 275-282. Recent data has shown that A-rich regions in the structured region of the hRSV M2 mRNA are essential for successful initiation of ORF-2 translation and are thought to allow the essential cellular protein DDX3X to bind (P. Gould and A. Easton, personal communication).

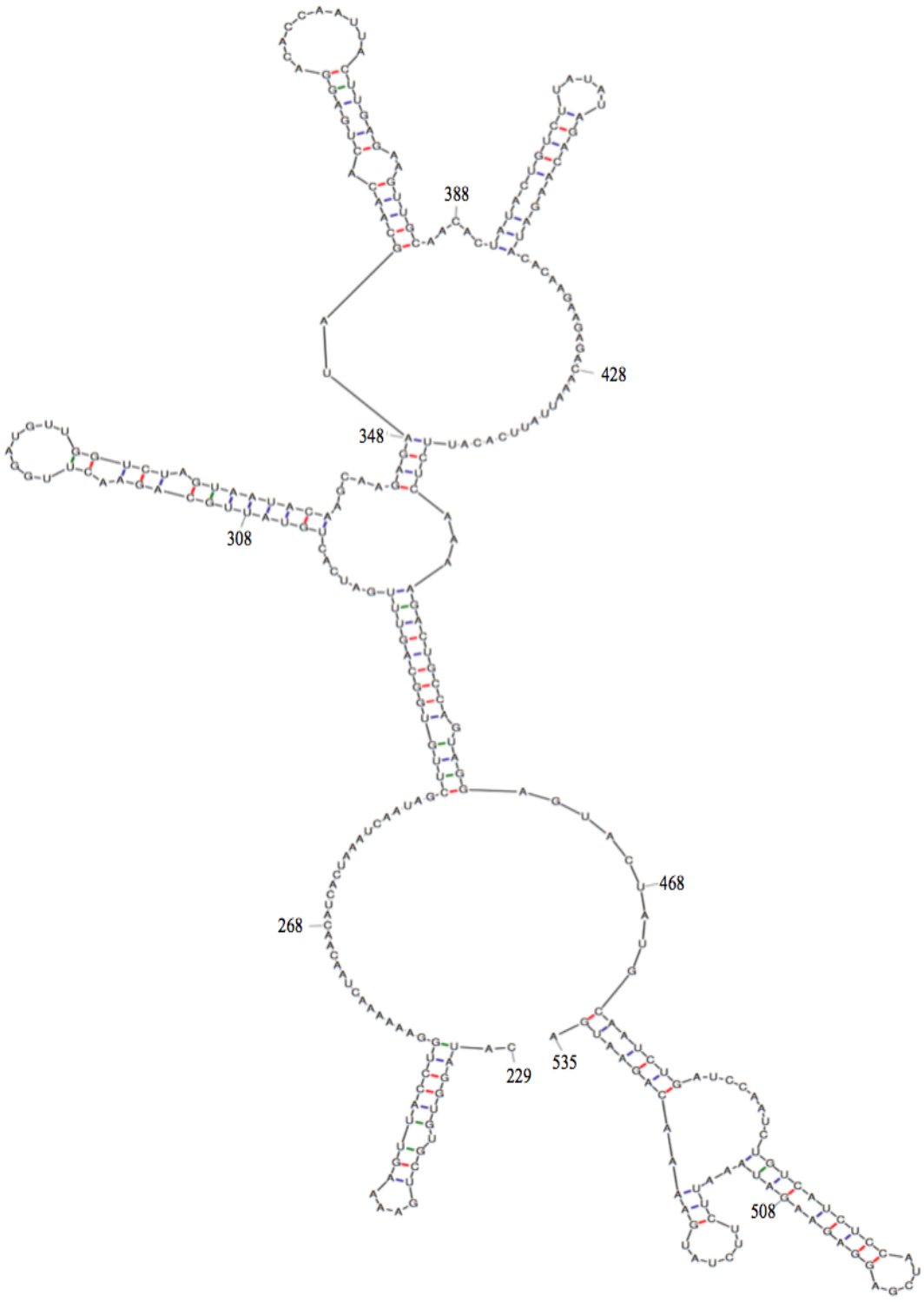


Fig. 6.8: Predicted mRNA structure generated by Mfold from position 228 to 535
Predicted mRNA structure generated by Mfold. Sequence inputted from position 228 to 535 from the beginning of the M2 gene transcript. ΔG=-56.63

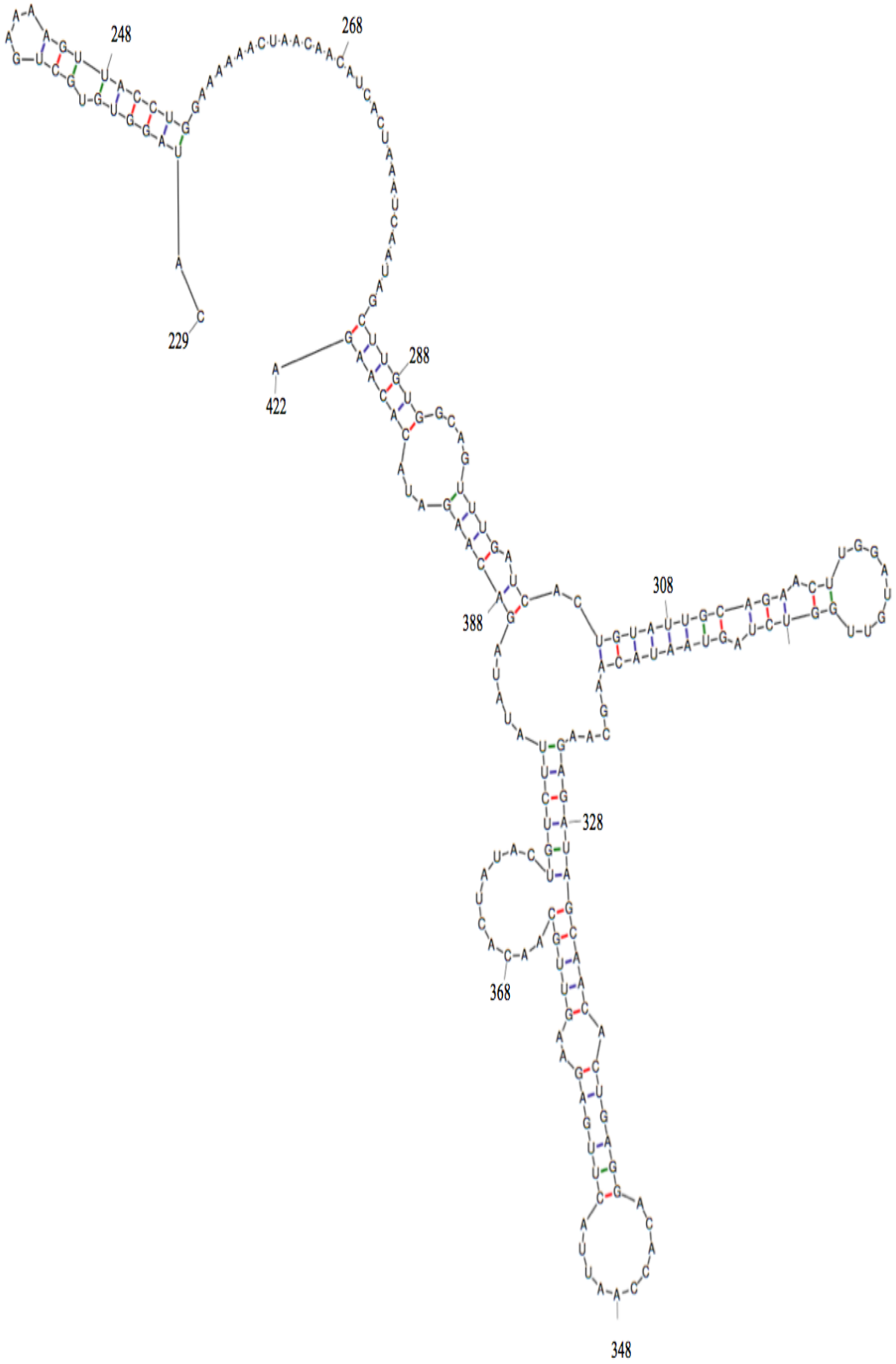


Fig. 6.9: Predicted mRNA structure generated by Mfold from position 228 to 422
Predicted PVM M2 mRNA structure generated by Mfold. The sequence analysed contained nucleotides 228 to 422 from the M2 gene transcript. The predicted ΔG=-35.74.

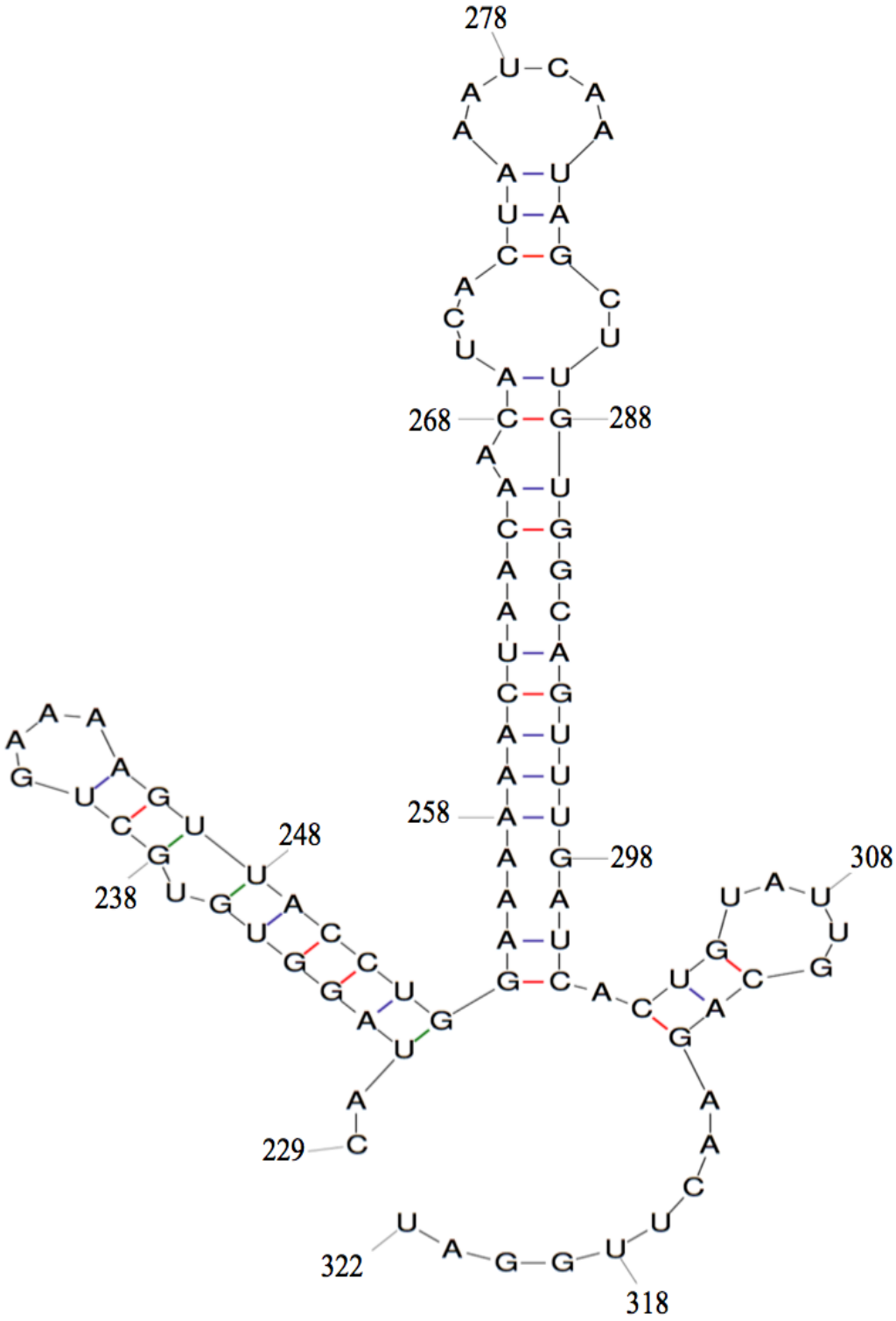


Fig. 6.10: Predicted mRNA structure generated by Mfold from position 228 to 322
Predicted mRNA structure generated by Mfold. Sequence inputted from position 228 to 322 from the beginning of the plasmid M2 gene transcript. ΔG=-17.50

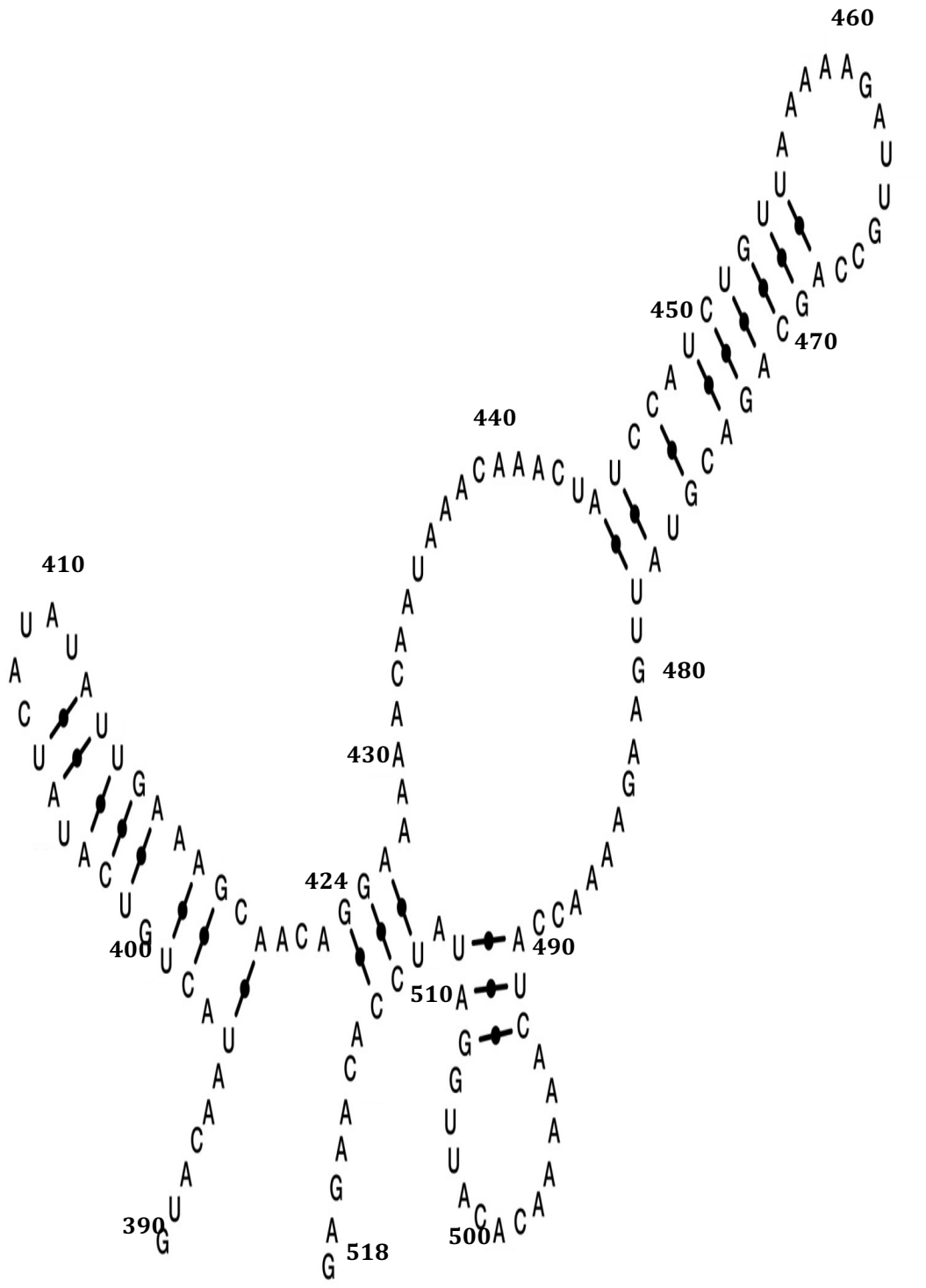


Fig. 6.11: mRNA secondary structure found in hRSV M2 gene transcript
mRNA secondary structure identified in hRSV M2 ORF-1 through SHAPE mapping (P. Gould and A. Easton, personal communication) Structure is critical for the initiation of translation of M2 ORF-2. Stem I (residues 395-421), Stem II (residues 425-511), Stem III (residues 445-478) and Stem IV (490-507)

6.6.2. Investigation into the involvement of a predicted secondary structure in translation of the PVM M2 ORF-2

The structural predictive software Mfold predicted the presence of a stem loop structure from position 254-301 in the PVM M2 gene transcript (Fig. 6.10). As well as the predicted structure being in the location of a putative secondary structure identified in Section 6.5, this was also located in a region that was identified as being critical to the initiation of translation of the M2 ORF-2 in PVM (Section 6.4).

Two plasmids were designed to disrupt this predicted stem loop structure to investigate if this affected translation of the PVM M2 ORF-2. Both plasmids were in the same format as the plasmid pPVMWC (Fig. 6.12A) and the predicted secondary structure was disrupted by mutating three G residues, located at nucleotide positions 284, 288 and at 298, that base paired in the stem structure. Using Mfold, it was predicted that these altered residues when mutated, would disrupt the stem loop structure, as shown in Fig. 6.12B. Point mutations were incorporated into the ORF-1 sequence by mutating these three G residues to C for the plasmid pPVMStructuralC (Fig. 6.13A) or U for plasmid pPVMStructuralU (Fig. 6.13.B). A decrease in CAT protein expression would suggest that the predicted secondary structure identified is likely to be involved in the initiation of translation of ORF-2.

A third plasmid (pPVMStructuralG/C) in the same format as pPVMWC (Fig. 6.12A), was also created to confirm the presence of the predicted structure generated by Mfold. As seen in Fig. 6.13C, the G/C base pairs found at these three sites were switched. If this predicted secondary structure was involved in the initiation of translation of ORF-2, then CAT expression would not be altered by this new set of point mutations as the structure would be retained.

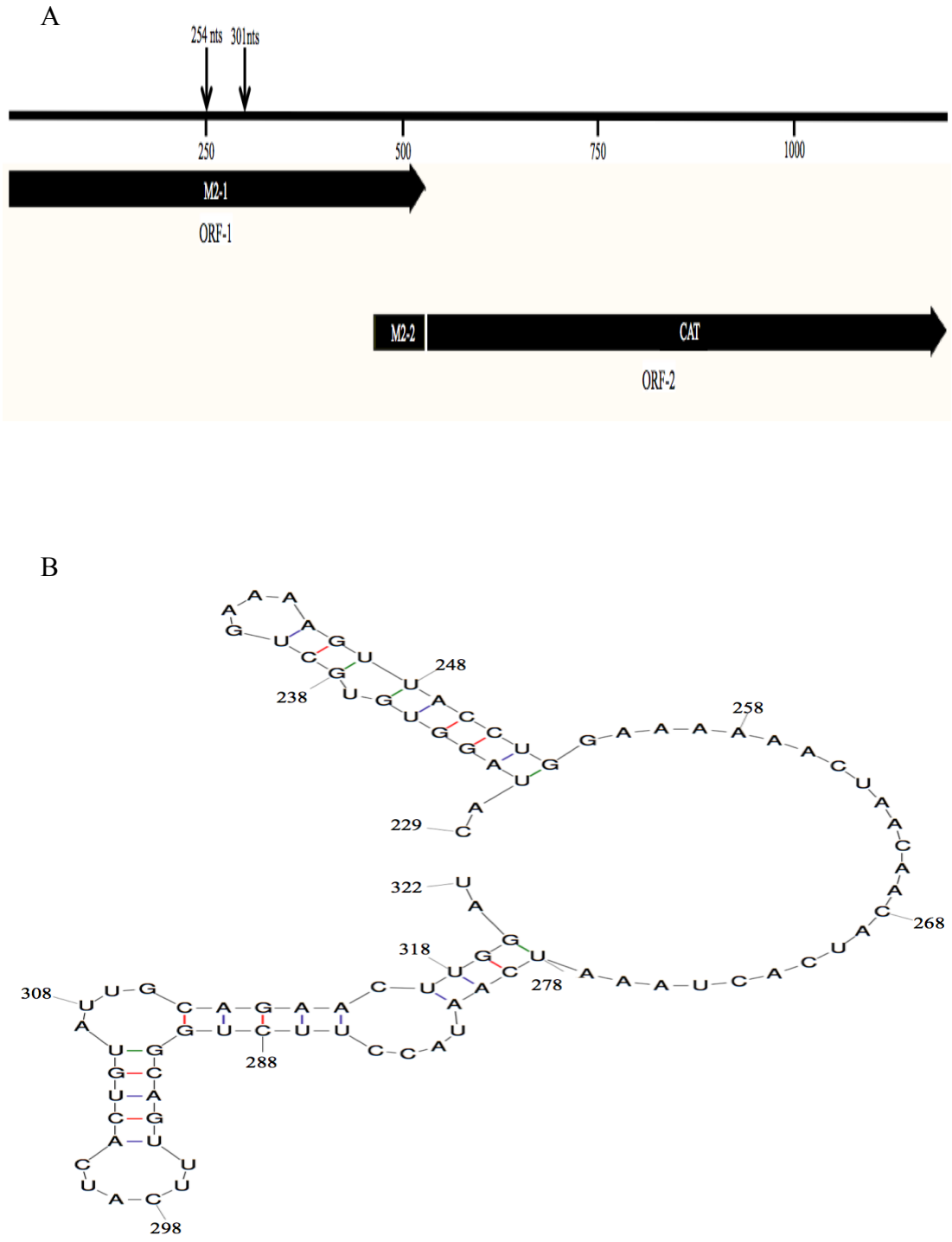


Fig. 6.12: Predicted structure of mutated PVM M2 mRNA structure
(A) Schematic representing the reporter plasmids pPVMStructuralC, pPVMStructuralU and pPVMStructuralG/C. Arrows represent the location for the beginning and the end of the predicted stem loop structure located between residues 254-301 on the PVM M2 gene transcript (B) Predicted mutated mRNA structure generated by Mfold. The sequence from position 229 to 322 of the PVM M2 mRNA containing point mutations (G to C) at nucleotide positions 284, 288 and 298 are present. The predicted ΔG=-11.35.

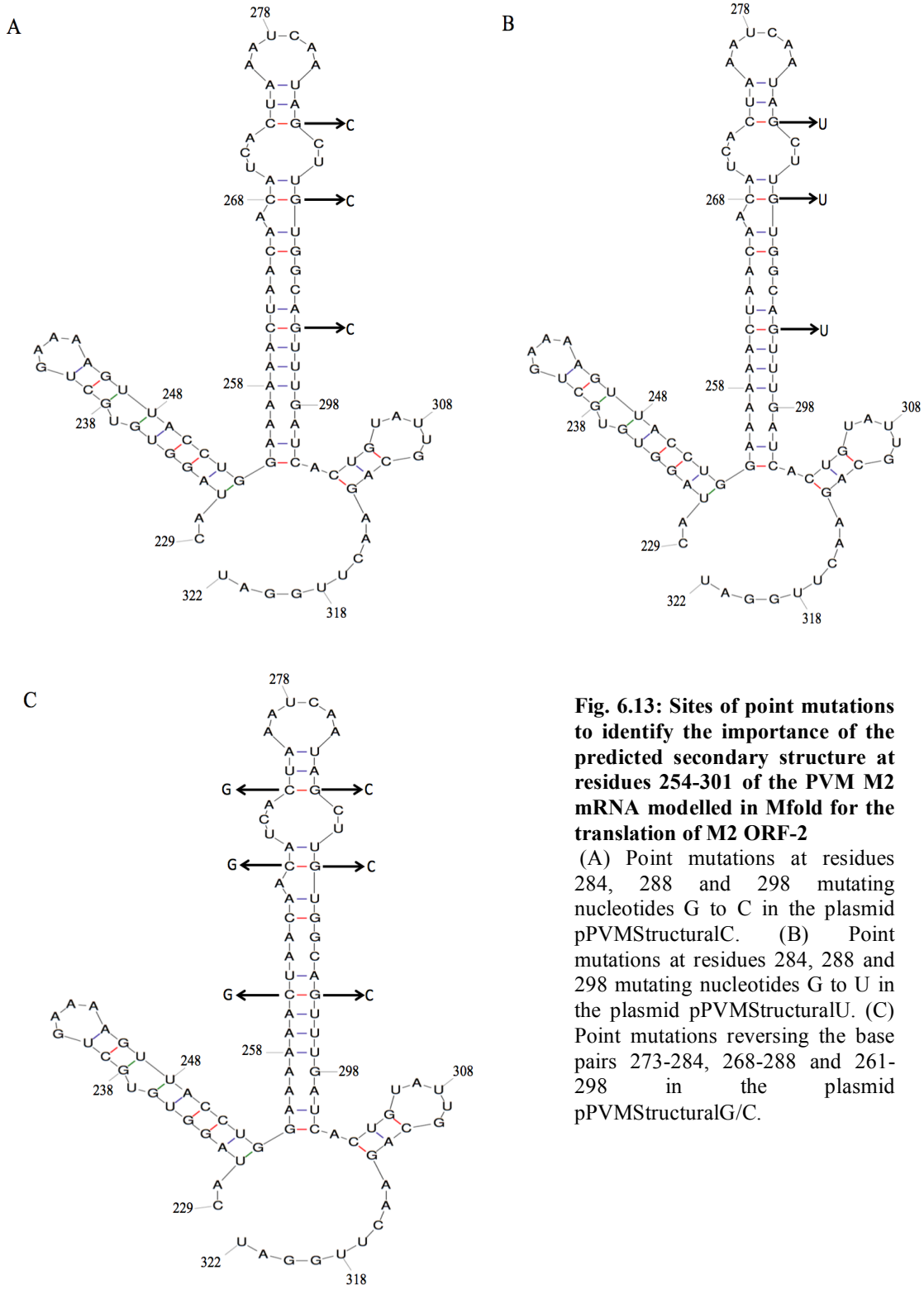


Fig. 6.13: Sites of point mutations to identify the importance of the predicted secondary structure at residues 254-301 of the PVM M2 mRNA modelled in Mfold for the translation of M2 ORF-2

(A) Point mutations at residues 284, 288 and 298 mutating nucleotides G to C in the plasmid pPVMStructuralC. (B) Point mutations at residues 284, 288 and 298 mutating nucleotides G to U in the plasmid pPVMStructuralU. (C) Point mutations reversing the base pairs 273-284, 268-288 and 261-298 in the plasmid pPVMStructuralG/C.

Construction of plasmids pPVMStructuralC, pPVMStructuralU and pPVMStructutalG/C

Two-step PCR.

The plasmids pPVMStructuralC, pPVMStructuralU and pPVMStructuralG/C had their respective point mutations mutated into M2 ORF-1 in the plasmid pPVMWC using two-step PCR. A diagram of the process is shown in Fig. 6.14. For each plasmid two PCR amplicons were initially generated in two separate PCR reactions in the first round of PCR using primers 1 and 3 (amplicon 1) and 2 and 4 (amplicon 2) (Fig. 6.14). Each amplicon had the mutations incorporated via primers 2 and 3 into the 5′ or 3′ end of the amplicon. The primers 2 and 3 were the reverse complement of the other so both amplicons contained a region of sequence homology. A second PCR reaction was performed with both amplicons and the external primers (1 and 4) from each of the first step reactions (Fig. 6.14). As the two amplicons contained regions of sequence homology they annealed during the reaction to generate the desired full length product. The newly generated amplicon was further amplified by the two external primers 1 and 4.

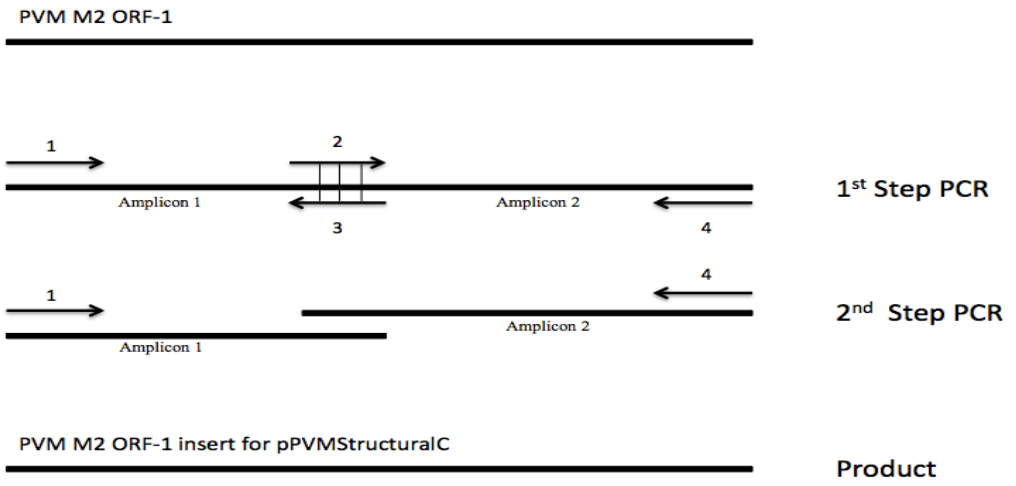


Fig. 6.14: Schematic of two-step PCR method to incorporate point mutations into pPVMWC

Schematic representation of the PCR for the method two-step PCR. For the first step primers 1, and 3 (Amplicon 1) and 2 and 4 (Amplicon 2) were used to generate two amplicon products that overlapped each other. Primers 2 and 3 were used to introduce point mutations into each amplicon. For the second step, primers 1 and 4 were used to join the two initial amplicons with overlapping sequences together with the newly incorporated point mutations in place. Horizontal lines between primers 2 and 3 represents the sites of the three point mutations

The cloning and sequencing method for all plasmids can be seen in Section 5.3.1.1. For the first step of the two-step PCR, two amplicons were generated using PCR from the template pPVMWC for each plasmid. The primer pairs used to amplify the amplicons are found in Table 6.2. All amplicons were 422 bp in size. After the first step, amplicons were purified using a QIAGEN PCR purification kit as described in Section 2.3.13. For the second step for each plasmid, the two amplicons were joined via a second round of PCR using the primers T7-91/T3-down (Primers 1 and 4 Fig. 6.14), this generated a fragment of 1354 bp in size.

Plasmid	Primer Pairs for each amplicon
pPVMStructuralG/C	T7-91/PstrucGC R and T3-down/ PstrucGC F
pPVMStructuralC	T7-91/PstrucC R and T3-down/ PstrucC F
pPVMStructuralU	T7-91/PstrucT R and T3-down/ PstrucT F

Table 6.2: Table of primer pairs used for two-step PCR
Table of primer pairs used for in the PVM fragment for first step of the two-step PCR used to construct the plasmids pPVMStructuralG/C, pPVMStructuralC and pPVMStructuralU.

6.6.3. Disruption of a predicted secondary to investigate its involvement in M2 ORF-2 translation

The plasmids described in 6.6.2 were individually transfected into HEp-2 cells as described in Section 2.4.1. Control cells were individually transfected with either pPVMWC or the empty vector pBluescribe. The level of CAT protein expression was quantified using a CAT ELISA assay as described in Section 2.4.2. All values were normalized against the results from transfected empty pBluescribe plasmid. For each sample, the levels of CAT expressed have been expressed as a percentage of the positive control plasmid pPVMWC.

As shown in Fig. 6.15, mutating each of the G residues at positions 284, 288 and 298 to C in plasmid pPVMStructuralC had no significant effect on CAT reporter expression with a level of 98% of the positive control plasmid transfection ($p=0.82$). Mutating the G residues at positions 284, 288 and 298 to U in the plasmid PVMStructuralU, reduced CAT protein expression to 92% of the control which was not significantly different ($p=0.27$) to the level obtained with pPVMStructuralC

(p=0.49). Taken together, these data indicate either that the predicted structure is not present, it is present but was unaffected by the mutations or that it is present but integrity of the structure is not required for the initiation of translation of the ORF-2.

In the plasmid pPVMStructuralG/C three additional mutations at positions 261, 268 and 273 were introduced into the mutant pPVMStructuralC. These mutations were designed to reinstate the predicted stem structure in the M2 mRNA but with the G and C residues now located on the opposite strands of the stem. As seen in Fig. 6.15, these additional mutations completely ablated expression of CAT protein (p=0.0005 compared to the positive control). The dramatic effect of the three additional mutations at positions 261, 268 and 273 combined with the mutations at positions 284, 288 and 298 suggests that the predicted stem structure is unlikely to exist in the mRNA. The data show that the region contains critical residues or a critical structure that is disrupted in the pPVMStructuralG/C mutant but further mutational analysis will be required to clarify the situation.

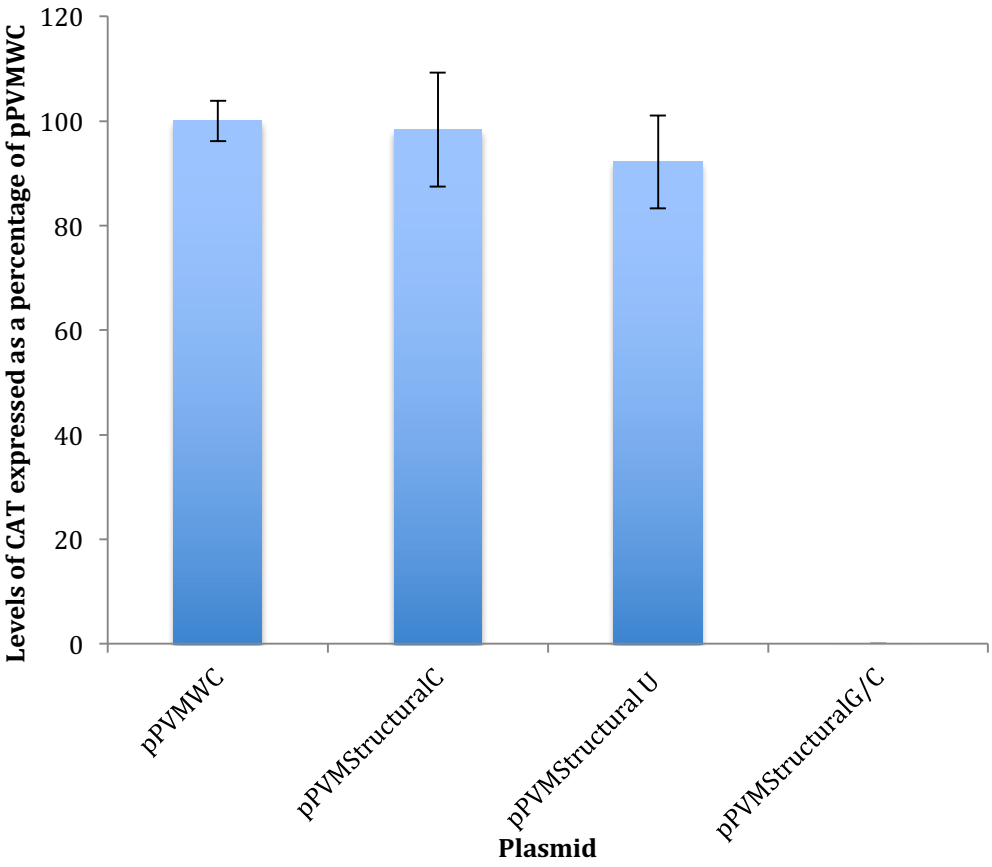


Fig. 6.15: Expression of CAT reporter protein from transfection of mutant plasmids into HEp-2 cells

Levels of expression of CAT protein in transfected HEp-2 cells from plasmids containing mutations designed to alter the structure of the PVM M2 mRNA as described in Section 6.6.2. Levels of CAT protein expression were normalised against the empty vector pBluescribe and are expressed as a percentage of the level obtained with the positive control plasmid pPVMWC. Error bars are standard deviation.

6.7. Conclusions

Data acquired from this chapter has further characterised the mechanism of coupled translation for initiation of translation of PVM M2 ORF-2. Chimeric reporter plasmids identified a region of sequence between positions 229 and 339 in the M2 ORF-1 that was used by the ribosome to initiate translation of M2 ORF-2 and was essential for successful translation. Removal of this region of sequence inhibited the mechanism of coupled translation.

This reliance on an upstream sequence has also been reported in the mechanism of coupled translation used to translate the hRSV M2 ORF-2 (Gould and Easton, 2005). In this example, this region of sequence was found to contain a structured region that when removed inhibited translation of M2 ORF-2 (P. Gould and A. Easton, personal communication). It was therefore decided to investigate if a structured region was present in this region of critical sequence for PVM M2 ORF-2. *In vitro* transcription reactions were therefore used to identify the presence of a large secondary structure through the production of premature termination products. Using this technique, a large stable RNA structure was identified during transcription of amplicons generated from the PVM reporter plasmid pPVMWC. This structure began inside the M2 ORF-1 sequence that was critical for translation of PVM M2 ORF-2. Mfold modeling software was used to characterise this structure by inputting this critical sequence for M2 ORF-2 translation. The software predicted a structure that consisted of three stem loops similar to the secondary structure critical for hRSV M2 ORF-2 translation (P. Gould and A. Easton, personal communication).

Although this structure was identified in the region critical for translation of M2 ORF-2, it was still unknown whether this structure was a factor in ORF-2 translation. A set of mutations were therefore made in plasmids that used the same reporter system as used in other plasmids in this chapter to measure ORF-2 expression. These mutations were designed both to disrupt the structure and retain the predicted structure, in an attempt to investigate if this predicted structure was a factor in ORF-2 translation. Although the plasmids that were designed to disrupt the structure had no effect on ORF-2 translation, translation of ORF-2 was inhibited for the plasmid designed to mutate key residues but retain the predicted structure. This plasmid had

six point mutations incorporated into this critical region located in the M2 ORF-1 sequence.

From these data, it is possible to conclude that PVM likely uses a similar mechanism to coupled translation in hRSV, where translation is reliant on an upstream structured region. It is likely that these point mutations that inhibited translation of ORF-2 disrupted the actual secondary structure located in this region. It is therefore likely that the secondary structure predicted by Mfold is not accurate and the secondary structure identified using *in vitro* transcription has an alternate and unknown structure.

Chapter 7
Discussion

7.1. Transcriptional profiles of bRSV and hRSV

Using transcriptional profiles to measure viral mRNA abundance in hRSV and bRSV infected cells, it was observed that a large proportion of viral mRNAs in both viruses at both 4 and 8 hpi had relative mRNA abundances consistent with the polar gradient model of transcription. For example, between the mRNAs expressed from the N to the G gene in both viruses at both time points, with the exception of the hRSV P gene at 4 hpi, there was a progressive decrease in viral mRNA abundance in the same sequential order as that of the genes from the 3´ promoter. However, there were also examples where the abundance of some viral mRNAs did not follow the expected gradient. For example, an increase in mRNA abundance was seen in both viruses at both time points between the G to F and F and M2 genes with the exception of bRSV M2 gene at 4 hpi. The reason for this variation at 4 hpi is unknown. As the mRNA levels measured were those at steady state and the data did not indicate true transcription rates of the genes, it is likely that this divergence from the polar gradient is due to the mRNA half-life of viral mRNAs differing from one another. While there is currently no evidence to support differing half-lives between viral mRNAs (Pennica et al., 1979). However, in reports where a divergence from the polar gradient model was reported, it was considered that this may be due to differences in mRNA half-life (Aljabr et al., 2016; Krempl et al., 2002). For cellular mRNAs, half-lives can differ significantly and the rates of decay are dependent on multiple factors such as sequence and size of mRNAs (Yang et al., 2003). As the half-life for viral mRNAs is dictated by cellular decay pathways (Tycowski et al., 2012), it is quite possible that the same factors have a similar influence on viral mRNA half-lives.

The data also shows that mRNA abundance may be consistent with the polar gradient model at one time point but not another. For example between the NS2 and N genes at 8 hpi but not at 4 hpi in both viruses, there was an increase in mRNA abundance (4.8% increase in hRSV and 3.8% increase in bRSV). This suggests that the mRNA half-life changes during infection. In agreement with this, at 24 and 48 hpi the M and G genes had the highest levels of mRNA abundance (Aljabr et al., 2016) but at 4 and 8 hpi, the M mRNA was closer to the median mRNA abundance level and the G mRNA had the second lowest abundance at both time points. This was also seen with the NS1 gene, where the levels of mRNA were significantly

lower as a proportion of total mRNA at 24 and 48 hpi (Aljabr et al., 2016) compared to the levels at 4 and 8 hpi reported here (Section 3.4).

These data suggest that if the half-life differs for each viral mRNA at different stages of infection, then control of viral protein expression is unlikely to occur based solely on mRNA abundance. Therefore, it is likely that translational regulation of virus mRNAs also plays a role in determining levels of protein.

While transcriptional profiling provides evidence of the steady state levels of mRNA, it does not directly address the question of whether virus transcription in pneumoviruses follows the polar gradient model. To address this directly, it will be necessary to also investigate viral mRNA half-life and to look at the levels of nascent RNA at different times during infection.

7.2. Translational profile of bRSV and hRSV

It is clear that from translational efficiencies of the individual mRNAs, translational regulation plays a role in gene expression in both bRSV and hRSV. In both viruses, only the G mRNA was highly translated (upregulated) and the L mRNA poorly translated (downregulated) at both 4 and 8 hpi. The P mRNA was also translated efficiently in both viruses at all time points with the exception of hRSV P mRNA extracted at 4 hpi. For the SH mRNAs and M2 ORF-1 for bRSV and the NS1 and SH mRNAs for hRSV the translation efficiencies were regulated at both time points.

Further evidence for the presence of translational regulation was seen for the hRSV SH mRNA at 8 hpi. The mRNA abundance for this mRNA was at very low levels of 1.4% in contrast to 9.4% for the bRSV SH mRNA. However, the translational efficiency of the hRSV had a value of 6.57 (indicating upregulation), while the more abundant bRSV SH mRNA had a translational efficiency of 0.53 (indicating downregulation). This suggests that the level of the SH protein for these viruses is achieved by a balance between mRNA levels (which likely differs as a result of different stabilities), and translation efficiencies. In both viruses there was also evidence of changes in mRNA translation efficiencies for specific mRNAs at different time points. This is seen with the bRSV NS1 mRNA and the hRSV N and

NS2 mRNAs (Table 3.3). This may be a mechanism to control the levels of protein at different stages of infection due to changing demand for the protein.

For the bRSV M2 ORF-1, regulation may occur due to the non-canonical mechanism used for translation M2 ORF-2 affecting translation of M2 ORF-1. This is unsurprising, since initiation of translation for M2 ORF-2 appears to take place inside M2 ORF-1, which is likely to disrupt translation of M2 ORF-1 (see later discussion).

It is currently understood that regulation of translation primarily occurs at the initiation stage of translation (Fütterer et al., 1997; Jacks and Varmus, 1985; Tsukiyama-Kohara et al., 1992). It is likely, with the exception of the M2 mRNA, that both bRSV and hRSV regulate translation through a canonical initiation pathway. The precise factors determining translation efficiencies of mRNAs are not known. One possible factor is that the virus mRNAs compete for cellular resources present in limiting concentrations and that some mRNAs are more efficient at binding to these. An example of limiting resources may be specific initiation factors that become limiting in the face of the large increase in overall mRNA levels that occur during virus infection. This may also be affected by the different stages of infection where there is a change in cellular metabolism. For example, a decrease in the levels of eIF4F (containing eIF4A) at later stages of infection may decrease successful initiation of mRNAs with highly structured 5′ UTRs. Another possible factor is the codon usage. In this, mRNAs containing sequences for multiple rare codons are likely to have lower levels of translation due to a relative shortage of rare tRNAs. In this case, it is not clear if the translation efficiency would be likely to alter throughout infection. Further analysis of the sequence of each mRNA may clarify this. A specific factor relevant for pneumovirus mRNAs is that the viruses can produce polycistronic mRNAs and the proportion of each of these is specific for each gene junction and is different for different virus strains as has been reported for the P mRNA in the related measles virus (Section 1.6.3) (Liston and Briedis, 1995). This could indirectly affect translation, as translation of the downstream ORF is likely to be rare. The frequency of polycistronic mRNAs in pneumoviruses has not been established.

An important consideration is that the data acquired for bRSV was obtained from a single experiment and replicates are required to provide confidence for assessment of mRNAs where the differences in translation efficiencies were small. This will also provide additional repeats for transcriptional profile data of bRSV. In addition, it will be necessary to confirm the translational regulation for specific mRNAs using biochemical analyses.

7.3. The mechanism of translation of bRSV M2 ORF-2

The hRSV M2 ORF-2 has been shown to be translated using a novel process of coupled translation termination/initiation (Section 1.4.2). It was anticipated that the very closely related and highly homologous bRSV would use the same process and this was investigated using a reporter gene approach that was used to characterise the hRSV system. Initially it was shown that the AUG codon at position 563 located in the M2 ORF-1/ORF-2 overlap region was used primarily to initiate translation of the bRSV M2 ORF-2 (Section 4.4) though initiation could occur at 20-fold lower levels at the AUG codon at position 554 if the preferred AUG codon was mutated. If the AUG at codon 554 was also used as a result of leaky scanning (Jacks et al., 1988b) this would result in an M2-2 protein that contained an additional 3 amino terminal residues compared to the product initiated at the preferred AUG 563.

While in principle ribosomal profiling data could be used to identify where ribosomes are initiating translation of bRSV M2 ORF-2, the levels of translation of M2 ORF-2 were extremely low making it difficult to distinguish initiating ribosomes in the M2 ORF-1/ORF-2 overlap region from elongating ribosomes translating M2 ORF-1. It is possible that the levels of translation of M2 ORF-2 increase at later stages of infection as the requirement for M2-2 protein increases with the switch from viral transcription to replication (Bermingham and Collins, 1999). Ribosomal profiling of samples extracted between 12 to 15 hpi, when levels of translation of M2 ORF-2 may be higher may prove more suitable for identification of the AUG codons used for translation.

As described in Chapter 5 the Kozak sequence surrounding the M2 ORF-2 initiation codon was a factor in successful initiation. This is commonly seen in other non-canonical scanning mechanisms (Racine et al., 2007; Wise et al., 2009) but not in

members of the *Caliciviridae* family that use a form of coupled translation to access an overlapping ORF (Luttermann and Meyers, 2014). This may also explain the preference for initiation at position 563 instead at position 554, as the preferred AUG codon has a stronger Kozak sequence. However mutation of the Kozak sequence at AUG 563 suggested that this was not the only factor in initiation site preference (Tikole and Sankararamakrishnan, 2008). Mutation of the Kozak sequence adjacent to the AUG at position 563 to introduce an optimal sequence reduced translation of M2 ORF-2. This suggests that the sequence surrounding the AUG codon is already optimal for translation of M2 ORF-2.

The insertion of two separate artificial AUG codons for M2 ORF-2 translation led to a failure to initiate translation when both native start codons for M2 ORF-2 were mutated (Section 5.4). The Kozak sequences for these artificial codons were stronger than the Kozak sequence for the AUG codon at position 554. The failure of these optimised codons to initiate translation of ORF-2 suggests that the location of the codon may also be important for successful translation. It also suggests that the Kozak sequence is not an essential factor in the ribosome choosing the location of initiation of M2 ORF-2. Further experiments to establish whether the location of the AUG codon is a factor would involve moving the sequence of the preferred AUG codon including the surrounding Kozak sequence further downstream and would indicate if the position of this codon relative to some other region may be important.

The finding that the Kozak sequence is not the sole factor in the site of initiation of translation of the M2 ORF-2 suggests that a scanning mechanism such as leaky scanning is not used to access ORF-2 as the 'strength' of the Kozak sequence is not a priority. This is also consistent with the data obtained from the plasmids containing the artificial start codons. The data suggests that the preference for the AUG codon at position 563 is based both on its location and on the nature of the upstream sequence with the latter acting in an unknown role.

A key characteristic of coupled translation termination/initiation common to all examples, is that the upstream ORF termination codon is located sufficiently close to the initiation codon for the downstream ORF to allow reinitiation (Ahmadian et al., 2000; Gould et al., 2014; Gould and Easton, 2007; Luttermann and Meyers, 2007).

Moving the terminating codon to increase this distance either significantly reduces or completely inhibits the translation of the second ORF. Mutating the bRSV M2 ORF-1 stop codon and increasing the ORF-1/ORF-2 overlap region to 92 nts had no effect on translation of ORF-2 (Section 4.5). In these experiments the increased length of the bRSV M2 ORF-1/ORF-2 overlap region was significantly longer than the size of the M2 ORF-1/ORF-2 overlap region required to inhibit translation of M2 ORF-2 in other members of the *Pneumoviridae* family (Ahmadian et al., 2000; Gould and Easton, 2007). This indicates that bRSV does not use coupled translation to access the M2 ORF-2. Ribosomal profiling provided further evidence that initiation of translation of bRSV M2 ORF-2 was not achieved using coupled translation (Section 4.6). In these data it was clear there were no significant examples of ribosomes stalling in a similar manner observed in hRSV (Phillip Gould and Andrew Easton, personal communication) or that the translation between the two ORFs were linked.

The data reported here shows that bRSV is the first member of the *Orthopneumovirus* family that does not use the mechanism of coupled translation. While there was a report that the *Metapneumovirus* HMPV does not use coupled translation to access the M2 ORF-2, this was not proved conclusively. As described in Section 1.6.1, translation of the HMPV M2 ORF-2 was not inhibited when translation of M2 ORF-1 was prevented. This was interpreted to suggest that HMPV M2 ORF-2 is not reliant on translation of M2 ORF-1. However the key experiment of mutating the M2 ORF-1 termination codon and observing the effect on M2 ORF-2 expression was not performed (Buchholz et al., 2005). Since preventing M2 ORF-1 translation leaves open the possibility that with these mutants the M2 ORF-2 was accessed by leaky scanning, it remains possible that coupled translation is used for this mRNA.

It is likely that a non-canonical mechanism is used to initiate translation of the bRSV M2 ORF-2. It is unlikely that leaky scanning is used, as a ribosome initiating translation at the preferred AUG codon at position 563 would have to bypass 14 other potential initiation codons, 3 of which have similar strength Kozak sequences. Although the distance from the preferred start codon is not too far for leaky scanning to occur (Fütterer et al., 1997), the efficiency of leaky scanning decreases with distance. This is further supported by transfection and expression of ORF-2 from the

plasmid pbRSVWC14 (Section 5.6). With this plasmid, the ribosome would have to bypass a total of 22 codons and travel 1371 nts to reach the desired AUG codon. This is over 1.5 times the largest distance reported to be used by leaky scanning, making it extremely unlikely (Fütterer et al., 1997).

The absence of a fusion protein in cells transfected with the plasmid pbRSVWC42 suggests that programmed ribosomal frameshifting is not used (Section 5.2). There was also no evidence of a slippery sequence surrounding the site of initiation. Although it is important to note that in cases of -2 and +1 frameshifting, the slippery sequence appears to be case specific (Matsufuji et al., 1996; Olivier et al., 2008). However the majority of -2 and +1 mechanisms are still reliant on downstream structured regions in the vicinity of the slippery sequence (Ivanov and Atkins, 2007). In the creation of the reporter plasmids the sequence directly downstream of the initiation site for M2 ORF-2 was replaced with CAT coding sequence and the continued expression of the M2 ORF-2 in these plasmids indicates that such an essential slippery sequence is not required for M2 ORF-2 expression in bRSV.

It is also unlikely that ribosomal shunting is used to translate M2 ORF-2 by a ribosome translating a short ORF in the 5′ UTR in order to retain certain initiation factors for a jump, as the M2 mRNA 5′UTR is only 9 nts long. Similarly if shunting occurred after termination of M2 ORF-1, mutating the ORF-1 stop codon thereby changing the donor site, also had no effect on translation, suggesting this mechanism is not used in the bRSV M2 mRNA. It has been shown for the late adenovirus mRNAs, that instead of the ribosome using initiation factors gained from translating a short ORF to bypass the 5′ UTR, the ribosome can utilise complementary sequences to the 3′ end of the 18S rRNA to facilitate the jump from the 5′ UTR to the internal initiation position (Yueh and Schneider, 2000). While a short sequence complementary to the 3′ end of the 18S rRNA was identified in the bRSV M2 ORF-1, mutating this sequence to either improve or reduce the complementarity of this sequence had no effect on the translation of ORF-2. Finally, altering the M2 ORF-1 sequence with the addition of 708 nts to ORF-1 and a new 5′ UTR in the reporter plasmid pbRSVWC14, did not reduce or eliminate translation of ORF-2 (Fig. 5.16). It is therefore unlikely that ribosomal shunting is used to initiate translation of M2 ORF-2.

The use of chimeric reporter assays identified a region upstream of the M2 ORF-1/ORF-2 overlap region that was critical for translation of M2 ORF-2 (Fig. 5.16). Studies with bicistronic reporter assays to test this critical region of sequence suggest that an internal initiation mechanism is likely to be used by bRSV for initiation of translation of M2 ORF-2 (Section 5.6). The sequence between position 210 and 563 was able to facilitate internal translation when placed in the context of an unrelated ORF (Section 5.8). While a stable secondary structure is not always required for an IRES, and in some cases can be inhibitory (Filbin and Kieft, 2009), such regions are frequently identified in association with internal initiation processes. However, there was no evidence of a stable secondary structure in this region (Section 5.5.2.1). Further work is required to establish if the M2 IRES-like region loads the ribosome directly onto the AUG codon or requires the ribosome to scan a distance downstream (Jaafar et al., 2016). Further work is also required to ensure that plasmids used to test for internal translation contained no cryptic promoters. Although it is important to note that there is currently no evidence of T7 promoters acting as cryptic promoters when transfected into eukaryotic cells that have not been infected with vaccinia virus that expresses T7 RNA polymerase (Ray et al., 2006, Dhar et al., 2007).

An unusual feature is the putative stem loop structure identified between the IRES-like sequence and the initiation codon for M2 ORF-2, that reduces the efficiency of translation for M2 ORF-2 (Section 5.5.1.2). Structural predictions suggest that most of the M2 ORF-1/ORF-2 overlap region, including the preferred initiation codon for M2 ORF-2, is base paired in the stem of this structure. A point mutation inserted into this sequence designed to disrupt this structure without altering the Kozak sequence significantly improved ORF-2 expression. The increase in ORF-2 translation may result as a consequence of the ribosome not having to unwind the structure to access the preferred AUG codon for translation of ORF-2. This is consistent with studies showing that structured regions can inhibit successful initiation and require unwinding before efficient initiation can occur (Oppenheim and Yanofsky, 1980). To further characterise this inhibitory structure, additional mutations to further destabilise it, in conjunction with biochemical analyses, will clarify the role and potential function of this region.

7.4. Coupled translation termination/initiation for initiation of translation of PVM M2 ORF-2

It has been shown previously that the PVM M2 ORF-2 is accessed by coupled translation (Gould and Easton, 2007). Using chimeric reporter plasmids based on the PVM M2 mRNA it was shown that a region of sequence upstream of the PVM M2 ORF-1/ORF-2 overlap region between positions 229 and 332 was required for translation of ORF-2. When this region was removed, translation of M2 ORF-2 was reduced to basal levels similar to those observed when the M2 ORF-1 terminating stop codon was mutated (Gould and Easton, 2007) (Section 6.3). The progress of T7 RNA polymerase was reduced during *in vitro* transcription across this region strongly suggesting that it contained a secondary structure as had been demonstrated for the hRSV M2 mRNA (Gould and Easton, 2005). The Mfold software predicted a secondary structure consisting of three stem loops that was similar to the structured region critical for hRSV M2 ORF-2 translation (P. Gould and A. Easton, personal communication). Mutations designed to alter the predicted structure led to complete inhibition of ORF-2 translation (Section 6.6). This is similar to the situation for hRSV M2 ORF-2 translation (Gould and Easton, 2005).

This evidence combined with a highly similar mechanism of coupled translation reported in hRSV (reviewed in Section 1.4.2), makes it highly likely that the secondary structure in PVM acts in a similar mechanism to the hRSV structure (Section 1.4.2). The secondary structure in the hRSV M2 mRNA appears to be more stable than the structure in the PVM M2 mRNA and this may play a role in the 10 fold difference in the levels of M2 ORF-2 expression between the two different viruses (Gould and Easton, 2007). However, several additional factors such as size of the M2 ORF-1/ORF-2 overlap region may also play a role and further investigation is required to define the reasons for the differences between the two virus systems. Such studies would involve additional mutational analysis within the region to disrupt and map the critical structure. Biochemical analysis such as SHAPE mapping would also further define the secondary structure. Ribosomal profiling would then be used to identify if the secondary structure is used by PVM in the same manner as in hRSV (P. Gould and A. Easton, personal communication).

7.5. Concluding remarks

Current understanding of the control of gene expression in pneumoviruses is limited. The data presented here indicates that there are both transcriptional and translational regulatory factors that play a role in determining the final level of expression of gene products. It is also clear that although many similarities are shared between the closely related bRSV, PVM and hRSV, the mechanism of translation of M2 ORF-2 differs between these viruses. While hRSV and PVM appear to use very similar processes to access and initiate translation of M2 ORF-2, a different mechanism is used for translation of bRSV M2 ORF-2. Translation of the bRSV M2 ORF-2 requires an upstream sequence located within the M2-1 ORF, which is similar to PVM and hRSV. However for bRSV, this sequence appears to direct ribosomes to initiate translation internally within the M2 mRNA rather than to promote coupled translation termination/initiation that is seen in the other *Orthopneumoviruses*. Whilst the outcome of these different processes is the same with the expression of ORF-2, the data demonstrates that the viruses can achieve this by using different mechanisms. This raises fundamental questions about the evolutionary relationships of the viruses.

Chapter 8
References

Abraham, G., and Banerjee, A.K. (1976). Sequential transcription of the genes of vesicular stomatitis virus. Proc Natl Acad Sci U S A *73*, 1504-1508.

Ahmadian, G., Randhawa, J.S., and Easton, A.J. (2000). Expression of the ORF-2 protein of the human respiratory syncytial virus M2 gene is initiated by a ribosomal termination-dependent reinitiation mechanism. EMBO J *19*, 2681-2689.

Aksoy, S., Squires, C.L., and Squires, C. (1984). Translational coupling of the trpB and trpA genes in the Escherichia coli tryptophan operon. J Bacteriol *157*, 363-367.

Aljabr, W., Touzelet, O., Pollakis, G., Wu, W., Munday, D.C., Hughes, M., Hertz-Fowler, C., Kenny, J., Fearns, R., Barr, J.N.*, et al.* (2016). Investigating the Influence of Ribavirin on Human Respiratory Syncytial Virus RNA Synthesis by Using a High-Resolution Transcriptome Sequencing Approach. J Virol *90*, 4876-4888.

Alvarez, R., Lwamba, H.M., Kapczynski, D.R., Njenga, M.K., and Seal, B.S. (2003). Nucleotide and predicted amino acid sequence-based analysis of the avian metapneumovirus type C cell attachment glycoprotein gene: phylogenetic analysis and molecular epidemiology of U.S. pneumoviruses. J Clin Microbiol *41*, 1730-1735.

Asenjo, A., and Villanueva, N. (2016). Phosphorylation of the human respiratory syncytial virus P protein mediates M2-2 regulation of viral RNA synthesis, a process that involves two P proteins. Virus Res *211*, 117-125.

Bakker, S.E., Duquerroy, S., Galloux, M., Loney, C., Conner, E., Eléouët, J.F., Rey, F.A., and Bhella, D. (2013). The respiratory syncytial virus nucleoprotein-RNA complex forms a left-handed helical nucleocapsid. J Gen Virol *94*, 1734-1738.

Balvay, L., Soto Rifo, R., Ricci, E.P., Decimo, D., and Ohlmann, T. (2009). Structural and functional diversity of viral IRESes. Biochim Biophys Acta *1789*, 542-557.

Barik, S. (1993). The structure of the 5' terminal cap of the respiratory syncytial virus mRNA. J Gen Virol *74 (Pt 3)*, 485-490.

Behera, A.K., Matsuse, H., Kumar, M., Kong, X., Lockey, R.F., and Mohapatra, S.S. (2001). Blocking intercellular adhesion molecule-1 on human epithelial cells decreases respiratory syncytial virus infection. Biochem Biophys Res Commun *280*, 188-195.

Bermingham, A., and Collins, P.L. (1999). The M2-2 protein of human respiratory syncytial virus is a regulatory factor involved in the balance between RNA replication and transcription. Proc Natl Acad Sci U S A *96*, 11259-11264.

Biswas, P., Jiang, X., Pacchia, A.L., Dougherty, J.P., and Peltz, S.W. (2004). The human immunodeficiency virus type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target for antiviral therapy. J Virol *78*, 2082-2087.

Bitko, V., Shulyayeva, O., Mazumder, B., Musiyenko, A., Ramaswamy, M., Look, D.C., and Barik, S. (2007). Nonstructural proteins of respiratory syncytial virus suppress premature apoptosis by an NF-kappaB-dependent, interferon-independent mechanism and facilitate virus growth. J Virol *81*, 1786-1795.

Blodörn, K., Hägglund, S., Fix, J., Dubuquoy, C., Makabi-Panzu, B., Thom, M., Karlsson, P., Roque, J.L., Karlstam, E., Pringle, J.*, et al.* (2014). Vaccine safety and efficacy evaluation of a recombinant bovine respiratory syncytial virus (BRSV) with deletion of the SH gene and subunit vaccines based on recombinant human RSV proteins: N-nanorings, P and M2-1, in calves with maternal antibodies. PLoS One *9*, e100392.

Blodörn, K., Hägglund, S., Gavier-Widen, D., Eléouët, J.F., Riffault, S., Pringle, J., Taylor, G., and Valarcher, J.F. (2015). A bovine respiratory syncytial virus model with high clinical expression in calves with specific passive immunity. BMC Vet Res *11*, 76.

Blondot, M.L., Dubosclard, V., Fix, J., Lassoued, S., Aumont-Nicaise, M., Bontems, F., Eléouët, J.F., and Sizun, C. (2012). Structure and functional analysis of the RNA- and viral phosphoprotein-binding domain of respiratory syncytial virus M2-1 protein. PLoS Pathog *8*, e1002734.

Blount, R.E., Morris, J.A., and Savage, R.E. (1956). Recovery of cytopathogenic agent from chimpanzees with coryza. Proc Soc Exp Biol Med *92*, 544-549.

Borchers, A.T., Chang, C., Gershwin, M.E., and Gershwin, L.J. (2013). Respiratory syncytial virus--a comprehensive review. Clin Rev Allergy Immunol *45*, 331-379.

Brierley, I., Boursnell, M.E., Binns, M.M., Bilimoria, B., Blok, V.C., Brown, T.D., and Inglis, S.C. (1987). An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. EMBO J *6*, 3779-3785.

Brierley, I., Digard, P. and Inglis, S. C. (1989) Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell,* 57(4), 537-47.

Brock, L.G., Karron, R.A., Krempl, C.D., Collins, P.L., and Buchholz, U.J. (2012). Evaluation of pneumonia virus of mice as a possible human pathogen. J Virol *86*, 5829-5843.

Buchholz, U.J., Biacchesi, S., Pham, Q.N., Tran, K.C., Yang, L., Luongo, C.L., Skiadopoulos, M.H., Murphy, B.R., and Collins, P.L. (2005). Deletion of M2 gene open reading frames 1 and 2 of human metapneumovirus: effects on RNA synthesis, attenuation, and immunogenicity. J Virol *79*, 6588-6597.

Buchholz, U.J., Finke, S., and Conzelmann, K.K. (1999). Generation of bovine respiratory syncytial virus (BRSV) from cDNA: BRSV NS2 is not essential for virus replication in tissue culture, and the human RSV leader region acts as a functional BRSV genome promoter. J Virol *73*, 251-259.

Burke, E., Dupuy, L., Wall, C., and Barik, S. (1998). Role of cellular actin in the gene expression and morphogenesis of human respiratory syncytial virus. Virology *252*, 137-148.

Burke, E., Mahoney, N.M., Almo, S.C., and Barik, S. (2000). Profilin is required for optimal actin-dependent transcription of respiratory syncytial virus genome RNA. J Virol *74*, 669-675.

Caliskan, N., Peske, F., and Rodnina, M.V. (2015). Changed in translation: mRNA recoding by -1 programmed ribosomal frameshifting. Trends Biochem Sci *40*, 265-274.

Cartee, T.L., and Wertz, G.W. (2001). Respiratory syncytial virus M2-1 protein requires phosphorylation for efficient function and binds viral RNA during infection. J Virol *75*, 12188-12197.

Castaño, A., Ruiz, L., and Hernández, C. (2009). Insights into the translational regulation of biologically active open reading frames of Pelargonium line pattern virus. Virology *386*, 417-426.

Cattaneo, R., Rebmann, G., Schmid, A., Baczko, K., ter Meulen, V., and Billeter, M.A. (1987). Altered transcription of a defective measles virus genome derived from a diseased human brain. EMBO J *6*, 681-688.

Chappell, S.A., Dresios, J., Edelman, G.M., and Mauro, V.P. (2006). Ribosomal shunting mediated by a translational enhancer element that base pairs to 18S rRNA. Proc Natl Acad Sci U S A *103*, 9488-9493.

Cheng, X., Park, H., Zhou, H., and Jin, H. (2005). Overexpression of the M2-2 protein of respiratory syncytial virus inhibits viral replication. J Virol *79*, 13943-13952.

Collins, P.L., Fearns, R., and Graham, B.S. (2013). Respiratory syncytial virus: virology, reverse genetics, and pathogenesis of disease. Curr Top Microbiol Immunol *372*, 3-38.

Collins, P.L., and Graham, B.S. (2008). Viral and host factors in human respiratory syncytial virus pathogenesis. J Virol *82*, 2040-2055.

Collins, P.L., Hill, M.G., Cristina, J., and Grosfeld, H. (1996). Transcription elongation factor of respiratory syncytial virus, a nonsegmented negative-strand RNA virus. Proc Natl Acad Sci U S A *93*, 81-85.

Collins, P.L., and Mottet, G. (1993). Membrane orientation and oligomerization of the small hydrophobic protein of human respiratory syncytial virus. J Gen Virol *74 (Pt 7)*, 1445-1450.

Collins, P.L., Olmsted, R.A., Spriggs, M.K., Johnson, P.R., and Buckler-White, A.J. (1987). Gene overlap and site-specific attenuation of transcription of the viral polymerase L gene of human respiratory syncytial virus. Proc Natl Acad Sci U S A *84*, 5134-5138.

Collins, P.L., and Wertz, G.W. (1983). cDNA cloning and transcriptional mapping of nine polyadenylylated RNAs encoded by the genome of human respiratory syncytial virus. Proc Natl Acad Sci U S A *80*, 3208-3212.

Connor, J. H. and Lyles, D. S. (2002). Vesicular stomatitis virus infection alters the eIF4F translation initiation complex and causes dephosphorylation of the eIF4E binding protein 4E-BP1. J Virol, 76(20), pp. 10177-87.

Corti, D., Bianchi, S., Vanzetta, F., Minola, A., Perez, L., Agatic, G., Guarino, B., Silacci, C., Marcandalli, J., Marsland, B.J.*, et al.* (2013). Cross-neutralization of four paramyxoviruses by a human monoclonal antibody. Nature *501*, 439-443.

Cowton, V.M., McGivern, D.R., and Fearns, R. (2006). Unravelling the complexities of respiratory syncytial virus RNA synthesis. J Gen Virol *87*, 1805-1821.

Das, A., and Yanofsky, C. (1989). Restoration of a translational stop-start overlap reinstates translational coupling in a mutant trpB'-trpA gene pair of the Escherichia coli tryptophan operon. Nucleic Acids Res *17*, 9333-9340.

de Breyne, S., Simonet, V., Pelet, T., and Curran, J. (2003). Identification of a cis-acting element required for shunt-mediated translational initiation of the Sendai virus Y proteins. Nucleic Acids Res *31*, 608-618.

Dever, T.E., and Green, R. (2012). The elongation, termination, and recycling phases of translation in eukaryotes. Cold Spring Harb Perspect Biol *4*, a013706.

Dhar, D., Roy, S. and Das, S. (2007). Translational control of the interferon regulatory factor 2 mRNA by IRES element. Nucleic Acids Res 35, pp. 5409-21.

Dinman, J.D. (2012). Mechanisms and implications of programmed translational frameshifting. Wiley Interdiscip Rev RNA *3*, 661-673.

Dyer, K.D., Garcia-Crespo, K.E., Glineur, S., Domachowske, J.B., and Rosenberg, H.F. (2012). The Pneumonia Virus of Mice (PVM) model of acute respiratory infection. Viruses *4*, 3494-3510.

El Omari, K., Dhaliwal, B., Ren, J., Abrescia, N.G., Lockyer, M., Powell, K.L., Hawkins, A.R., and Stammers, D.K. (2011). Structures of respiratory syncytial virus nucleocapsid protein from two crystal forms: details of potential packing interactions in the native helical form. Acta Crystallogr Sect F Struct Biol Cryst Commun *67*, 1179-1183.

Fearns, R., and Collins, P.L. (1999a). Model for polymerase access to the overlapped L gene of respiratory syncytial virus. J Virol *73*, 388-397.

Fearns, R., and Collins, P.L. (1999b). Role of the M2-1 transcription antitermination protein of respiratory syncytial virus in sequential transcription. J Virol *73*, 5852-5864.

Fernández, I.S., Bai, X.C., Murshudov, G., Scheres, S.H., and Ramakrishnan, V. (2014). Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. Cell *157*, 823-831.

Filbin, M.E., and Kieft, J.S. (2009). Toward a structural understanding of IRES RNA function. Curr Opin Struct Biol *19*, 267-276.

Firth, A.E., Jagger, B.W., Wise, H.M., Nelson, C.C., Parsawar, K., Wills, N.M., Napthine, S., Taubenberger, J.K., Digard, P., and Atkins, J.F. (2012). Ribosomal frameshifting used in influenza A virus expression occurs within the sequence UCC_UUU_CGU and is in the +1 direction. Open Biol *2*, 120109.

Fleming, D.M., Pannell, R.S., and Cross, K.W. (2005). Mortality in children from influenza and respiratory syncytial virus. J Epidemiol Community Health *59*, 586-590.

Fletcher, J.N., Smyth, R.L., Thomas, H.M., Ashby, D., and Hart, C.A. (1997). Respiratory syncytial virus genotypes and disease severity among children in hospital. Arch Dis Child *77*, 508-511.

Flower, A.M., and McHenry, C.S. (1990). The gamma subunit of DNA polymerase III holoenzyme of Escherichia coli is produced by ribosomal frameshifting. Proc Natl Acad Sci U S A *87*, 3713-3717.

Fuller, F., Bhown, A.S., and Bishop, D.H. (1983). Bunyavirus nucleoprotein, N, and a non-structural protein, NSS, are coded by overlapping reading frames in the S RNA. J Gen Virol *64 (Pt 8)*, 1705-1714.

Fütterer, J., and Hohn, T. (1991). Translation of a polycistronic mRNA in the presence of the cauliflower mosaic virus transactivator protein. EMBO J *10*, 3887-3896.

Fütterer, J., Rothnie, H.M., Hohn, T., and Potrykus, I. (1997). Rice tungro bacilliform virus open reading frames II and III are translated from polycistronic pregenomic RNA by leaky scanning. J Virol *71*, 7984-7989.

Gaddum, R.M., Cook, R.S., Furze, J.M., Ellis, S.A., and Taylor, G. (2003). Recognition of bovine respiratory syncytial virus proteins by bovine CD8+ T lymphocytes. Immunology *108*, 220-229.

Galloux, M., Tarus, B., Blazevic, I., Fix, J., Duquerroy, S., and Eléouët, J.F. (2012). Characterization of a viral phosphoprotein binding site on the surface of the respiratory syncytial nucleoprotein. J Virol *86*, 8375-8387.

Garbelli, A., Beermann, S., Di Cicco, G., Dietrich, U., and Maga, G. (2011). A motif unique to the human DEAD-box protein DDX3 is important for nucleic acid binding, ATP hydrolysis, RNA/DNA unwinding and HIV-1 replication. PLoS One *6*, e19810.

García, J., García-Barreno, B., Vivo, A., and Melero, J.A. (1993). Cytoplasmic inclusions of respiratory syncytial virus-infected cells: formation of inclusion bodies in transfected cells that coexpress the nucleoprotein, the phosphoprotein, and the 22K protein. Virology *195*, 243-247.

Gershwin, L.J. (2012). Immunology of bovine respiratory syncytial virus infection of cattle. Comp Immunol Microbiol Infect Dis *35*, 253-257.

Ghattas, I.R., Sanes, J.R., and Majors, J.E. (1991). The encephalomyocarditis virus internal ribosome entry site allows efficient coexpression of two genes from a recombinant provirus in cultured cells and in embryos. Mol Cell Biol *11*, 5848-5859.

Ghildyal, R., Ho, A., and Jans, D.A. (2006). Central role of the respiratory syncytial virus matrix protein in infection. FEMS Microbiol Rev *30*, 692-705.

Glazier, K., Raghow, R., and Kingsbury, D.W. (1977). Regulation of Sendai virus transcription: evidence for a single promoter in vivo. J Virol *21*, 863-871.

Gomez, R.S., Guisle-Marsollier, I., Bohmwald, K., Bueno, S.M., and Kalergis, A.M. (2014). Respiratory Syncytial Virus: Pathology, therapeutic drugs and prophylaxis. Immunol Lett.

González-Reyes, L., Ruiz-Argüello, M.B., García-Barreno, B., Calder, L., López, J.A., Albar, J.P., Skehel, J.J., Wiley, D.C., and Melero, J.A. (2001). Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion. Proc Natl Acad Sci U S A *98*, 9859-9864.

Goodman, R.P., Freret, T.S., Kula, T., Geller, A.M., Talkington, M.W., Tang-Fernandez, V., Suciu, O., Demidenko, A.A., Ghabrial, S.A., Beach, D.H.*, et al.* (2011). Clinical isolates of Trichomonas vaginalis concurrently infected by strains of up to four Trichomonasvirus species (Family Totiviridae). J Virol *85*, 4258-4270.

Gould, P.S., Dyer, N.P., Croft, W., Ott, S., and Easton, A.J. (2014). Cellular mRNAs access second ORFs using a novel amino acid sequence-dependent coupled translation termination-reinitiation mechanism. RNA *20*, 373-381.

Gould, P.S., and Easton, A.J. (2005). Coupled translation of the respiratory syncytial virus M2 open reading frames requires upstream sequences. J Biol Chem *280*, 21972-21980.

Gould, P.S., and Easton, A.J. (2007). Coupled translation of the second open reading frame of M2 mRNA is sequence dependent and differs significantly within the subfamily Pneumovirinae. J Virol *81*, 8488-8496.

Gradi, A., Svitkin, Y. V., Imataka, H. and Sonenberg, N. (1998). Proteolysis of human eukaryotic translation initiation factor eIF4GII, but not eIF4GI, coincides with the shutoff of host protein synthesis after poliovirus infection. Proc Natl Acad Sci U S A 95, pp. 11089-94.

Grosfeld, H., Hill, M.G., and Collins, P.L. (1995). RNA replication by respiratory syncytial virus (RSV) is directed by the N, P, and L proteins; transcription also occurs under these conditions but requires RSV superinfection for efficient synthesis of full-length mRNA. J Virol *69*, 5677-5686.

Groskreutz, D. J., Babor, E. C., Monick, M. M., Varga, S. M. and Hunninghake, G. W. (2010). Respiratory syncytial virus limits alpha subunit of eukaryotic translation initiation factor 2 (eIF2alpha) phosphorylation to maintain translation and viral replication. J Biol Chem 285, pp. 24023-31.

Grzegorski, S.J., Chiari, E.F., Robbins, A., Kish, P.E., and Kahana, A. (2014). Natural variability of Kozak sequences correlates with function in a zebrafish model. PLoS One *9*, e108475.

Guo, L.H., Sun, L., Chiba, S., Araki, H., and Suzuki, N. (2009). Coupled termination/reinitiation for translation of the downstream open reading frame B of the prototypic hypovirus CHV1-EP713. Nucleic Acids Res *37*, 3645-3659.

Gussin, G.N. (1966). Three complementation groups in bacteriophage R1 (Journal of Molecular Biology), pp. 435-453.

Guvenel, A.K., Chiu, C., and Openshaw, P.J. (2014). Current concepts and progress in RSV vaccine development. Expert Rev Vaccines *13*, 333-344.

Hardy, R.W., and Wertz, G.W. (1998). The product of the respiratory syncytial virus M2 gene ORF1 enhances readthrough of intergenic junctions during viral transcription. J Virol *72*, 520-526.

Hastie, M.L., Headlam, M.J., Patel, N.B., Bukreyev, A.A., Buchholz, U.J., Dave, K.A., Norris, E.L., Wright, C.L., Spann, K.M., Collins, P.L.*, et al.* (2012). The human respiratory syncytial virus nonstructural protein 1 regulates type I and type II interferon pathways. Mol Cell Proteomics *11*, 108-127.

Haβ, M., Luttermann, C., and Meyers, G. (2014). Feline calicivirus can tolerate gross changes of its minor capsid protein expression levels induced by changing translation reinitiation frequency or use of a separate VP2-coding mRNA. PLoS One *9*, e102254.

Hellmuth, K., Rex, G., Surin, B., Zinck, R., and McCarthy, J.E. (1991). Translational coupling varying in efficiency between different pairs of genes in the central region of the atp operon of Escherichia coli. Mol Microbiol *5*, 813-824.

Hendricks, D.A., Baradaran, K., McIntosh, K., and Patterson, J.L. (1987). Appearance of a soluble form of the G protein of respiratory syncytial virus in fluids of infected cells. J Gen Virol *68 (Pt 6)*, 1705-1714.

Hodges, E.N., Heinrich, B.S., and Connor, J.H. (2012). A vesiculovirus showing a steepened transcription gradient and dominant trans-repression of virus transcription. J Virol *86*, 8884-8889.

Horiuchi, K., Lodish, H.F., and Zinder, N.D. (1966). Mutants of the bacteriophage f2. VI. Homology of temperature-sensitive and host-dependent mutants. Virology *28*, 438-447.

Horvath, C.M., Williams, M.A., and Lamb, R.A. (1990). Eukaryotic coupled translation of tandem cistrons: identification of the influenza B virus BM2 polypeptide. EMBO J *9*, 2639-2647.

Howard, M.T., Gesteland, R.F., and Atkins, J.F. (2004). Efficient stimulation of site-specific ribosome frameshifting by antisense oligonucleotides. RNA *10*, 1653-1661.

ICTV (2016). International Committee on Taxonomy of Viruses.

Ingolia, N.T. (2010). Genome-wide translational profiling by ribosome footprinting. Methods Enzymol *470*, 119-142.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nat Protoc *7*, 1534-1550.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2013). Genome-wide annotation and quantitation of translation by ribosome profiling. Curr Protoc Mol Biol *Chapter 4*, Unit 4.18.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218-223.

Ivanov, I.P., and Atkins, J.F. (2007). Ribosomal frameshifting in decoding antizyme mRNAs from yeast and protists to humans: close to 300 cases reveal remarkable diversity despite underlying conservation. Nucleic Acids Res *35*, 1842-1858.

Jaafar, Z.A., Oguro, A., Nakamura, Y., and Kieft, J.S. (2016). Translation initiation by the hepatitis C virus IRES requires eIF1A and ribosomal complex remodeling. Elife *5*.

Jacks, T., Madhani, H.D., Masiarz, F.R., and Varmus, H.E. (1988a). Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. Cell *55*, 447-458.

Jacks, T., Power, M.D., Masiarz, F.R., Luciw, P.A., Barr, P.J., and Varmus, H.E. (1988b). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. Nature *331*, 280-283.

Jacks, T., and Varmus, H.E. (1985). Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. Science *230*, 1237-1242.

Jackson, R.J., Hellen, C.U., and Pestova, T.V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol *11*, 113-127.

Jang, S.K., Kräusslich, H.G., Nicklin, M.J., Duke, G.M., Palmenberg, A.C., and Wimmer, E. (1988). A segment of the 5' nontranslated region of

encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. J Virol *62*, 2636-2643.

Jayakar, H.R., and Whitt, M.A. (2002). Identification of two additional translation products from the matrix (M) gene that contribute to vesicular stomatitis virus cytopathology. J Virol *76*, 8011-8018.

Johnson, P.R., and Collins, P.L. (1989). The 1B (NS2), 1C (NS1) and N proteins of human respiratory syncytial virus (RSV) of antigenic subgroups A and B: sequence conservation and divergence within RSV genomic RNA. J Gen Virol *70 (Pt 6)*, 1539-1547.

Kang, S.T., Leu, J.H., Wang, H.C., Chen, L.L., Kou, G.H., and Lo, C.F. (2009). Polycistronic mRNAs and internal ribosome entry site elements (IRES) are widely used by white spot syndrome virus (WSSV) structural protein genes. Virology *387*, 353-363.

Khong, A., Bonderoff, J.M., Spriggs, R.V., Tammpere, E., Kerr, C.H., Jackson, T.J., Willis, A.E., and Jan, E. (2016). Temporal Regulation of Distinct Internal Ribosome Entry Sites of the Dicistroviridae Cricket Paralysis Virus. Viruses *8*.

Kieft, J.S. (2008). Viral IRES RNA structures and ribosome interactions. Trends Biochem Sci *33*, 274-283.

Kolupaeva, V.G., Lomakin, I.B., Pestova, T.V., and Hellen, C.U. (2003). Eukaryotic initiation factors 4G and 4A mediate conformational changes downstream of the initiation codon of the encephalomyocarditis virus internal ribosomal entry site. Mol Cell Biol *23*, 687-698.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell *44*, 283-292.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res *15*, 8125-8148.

Kozak, M. (1997). Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. EMBO J *16*, 2482-2492.

Kozak, M. (2002). Pushing the limits of the scanning mechanism for initiation of translation. Gene *299*, 1-34.

Krempl, C., Murphy, B.R., and Collins, P.L. (2002). Recombinant respiratory syncytial virus with the G and F genes shifted to the promoter-proximal positions. J Virol *76*, 11931-11942.

Krempl, C.D., Lamirande, E.W., and Collins, P.L. (2005). Complete sequence of the RNA genome of pneumonia virus of mice (PVM). Virus Genes *30*, 237-249.

Krzyzaniak, M.A., Zumstein, M.T., Gerez, J.A., Picotti, P., and Helenius, A. (2013). Host cell entry of respiratory syncytial virus involves macropinocytosis followed by proteolytic activation of the F protein. PLoS Pathog *9*, e1003309.

Kwilas, S., Liesman, R.M., Zhang, L., Walsh, E., Pickles, R.J., and Peeples, M.E. (2009). Respiratory syncytial virus grown in Vero cells contains a truncated attachment protein that alters its infectivity and dependence on glycosaminoglycans. J Virol *83*, 10710-10718.

Latorre, P., Kolakofsky, D., and Curran, J. (1998). Sendai virus Y proteins are initiated by a ribosomal shunt. Mol Cell Biol *18*, 5021-5031.

Lee, W.J., Kim, Y.J., Kim, D.W., Lee, H.S., Lee, H.Y., and Kim, K. (2012). Complete genome sequence of human respiratory syncytial virus genotype A with a 72-nucleotide duplication in the attachment protein G gene. J Virol *86*, 13810-13811.

Li, H., Havens, W.M., Nibert, M.L., and Ghabrial, S.A. (2011). RNA sequence determinants of a coupled termination-reinitiation strategy for downstream open reading frame translation in Helminthosporium victoriae virus 190S and other victoriviruses (Family Totiviridae). J Virol *85*, 7343-7352.

Li, H., Havens, W.M., Nibert, M.L., and Ghabrial, S.A. (2015a). An RNA cassette from Helminthosporium victoriae virus 190S necessary and sufficient for stop/restart translation. Virology *474*, 131-143.

Li, Y., Jain, N., Limpanawat, S., To, J., Quistgaard, E.M., Nordlund, P., Thanabalu, T., and Torres, J. (2015b). Interaction between human BAP31 and respiratory syncytial virus small hydrophobic (SH) protein. Virology *482*, 105-110.

Licis, N., van Duin, J., Balklava, Z., and Berzins, V. (1998). Long-range translational coupling in single-stranded RNA bacteriophages: an evolutionary analysis. Nucleic Acids Res *26*, 3242-3246.

Liljeroos, L., Krzyzaniak, M.A., Helenius, A., and Butcher, S.J. (2013). Architecture of respiratory syncytial virus revealed by electron cryotomography. Proc Natl Acad Sci U S A *110*, 11133-11138.

Lin, C.G., and Lo, S.J. (1992). Evidence for involvement of a ribosomal leaky scanning mechanism in the translation of the hepatitis B virus pol gene from the viral pregenome RNA. Virology *188*, 342-352.

Liston, P., and Briedis, D.J. (1995). Ribosomal frameshifting during translation of measles virus P protein mRNA is capable of directing synthesis of a unique protein. J Virol *69*, 6742-6750.

Liuzzi, M., Mason, S.W., Cartier, M., Lawetz, C., McCollum, R.S., Dansereau, N., Bolger, G., Lapeyre, N., Gaudette, Y., Lagacé, L*., et al.* (2005). Inhibitors of respiratory syncytial virus replication target cotranscriptional mRNA guanylylation by viral RNA-dependent RNA polymerase. J Virol *79*, 13105-13115.

Llorente, M.T., García-Barreno, B., Calero, M., Camafeita, E., López, J.A., Longhi, S., Ferrón, F., Varela, P.F., and Melero, J.A. (2006). Structural analysis of the human respiratory syncytial virus phosphoprotein: characterization of an alpha-helical domain involved in oligomerization. J Gen Virol *87*, 159-169.

Lo, M.S., Brazas, R.M., and Holtzman, M.J. (2005). Respiratory syncytial virus nonstructural proteins NS1 and NS2 mediate inhibition of Stat2 expression and alpha/beta interferon responsiveness. J Virol *79*, 9315-9319.

Low, K.W., Tan, T., Ng, K., Tan, B.H., and Sugrue, R.J. (2008). The RSV F and G glycoproteins interact to form a complex on the surface of infected cells. Biochem Biophys Res Commun *366*, 308-313.

Lu, B., Ma, C.H., Brazas, R., and Jin, H. (2002). The major phosphorylation sites of the respiratory syncytial virus phosphoprotein are dispensable for virus replication in vitro. J Virol *76*, 10776-10784.

Luttermann, C., and Meyers, G. (2007). A bipartite sequence motif induces translation reinitiation in feline calicivirus RNA. J Biol Chem *282*, 7056-7065.

Luttermann, C., and Meyers, G. (2009). The importance of inter- and intramolecular base pairing for translation reinitiation on a eukaryotic bicistronic mRNA. Genes Dev *23*, 331-344.

Luttermann, C., and Meyers, G. (2014). Two alternative ways of start site selection in human norovirus reinitiation of translation. J Biol Chem *289*, 11739-11754.

Léger, M., Dulude, D., Steinberg, S.V., and Brakier-Gingras, L. (2007). The three transfer RNAs occupying the A, P and E sites on the ribosome are involved in viral programmed -1 ribosomal frameshift. Nucleic Acids Res *35*, 5581-5592.

Maclellan, K., Loney, C., Yeo, R.P., and Bhella, D. (2007). The 24-angstrom structure of respiratory syncytial virus nucleocapsid protein-RNA decameric rings. J Virol *81*, 9519-9524.

Mari, J., Poulos, B.T., Lightner, D.V., and Bonami, J.R. (2002). Shrimp Taura syndrome virus: genomic characterization and similarity with members of the genus Cricket paralysis-like viruses. J Gen Virol *83*, 915-926.

Martin, F., Ménétret, J.F., Simonetti, A., Myasnikov, A.G., Vicens, Q., Prongidi-Fix, L., Natchiar, S.K., Klaholz, B.P., and Eriani, G. (2016). Ribosomal 18S rRNA base pairs with mRNA during eukaryotic translation initiation. Nat Commun *7*, 12622.

Martínez-Azorín, F., Remacha, M., Martínez-Salas, E., and Ballesta, J.P. (2008). Internal translation initiation on the foot-and-mouth disease virus IRES is affected by ribosomal stalk conformation. FEBS Lett *582*, 3029-3032.

Martínez-Salas, E., Francisco-Velilla, R., Fernandez-Chamorro, J., Lozano, G., and Diaz-Toledano, R. (2015). Picornavirus IRES elements: RNA structure and host protein interactions. Virus Res *206*, 62-73.

Mason, S.W., Aberg, E., Lawetz, C., DeLong, R., Whitehead, P., and Liuzzi, M. (2003). Interaction between human respiratory syncytial virus (RSV) M2-1 and P proteins is required for reconstitution of M2-1-dependent RSV minigenome activity. J Virol *77*, 10670-10676.

Mastrangelo, P., and Hegele, R.G. (2013). RSV fusion: time for a new model. Viruses *5*, 873-885.

Matsufuji, S., Matsufuji, T., Wills, N.M., Gesteland, R.F., and Atkins, J.F. (1996). Reading two bases twice: mammalian antizyme frameshifting in yeast. EMBO J *15*, 1360-1370.

McCarthy, A.J., and Goodman, S.J. (2010). Reassessing conflicting evolutionary histories of the Paramyxoviridae and the origins of respiroviruses with Bayesian multigene phylogenies. Infect Genet Evol *10*, 97-107.

McCormick, C.J., Salim, O., Lambden, P.R., and Clarke, I.N. (2008). Translation termination reinitiation between open reading frame 1 (ORF1) and ORF2 enables capsid expression in a bovine norovirus without the need for production of viral subgenomic RNA. J Virol *82*, 8917-8921.

McLellan, J.S., Chen, M., Leung, S., Graepel, K.W., Du, X., Yang, Y., Zhou, T., Baxa, U., Yasuda, E., Beaumont, T.*, et al.* (2013a). Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. Science *340*, 1113-1117.

McLellan, J.S., Ray, W.C., and Peeples, M.E. (2013b). Structure and function of respiratory syncytial virus surface glycoproteins. Curr Top Microbiol Immunol *372*, 83-104.

McNamara, P.S., and Smyth, R.L. (2002). The pathogenesis of respiratory syncytial virus disease in childhood. Br Med Bull *61*, 13-28.

Melero, J.A., and Mas, V. (2015). The Pneumovirinae fusion (F) protein: A common target for vaccines and antivirals. Virus Res *209*, 128-135.

Meyers, G. (2003). Translation of the minor capsid protein of a calicivirus is initiated by a novel termination-dependent reinitiation mechanism. J Biol Chem *278*, 34051-34060.

Meyers, G. (2007). Characterization of the sequence element directing translation reinitiation in RNA of the calicivirus rabbit hemorrhagic disease virus. J Virol *81*, 9623-9632.

Mink, M.A., Stec, D.S., and Collins, P.L. (1991). Nucleotide sequences of the 3' leader and 5' trailer regions of human respiratory syncytial virus genomic RNA. Virology *185*, 615-624.

Mitra, R., Baviskar, P., Duncan-Decocq, R.R., Patel, D., and Oomens, A.G. (2012). The human respiratory syncytial virus matrix protein is required for maturation of viral filaments. J Virol *86*, 4432-4443.

Money, V.A., McPhee, H.K., Mosely, J.A., Sanderson, J.M., and Yeo, R.P. (2009). Surface features of a Mononegavirales matrix protein indicate sites of membrane interaction. Proc Natl Acad Sci U S A *106*, 4441-4446.

Morin, B., Kranzusch, P.J., Rahmeh, A.A., and Whelan, S.P. (2013). The polymerase of negative-stranded RNA viruses. Curr Opin Virol *3*, 103-110.

Nair, H., Nokes, D.J., Gessner, B.D., Dherani, M., Madhi, S.A., Singleton, R.J., O'Brien, K.L., Roca, A., Wright, P.F., Bruce, N.*, et al.* (2010). Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. Lancet *375*, 1545-1555.

Nakagawa, S., Niimura, Y., Gojobori, T., Tanaka, H., and Miura, K. (2008). Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. Nucleic Acids Res *36*, 861-871.

Namy, O., Moran, S.J., Stuart, D.I., Gilbert, R.J., and Brierley, I. (2006). A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. Nature *441*, 244-247.

Napthine, S., Lever, R.A., Powell, M.L., Jackson, R.J., Brown, T.D., and Brierley, I. (2009). Expression of the VP2 protein of murine norovirus by a translation termination-reinitiation strategy. PLoS One *4*, e8390.

Nikolaitchik, O.A., and Hu, W.S. (2014). Deciphering the role of the Gag-Pol ribosomal frameshift signal in HIV-1 RNA genome packaging. J Virol *88*, 4040-4046.

Noton, S.L., Cowton, V.M., Zack, C.R., McGivern, D.R., and Fearns, R. (2010). Evidence that the polymerase of respiratory syncytial virus initiates RNA replication in a nontemplated fashion. Proc Natl Acad Sci U S A *107*, 10226-10231.

Noton, S.L., Deflubé, L.R., Tremaglio, C.Z., and Fearns, R. (2012). The respiratory syncytial virus polymerase has multiple RNA synthesis activities at the promoter. PLoS Pathog *8*, e1002980.

Ogawa, A. (2013). Ligand-dependent upregulation of ribosomal shunting. Chembiochem *14*, 1539-1543, 1509.

Olivier, V., Blanchard, P., Chaouch, S., Lallemand, P., Schurr, F., Celle, O., Dubois, E., Tordo, N., Thiéry, R., Houlgatte, R.*, et al.* (2008). Molecular characterisation and phylogenetic analysis of Chronic bee paralysis virus, a honey bee virus. Virus Res *132*, 59-68.

Oppenheim, D.S., and Yanofsky, C. (1980). Translational coupling during expression of the tryptophan operon of Escherichia coli. Genetics *95*, 785-795.

Oshansky, C.M., Zhang, W., Moore, E., and Tripp, R.A. (2009). The host response and molecular pathogenesis associated with respiratory syncytial virus infection. Future Microbiol *4*, 279-297.

Palanimurugan, R., Scheel, H., Hofmann, K., and Dohmen, R.J. (2004). Polyamines regulate their synthesis by inducing expression and blocking degradation of ODC antizyme. EMBO J *23*, 4857-4867.

Pelletier, J., and Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Nature *334*, 320-325.

Pennica, D., Lynch, K.R., Cohen, P.S., and Ennis, H.L. (1979). Decay of vesicular stomatitis virus mRNAs in vivo. Virology *94*, 484-487.

Pérard, J., Leyrat, C., Baudin, F., Drouet, E., and Jamin, M. (2013). Structure of the full-length HCV IRES in solution. Nat Commun *4*, 1612.

Pestova, T.V., Hellen, C.U., and Shatsky, I.N. (1996). Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. Mol Cell Biol *16*, 6859-6869.

Petrov, A., Grosely, R., Chen, J., O'Leary, S.E., and Puglisi, J.D. (2016). Multiple Parallel Pathways of Translation Initiation on the CrPV IRES. Mol Cell *62*, 92-103.

Pisarev, A.V., Chard, L.S., Kaku, Y., Johns, H.L., Shatsky, I.N., and Belsham, G.J. (2004). Functional and structural similarities between the internal ribosome entry sites of hepatitis C virus and porcine teschovirus, a picornavirus. J Virol *78*, 4487-4497.

Plant, E.P., and Dinman, J.D. (2006). Comparative study of the effects of heptameric slippery site composition on -1 frameshifting among different eukaryotic systems. RNA *12*, 666-673.

Plant, E.P., Jacobs, K.L., Harger, J.W., Meskauskas, A., Jacobs, J.L., Baxter, J.L., Petrov, A.N., and Dinman, J.D. (2003). The 9-A solution: how mRNA pseudoknots promote efficient programmed -1 ribosomal frameshifting. RNA *9*, 168-174.

Plant, E.P., Rakauskaite, R., Taylor, D.R., and Dinman, J.D. (2010). Achieving a golden mean: mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. J Virol *84*, 4330-4340.

Plumet, S., Duprex, W.P., and Gerlier, D. (2005). Dynamics of viral RNA synthesis during measles virus infection. J Virol *79*, 6900-6908.

Pooggin, M.M., Rajeswaran, R., Schepetilnikov, M.V., and Ryabova, L.A. (2012). Short ORF-dependent ribosome shunting operates in an RNA picorna-like virus and a DNA pararetrovirus that cause rice tungro disease. PLoS Pathog *8*, e1002568.

Pooggin, M.M., Ryabova, L.A., He, X., Fütterer, J., and Hohn, T. (2006). Mechanism of ribosome shunting in Rice tungro bacilliform pararetrovirus. RNA *12*, 841-850.

Powell, M.L., Leigh, K.E., Pöyry, T.A., Jackson, R.J., Brown, T.D., and Brierley, I. (2011). Further characterisation of the translational termination-reinitiation signal of the influenza B virus segment 7 RNA. PLoS One *6*, e16822.

Powell, M.L., Napthine, S., Jackson, R.J., Brierley, I., and Brown, T.D. (2008). Characterization of the termination-reinitiation strategy employed in the expression of influenza B virus BM2 protein. RNA *14*, 2394-2406.

Pöyry, T.A., Kaminski, A., Connell, E.J., Fraser, C.S., and Jackson, R.J. (2007). The mechanism of an exceptional case of reinitiation after translation of a long ORF reveals why such events do not generally occur in mammalian mRNA translation. Genes Dev *21*, 3149-3162.

Racine, T., Barry, C., Roy, K., Dawe, S.J., Shmulevitz, M., and Duncan, R. (2007). Leaky scanning and scanning-independent ribosome migration on the tricistronic S1 mRNA of avian reovirus. J Biol Chem *282*, 25613-25622.

Racine, T., and Duncan, R. (2010). Facilitated leaky scanning and atypical ribosome shunting direct downstream translation initiation on the tricistronic S1 mRNA of avian reovirus. Nucleic Acids Res *38*, 7260-7272.

Ray, P. S., Grover, R. and Das, S. (2006). Two internal ribosome entry sites mediate the translation of p53 isoforms. EMBO Rep 7, pp. 404-10.

Ren, Q., Wang, Q.S., Firth, A.E., Chan, M.M., Gouw, J.W., Guarna, M.M., Foster, L.J., Atkins, J.F., and Jan, E. (2012). Alternative reading frame selection mediated by

a tRNA-like domain of an internal ribosome entry site. Proc Natl Acad Sci U S A *109*, E630-639.

Rex, G., Surin, B., Besse, G., Schneppe, B., and McCarthy, J.E. (1994). The mechanism of translational coupling in Escherichia coli. Higher order structure in the atpHA mRNA acts as a conformational switch regulating the access of de novo initiating ribosomes. J Biol Chem *269*, 18118-18127.

Rijnbrand, R., van der Straaten, T., van Rijn, P.A., Spaan, W.J., and Bredenbeek, P.J. (1997). Internal entry of ribosomes is directed by the 5' noncoding region of classical swine fever virus and is dependent on the presence of an RNA pseudoknot upstream of the initiation codon. J Virol *71*, 451-457.

Rixon, H.W., Brown, G., Murray, J.T., and Sugrue, R.J. (2005). The respiratory syncytial virus small hydrophobic protein is phosphorylated via a mitogen-activated protein kinase p38-dependent tyrosine kinase activity during virus infection. J Gen Virol *86*, 375-384.

Rosenberg, H.F., and Domachowske, J.B. (2008). Pneumonia virus of mice: severe respiratory infection in a natural host. Immunol Lett *118*, 6-12.

Rudraraju, R., Jones, B.G., Sealy, R., Surman, S.L., and Hurwitz, J.L. (2013). Respiratory syncytial virus: current progress in vaccine development. Viruses *5*, 577-594.

Ryabova, L.A., and Hohn, T. (2000). Ribosome shunting in the cauliflower mosaic virus 35S RNA leader is a special case of reinitiation of translation functioning in plant and animal systems. Genes Dev *14*, 817-829.

Sacco, R.E., McGill, J.L., Pillatzki, A.E., Palmer, M.V., and Ackermann, M.R. (2014). Respiratory syncytial virus infection in cattle. Vet Pathol *51*, 427-436.

Sasaki, J., and Nakashima, N. (1999). Translation initiation at the CUU codon is mediated by the internal ribosome entry site of an insect picorna-like virus in vitro. J Virol *73*, 1219-1226.

Sato, H., Masuda, M., Kanai, M., Tsukiyama-Kohara, K., Yoneda, M. and Kai, C. (2007). Measles virus N protein inhibits host translation by binding to eIF3-p40. J Virol 81, pp. 11569-76.

Schneider-Poetsch, T., Ju, J., Eyler, D.E., Dang, Y., Bhat, S., Merrick, W.C., Green, R., Shen, B., and Liu, J.O. (2010). Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. Nat Chem Biol *6*, 209-217.

Schümperli, D., McKenney, K., Sobieski, D.A., and Rosenberg, M. (1982). Translational coupling at an intercistronic boundary of the Escherichia coli galactose operon. Cell *30*, 865-871.

Shaikh, F.Y., and Crowe, J.E. (2013). Molecular mechanisms driving respiratory syncytial virus assembly. Future Microbiol *8*, 123-131.

Shaikh, F.Y., Utley, T.J., Craven, R.E., Rogers, M.C., Lapierre, L.A., Goldenring, J.R., and Crowe, J.E. (2012). Respiratory syncytial virus assembles into structured filamentous virion particles independently of host cytoskeleton and related proteins. PLoS One *7*, e40826.

Simoes, E.A. (1999). Respiratory syncytial virus infection. Lancet *354*, 847-852.

Soto-Rifo, R., Rubilar, P.S., Limousin, T., de Breyne, S., Décimo, D., and Ohlmann, T. (2012). DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. EMBO J *31*, 3745-3756.

Spann, K.M., Tran, K.C., Chi, B., Rabin, R.L., and Collins, P.L. (2004). Suppression of the induction of alpha, beta, and lambda interferons by the NS1 and NS2 proteins of human respiratory syncytial virus in human epithelial cells and macrophages [corrected]. J Virol *78*, 4363-4369.

Spotts, G.D., Patel, S.V., Xiao, Q., and Hann, S.R. (1997). Identification of downstream-initiated c-Myc proteins which are dominant-negative inhibitors of transactivation by full-length c-Myc proteins. Mol Cell Biol *17*, 1459-1468.

Spriggs, K.A., Cobbold, L.C., Jopling, C.L., Cooper, R.E., Wilson, L.A., Stoneley, M., Coldwell, M.J., Poncet, D., Shen, Y.C., Morley, S.J.*, et al.* (2009). Canonical

initiation factor requirements of the Myc family of internal ribosome entry segments. Mol Cell Biol *29*, 1565-1574.

Steitz, J.A. (1969). Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA. Nature *224*, 957-964.

Sweeney, T.R., Abaeva, I.S., Pestova, T.V., and Hellen, C.U. (2014). The mechanism of translation initiation on Type 1 picornavirus IRESs. EMBO J *33*, 76-92.

Tayyari, F., Marchant, D., Moraes, T.J., Duan, W., Mastrangelo, P., and Hegele, R.G. (2011). Identification of nucleolin as a cellular receptor for human respiratory syncytial virus. Nat Med *17*, 1132-1135.

Techaarpornkul, S., Barretto, N., and Peeples, M.E. (2001). Functional analysis of recombinant respiratory syncytial virus deletion mutants lacking the small hydrophobic and/or attachment glycoprotein gene. J Virol *75*, 6825-6834.

Techaarpornkul, S., Collins, P.L., and Peeples, M.E. (2002). Respiratory syncytial virus with the fusion protein as its only viral glycoprotein is less dependent on cellular glycosaminoglycans for attachment than complete virus. Virology *294*, 296-304.

Teng, M.N., and Collins, P.L. (2002). The central conserved cystine noose of the attachment G protein of human respiratory syncytial virus is not required for efficient viral infection in vitro or in vivo. J Virol *76*, 6164-6171.

Thiel, V., Ivanov, K.A., Putics, A., Hertzig, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E.J., Rabenau, H., Doerr, H.W.*, et al.* (2003). Mechanisms and enzymes involved in SARS coronavirus genome expression. J Gen Virol *84*, 2305-2315.

Thomas, L.H., Cook, R.S., Wyld, S.G., Furze, J.M., and Taylor, G. (1998). Passive protection of gnotobiotic calves using monoclonal antibodies directed at different epitopes on the fusion protein of bovine respiratory syncytial virus. J Infect Dis *177*, 874-880.

Thomas, L.H., Gourlay, R.N., Stott, E.J., Howard, C.J., and Bridger, J.C. (1982). A search for new microorganisms in calf pneumonia by the inoculation of gnotobiotic calves. Res Vet Sci *33*, 170-182.

Thomas, M.S., Bedwell, D.M., and Nomura, M. (1987). Regulation of alpha operon gene expression in Escherichia coli. A novel form of translational coupling. J Mol Biol *196*, 333-345.

Tikole, S., and Sankararamakrishnan, R. (2008). Prediction of translation initiation sites in human mRNA sequences with AUG start codon in weak Kozak context: A neural network approach. Biochem Biophys Res Commun *369*, 1166-1168.

Tran, K.C., Collins, P.L., and Teng, M.N. (2004). Effects of altering the transcription termination signals of respiratory syncytial virus on viral gene expression and growth in vitro and in vivo. J Virol *78*, 692-699.

Tran, T.L., Castagné, N., Dubosclard, V., Noinville, S., Koch, E., Moudjou, M., Henry, C., Bernard, J., Yeo, R.P., and Eléouët, J.F. (2009). The respiratory syncytial virus M2-1 protein forms tetramers and interacts with RNA and P in a competitive manner. J Virol *83*, 6363-6374.

Tremaglio, C.Z., Noton, S.L., Deflubé, L.R., and Fearns, R. (2013). Respiratory syncytial virus polymerase can initiate transcription from position 3 of the leader promoter. J Virol *87*, 3196-3207.

Tsukiyama-Kohara, K., Iizuka, N., Kohara, M., and Nomoto, A. (1992). Internal ribosome entry site within hepatitis C virus RNA. J Virol *66*, 1476-1483.

Tycowski, K.T., Shu, M.D., Borah, S., Shi, M., and Steitz, J.A. (2012). Conservation of a triple-helix-forming RNA stability element in noncoding and genomic RNAs of diverse viruses. Cell Rep *2*, 26-32.

Utley, T.J., Ducharme, N.A., Varthakavi, V., Shepherd, B.E., Santangelo, P.J., Lindquist, M.E., Goldenring, J.R., and Crowe, J.E. (2008). Respiratory syncytial virus uses a Vps4-independent budding mechanism controlled by Rab11-FIP2. Proc Natl Acad Sci U S A *105*, 10209-10214.

Vera-Otarola, J., Solis, L., Soto-Rifo, R., Ricci, E.P., Pino, K., Tischler, N.D., Ohlmann, T., Darlix, J.L., and López-Lastra, M. (2012). The Andes hantavirus NSs protein is expressed from the viral small mRNA by a leaky scanning mechanism. J Virol *86*, 2176-2187.

Wilson, J.E., Pestova, T.V., Hellen, C.U., and Sarnow, P. (2000a). Initiation of protein synthesis from the A site of the ribosome. Cell *102*, 511-520.

Wilson, J.E., Powell, M.J., Hoover, S.E., and Sarnow, P. (2000b). Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. Mol Cell Biol *20*, 4990-4999.

Wise, H.M., Foeglein, A., Sun, J., Dalton, R.M., Patel, S., Howard, W., Anderson, E.C., Barclay, W.S., and Digard, P. (2009). A complicated message: Identification of a novel PB1-related protein translated from influenza A virus segment 2 mRNA. J Virol *83*, 8021-8031.

Wynne, J.W., Shiell, B.J., Marsh, G.A., Boyd, V., Harper, J.A., Heesom, K., Monaghan, P., Zhou, P., Payne, J., Klein, R.*, et al.* (2014). Proteomics informed by transcriptomics reveals Hendra virus sensitizes bat cells to TRAIL-mediated apoptosis. Genome Biol *15*, 532.

Yang, E., van Nimwegen, E., Zavolan, M., Rajewsky, N., Schroeder, M., Magnasco, M., and Darnell, J.E. (2003). Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. Genome Res *13*, 1863-1872.

Yang, P., Zheng, J., Wang, S., Liu, P., Xie, M., and Zhao, D. (2015). Respiratory syncytial virus nonstructural proteins 1 and 2 are crucial pathogenic factors that modulate interferon signaling and Treg cell distribution in mice. Virology *485*, 223-232.

Yu, C.H., Teulade-Fichou, M.P., and Olsthoorn, R.C. (2014). Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. Nucleic Acids Res *42*, 1887-1892.

Yueh, A., and Schneider, R.J. (2000). Translation by ribosome shunting on adenovirus and hsp70 mRNAs facilitated by complementarity to 18S rRNA. Genes Dev *14*, 414-421.

Zamora, M., and Samal, S.K. (1992). Gene junction sequences of bovine respiratory syncytial virus. Virus Res *24*, 115-121.

Zinoviev, A., Hellen, C.U., and Pestova, T.V. (2015). Multiple mechanisms of reinitiation on bicistronic calicivirus mRNAs. Mol Cell *57*, 1059-1073.

Appendix A

Commands for scripts used to process raw next generation sequencing data generated from ribosomal profiling.

Scripts used to process raw sequencing data from libraries built using the technique of ribosomal profiling. Scripts are numbered in order of use. File containing raw sequencing data is initially downloaded from BaseSpace prior to being processed using Fastx_clipper. File when downloaded is in fastq format.

1. Clipper to remove sequence of 5´ primer
Script: Fastx_clipper

./fastx_clipper -a AGATCGGAAGAGCACACGTCT -l 25 -c -n -v -Q33 -i P4.fastq -o P4clipped.fastq

A Script used. Must have ./ to designate that file is in folder
A Sequence that will be removed
-I –i= input -o =output file. Must have – before
A Input and output file name must go after –i/-o respectively
25 Minimum length of read recommended by illumina

2. Trimmer to trim unwanted length of sequence
Script: Fastx_trimmer

./fastx_trimmer -Q33 -f 1 -i P4clipped.fastq >P4trimmed.fastq

A Script used. Must have ./ to designate that scirpt is in folder
-I –i= input >=output file
A Input and output file name must go after –i/> respectively

3. Building index genome
Scrip: Bowtie-build

First download genome from genbank in FASTA format. Rename and place in designated folder.

./bowtie-build bRSV.fna bRSV

A Script used. Must have ./ to designate that file is in folder
A Name of input genome file in FASTA format
A Output file name

4. Bowtie alignment
Script Bowtie

To align sample ensure trimmed file is in the same folder/directory

./bowtie -S brsvindex P4trimmed.fastq P4.sam
t
A Script used. Must have ./ to designate that file is in folder
A Name of genome in index form, only use the file name that is used in bowtie build, the programme will use all the files produced
A Input file name

A Output file name ensure the file name ends in .sam

5. Turn aligned file into compatible file for IGV
Script Samtools

IGV, the viewing tool will only recognise a certain file type with a correct index attached.

First .sam file needs to be converted into .bam file

./samtools view -bS -o P4.bam P4.sam

A Script used. Must have ./ to designate that file is in folder
A –o output code
A Output file must have .bam at end and must have –o before
A Input file name from bowtie alignment

Next turn .bam file into sorted file

./samtools sort -o P4.sorted P4.bam
A Script used. Must have ./ to designate that file is in folder
A –o output code
A Output file, must have .sorted at end and must have –o before
A Input file name with .bam

Finally an index must be created. This is for IGV to recognise the sorted file
./samtools index -b P4.sorted P4.bam.bai

A Script used. Must have ./ to designate that file is in folder
A Input file must have .sorted at end must have –o before
A Output file name with .bam.bai The file must also have the same name before as the sorted file.

In folder change the file names to the following:
FileName.sorted :change to: FileName.bam
FileName.bam :change to: FileNameunsorted.bam

This is to ensure IGV does not recognize the unsorted version.

6. Loading alignment in IGV
First, load genome in the FASTA format.
Next load aligned file. Choose the .sorted file that is renamed to .bam.

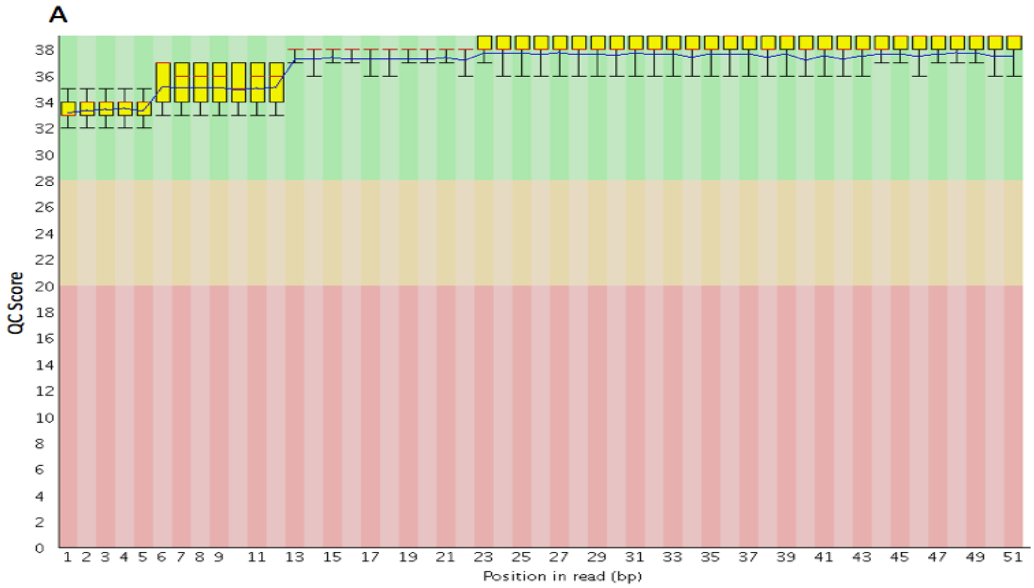
7. WIG files
Script WIG

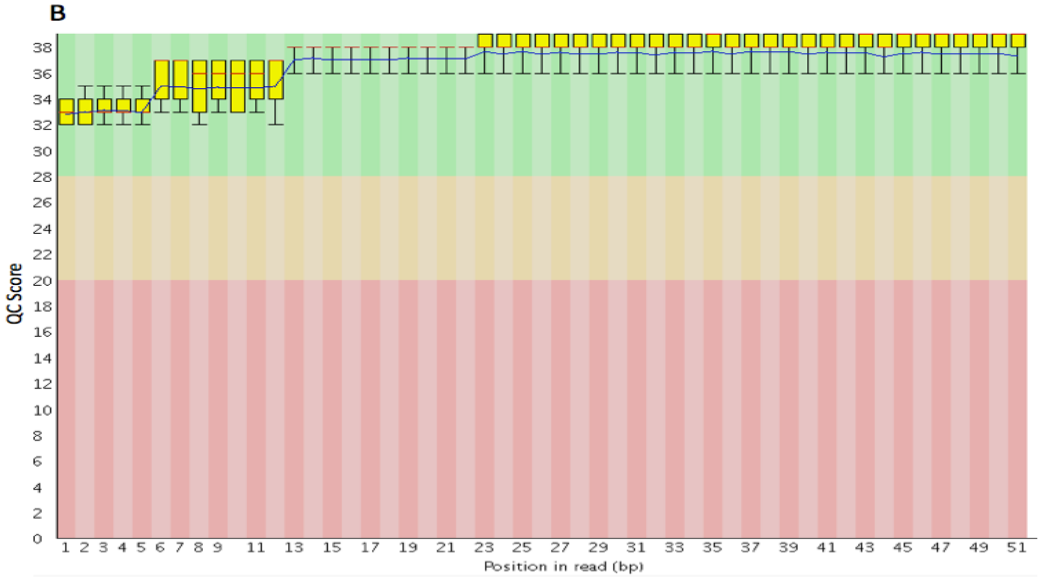
Used to view location of first nucleotide of each read aligned to genome
./orfAnalysis2 -w -p 0.01 –pt 100 -g brsv.rna_first -b P8.bam
A Script used. Must have ./ to designate to look in that file

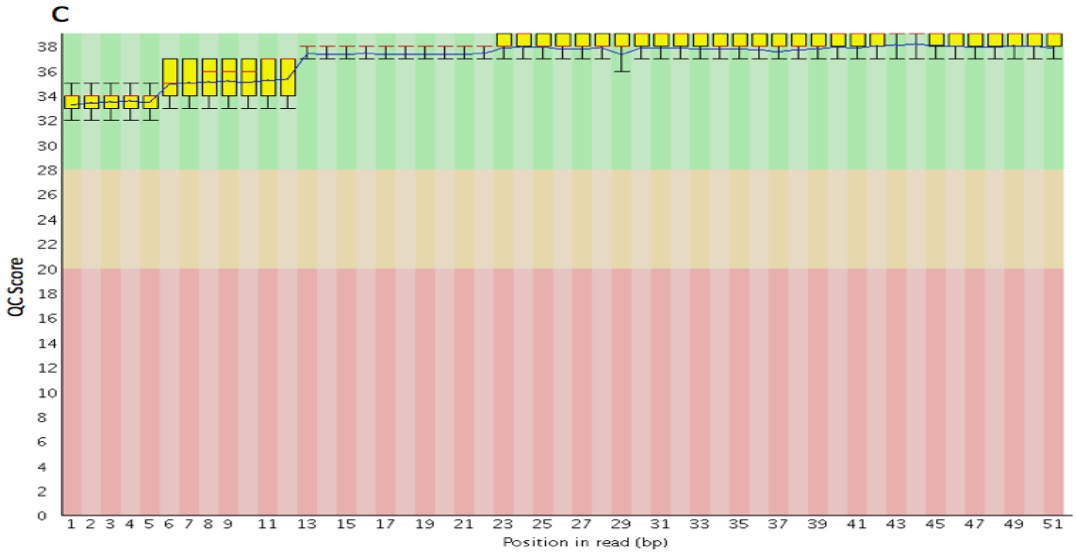
A Psuedo count, set between 0.01 and 0.3. Recommended to use 0.01
A Output file that is used to compare to genome. Ensure just the file name ends in .info
-b Always put before input file
A Input file should be the sorted version, that has been renamed to end in .bam.

Appendix B

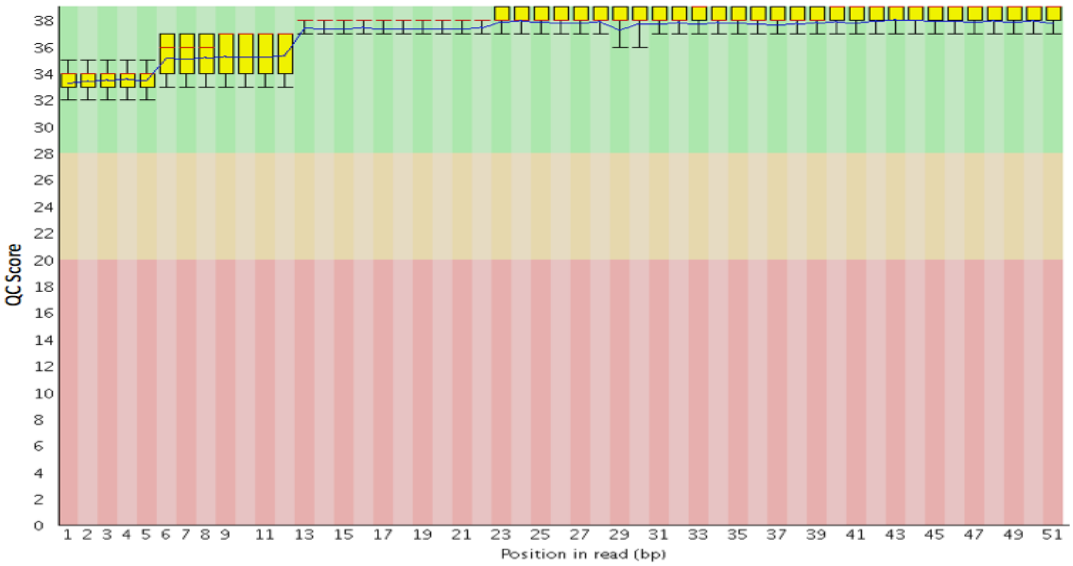
Additional figures.


A
QC Score
Position in read (bp)

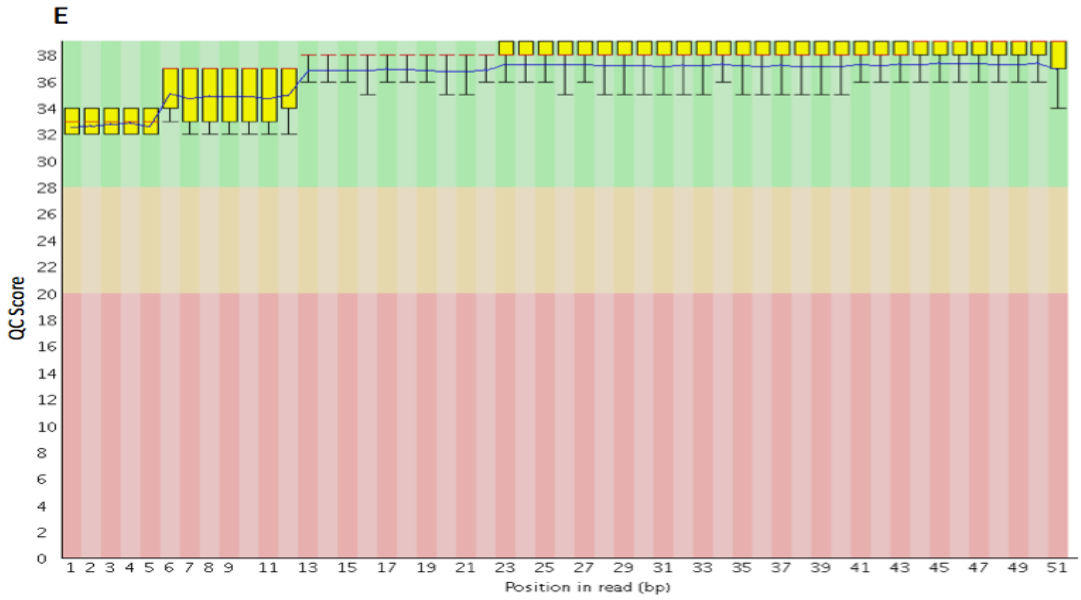

B
QC Score
Position in read (bp)


C
QC Score
Position in read (bp)

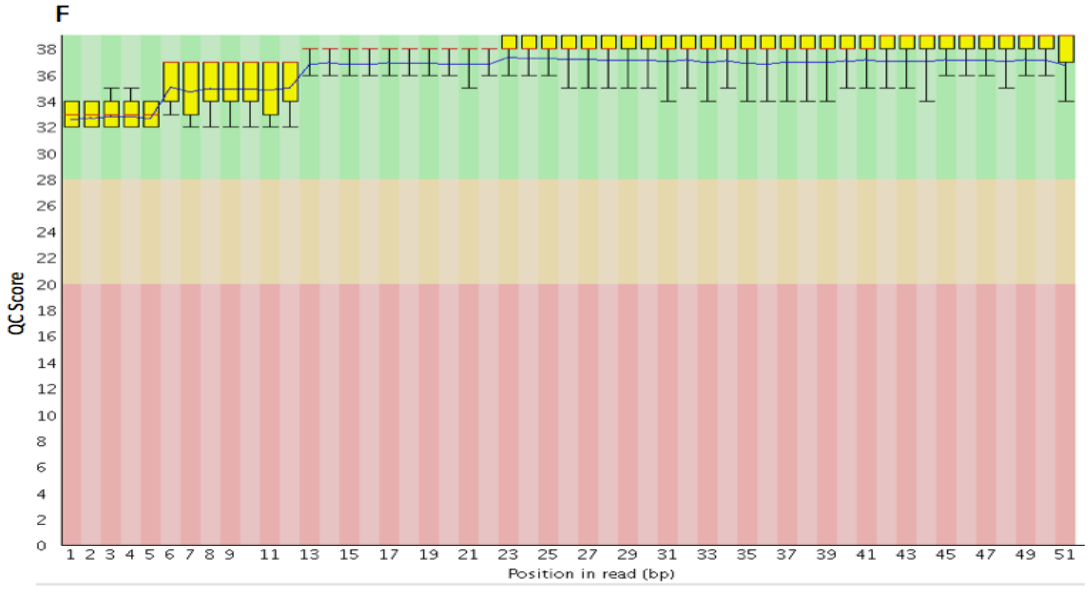
D

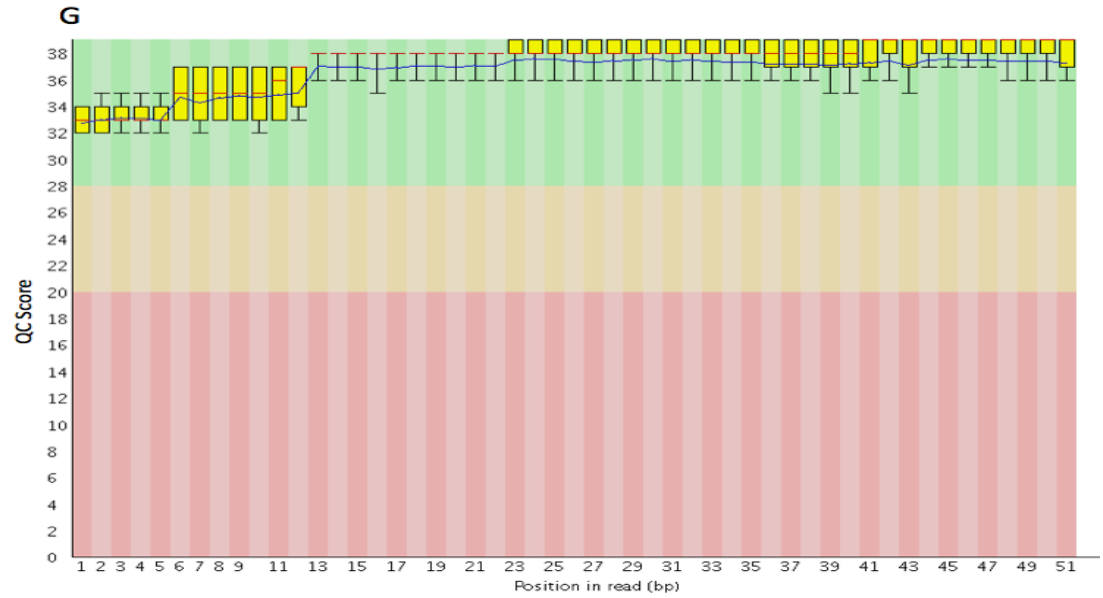

QC Score
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38
1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51
Position in read (bp)

E

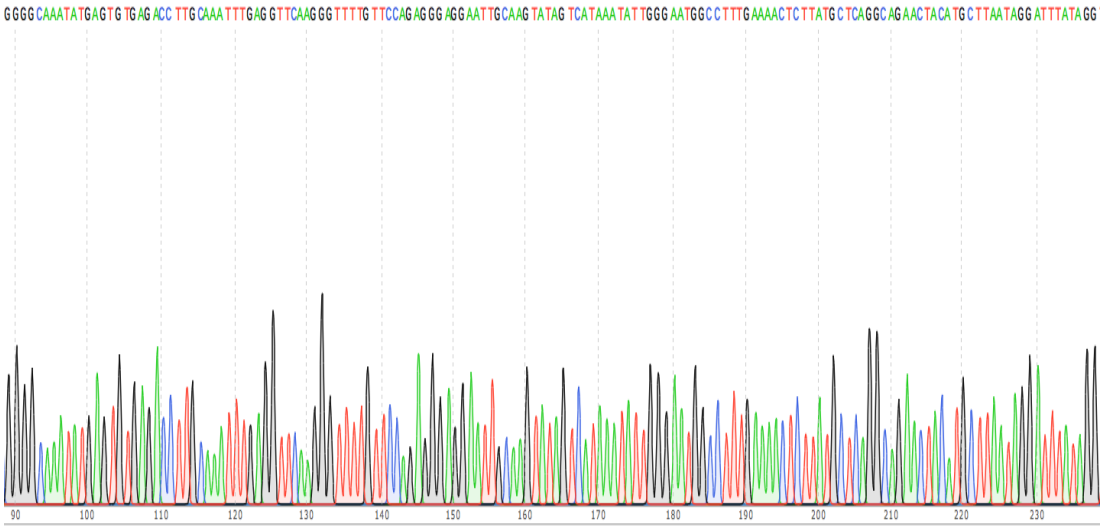

QC Score
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38
1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51
Position in read (bp)

F

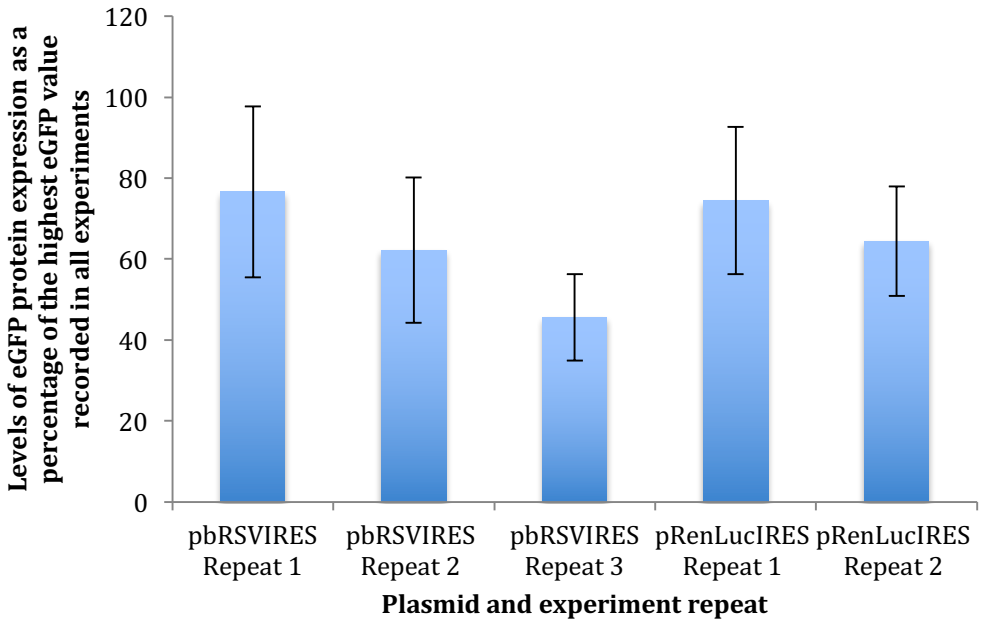

QC Score
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38
1 2 3 4 5 6 7 8 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51
Position in read (bp)



Appendix B1: QC plot of reads from sequencing libraries created in Chapter 3
QC plot of reads from sequencing libraries created in Chapter 3. Plot shows the quality of each nucleotide along unprocessed reads in each library. The software FastQC was used to test quality of sequencing reads. (A) FastQC plot from library hRSV4T. (B) hRSV8T. (C) hRSV4P. (D) hRSV8P. (E) bRSV4T. (F) bRSV8T, (G) bRSV4P

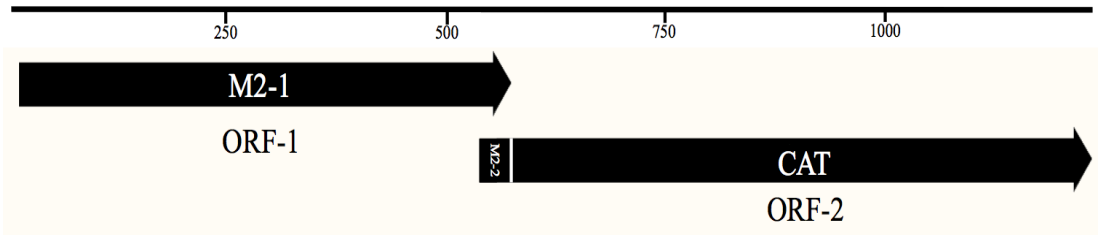


Appendix B2:Sequencing of the plasmid pbRSVM2
Sample of sanger sequencing from plasmid pbRSVM2 using GATC sequencing. Sequence ranges from position 1 to 150.



Appendix B3: eGFP protein expression for plasmids used to investigate if an internal initiation mechanism is used for initiation of translation of M2 ORF-2

Levels of expression of eGFP protein expressed from plasmids individually transfected into HEp-2 cells. Each repeat represents the average of three separate transfections of each plasmid. All experiments were performed during the same set of transfections. Results were normalised against the empty vector pBluescribe and are expressed as a percentage of the single plasmid transfection with the highest eGFP protein expression. Error bars indicate standard deviation.



Appendix B4: Schematic of the gene transcript transcribed from the plasmid pWildCAT

Schematic representation of the gene transcript transcribed from the hRSV plasmid pWildCAT used as control in *in vitro* transcription reactions. Including the two ORFs, ORF-1 (hRSV M2 ORF-1) and ORF-2 (hRSV M2 ORF-2/CAT ORF) which includes the ORF-1/ORF-2 overlap region

Appendix C

List of oligonucleotides.

Δ560CF

5′CAAGGTGACATAATTGTTAATGAACCAAATGAATAATCTAGAGAAAAA
AATC3′

Δ560CR

5′GATTTTTTTCTCTAGATTATTCATTTAGTTCATTAACAATTATGTCACCT
TG3′

Δ550F

5′GTAATAACCAAGGTGACATAATTATTAATGAACAAAATGAATAATC3′

Δ560GF

5′CAAGGTGACATAATTGTTAATGAACGAAATGAATAATCTAGAGAAAAA
AATC3′

Δ560GR

5′GATTTTTTTCTCTAGATTATTCATTTCGTTCATTAACAATTATGTCACCT
TG3′

Δ550R

5′GATTATTCATTTTGTTCATTAATAATTATGTCACCTTGGTTATTAC3′

Δ560UF

5′GATTTTTTTCTCTAGATTATTCATTTGGTTCATTAACAATTATGTCACCT3
′TG

Δ560UR

5′CAAGGTGACATAATTGTTAATGAACTAAATGAATAATCTAGAGAAAAA
AATC3′

5′AUGR

5′AGATTCTCTAGATTATTCATTTTGTTCAGTAACAATTATG3′

3′AUGR

5′AGATTCTCTAGATTATTCAGTTTGTTCATTAAC3′

AUG524F

5′CAAGGTGACATAATTGTTACTGTTGTTAATGAATAATCTAGAGAAAAA
AATCACTG3′

AUG557F

5′CAAGGTGACATAATTGTTACTGATGAAACTGAATAATCTAGAG3′

AUG524R

5′CAGTGATTTTTTTCTCTAGATTATTCATTAACAACAGTAACAATTATGTC
ACCTTG3′

AUG557R

5′CTCTAGATTATTCAGTTTCATCAGTAACAATTATGTCACCTTG3′

AUGSeqSF

5′CCAAGGTGACATAATTGTTACTGTTGTTAATGAATAATCTAGAGAAAA
AAATC3′

AUGSeqSR

5′CTCTAGATTATTCAGTTTCATCAGTAACAATTATGTCACCTTG3′

bRSV120R2

5′CAATAACATCATTTTCATCTTTGATTAGCAAAGAGGATATGCTGCAAGG
GTATTTTTCTG3′

bRSV120R1ST

5′TGCAAGGGTATTTTTCTGGGAAAATAATAATGTTGGAGTTGTTCATTTT
GTTCATTAAC3′

bRSV120R1WC

5′TGCAAGGGTATTTTTCTGGGAAAATAATAATGTTGGAGTTATTCATTTT
GTTCATTAAC3′

bRSV42

5′AGATTCTCTAGAGAAAATAATAATGTTGGAG3′

bRSVCATF

5′GTTGCCATCAAAAATCATGAGAACAAAATGAATAATCTAGAG3′

bRSVIRESF

5′AGATTCGCATGCGTTGCTATGAGTAAAC3′

bRSVSTOP

5′AGATTCTCTAGATTGTTCATTTTG3′

bRSVWTF

5′GGGAAAGGTACCATGGGGCAAATATGTCACG3′

bRSVWTR

5′AGATTCTCTAGATTATTCATTTTGTTCATTAACAATTAT3′

CAT120F1

5′CTAAGTCATCAAAATGTTCTTGACTACTTACAATTTCATCTAGAGAAAA
AAATCACTG3′

CAT120F2

5′ATATCCTCTTTGCTAATCAAAGATGAAAATGATGTTATTGTACTAAGTC
ATCAAAATG3′

CATF

5′GGGAAATCTAGAGAAAAAAATCACTG3′

CAT Hind III

5′ATAATAAAGCTTTTACGCCCCGCCCTGCCAC3′

CATInt

5′ACACGCCACATCTTGCG3′

CATR

5′GGGAAAGCATGCTTACGCCCCGCCCTGCCACTC3′

DoubleR

5′AGATTCTCTAGATTATTCAGTTTGTTCAGTAACAATTATG3′

eGFPIRESR

5′GTTTACTCATAGCAACACAGGCATGCTTACTTGTACAGCTCGTCCATGC
CGAG3′

eGFP Sph I

5′AATTAGCATGCTTGTACAGCTCGTCCATG3′

KozakF

5′GTAATAACCAAGGTGACATAATTGTTAGCCGCCACCATGGATAATCTA
GAGAAAAAAATC3′

KozakR

5′GATTTTTTTCTCTAGATTATCCATGGTGGCGGCTAACAATTATGTCACCT
TGGTTATTAC3′

KozakStopF

5′GTAATAACCAAGGTGACATAATTGTTAGCCGCCACCATGGACAATCTA
GAGAAAAAAATC3′

KozakStopR

5′GATTTTTTTCTCTAGATTGTCCATGGTGGCGGCTAACAATTATGTCACCT
TGGTTATTAC3′

M2 GTF

5′GGGAAAGGTACCGGGGCAAATATGTCACGAAG3′

M2 GTR

5′GGGAAAGCATGCTTTTAAATAACTATCTGTTAAG3′

pbRSV14F

5′GAATTAGCATGCACGAAGAAATCCCTGCAAATATG3′

pbRSV89F

5′GAATTAGCATGCCTTTGAATGGCCTCCACATGC3′

pbRSV209F

5′GAATTAGCATGCTAGAACAGAAGAATATGCATTG3′

pbRSV293F

5′GAATTAGCATGCCTGTGTTGCTATGAGTAAAC3′

pbRSV380F

5′GAATTAGCATGCGATAAGAATATATAACACAG3′

pbRSV527F

5′GAATTAGCATGCTAATAACCAAGGTGACATAATTG3′

PVM6 F

5′AATTAGCATGCGAGTGTGAGACCTTGCAAATTTG3′

PVM84 F

5′AATTAGCATGCGGAATGGCCTTTGAAAACTC3′

PVM204 F

5′AATTAGCATGCGACTGCTGAGTATGCCTTG3′

PVM370 F

5′AATTAGCATGCTTACTTGAGAAGTTGCAAC3′

PVM424 F

5′AATTAGCATGCGAGACAAATTATTCACATTC3′

PVM228

5′AATTAGCATGCCATAGGTGTGCTGAAAAGTTAC3′

PVM306

5′AATTAGCATGCATTGCAGAACTTGGATGTTGG3′

PstrucC F

5′CAACATCACTAAATCAATACCTTCTGGCACTTTGATCACTGTATTGCAG 3′′

PstrucC R

5′CTGCAATACAGTGATCAAAGTGCCAGAAGGTATTGATTTAGTGATGTTG 3′

PstrucGC F

5′GTTACCTGGAAAAAAGTAACAAGATCAGTAAATCAATACCTTCTGGCA CTTTGATCACTG3′

PstrucGC R

5′CAGTGATCAAAGTGCCAGAAGGTATTGATTTACTGATCTTGTTACTTTTTTCCAGGTAAC3′

PstrucT F

5′CATCACTAAATCAATATCTTTTGGCATTTTGATCACTGTATTGCAG3′

PstrucT R

5′CTGCAATACAGTGATCAAAATGCCAAAAGATATTGATTTAGTGATG3′

RenLucF

5′AGATTCGCATGCAACTTCTTAATTTACCAAAG3′

RenLucR

5′CTCTAGATTATTCATTTTGTTCTCATGATTTTTGATGGCAAC3′

ShuntKF

5′GGTAATAACAAGGTGACATAACACAAACCAAACAAAATGAATAACTCCAAC3′

ShuntKR

5′GTTGGAGTTATTCATTTTGTTTGGTTTGTGTTATGTCACCTTGTTATTACC3′

ShuntPF

5′GTAATAACAAGGTGACATAATTGTTTATGGAAATTATGAATAATCTAGAG3′

ShuntPR

5′CTCTAGATTATTCATAATTTCCATAAACAATTATGTCACCTTGTTATTAC3′

T3 down

5′CATGATTACGCCAAGCTCG3′

T7

5′TAATACGACTCACTATAG3′

T7-91

5′GGCGAAAGGGGGATGTGCTG3′