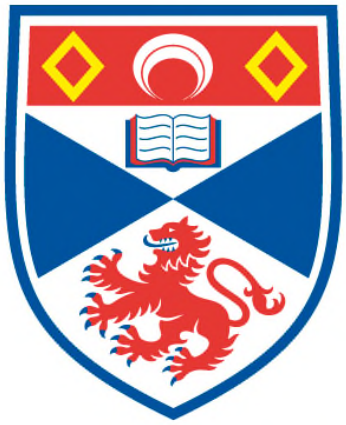
DEVELOPMENT OF A FOOT-AND-MOUTH DISEASE VIRUS REPLICON SYSTEM FOR THE STUDY OF RNA REPLICATION

Fiona Tulloch

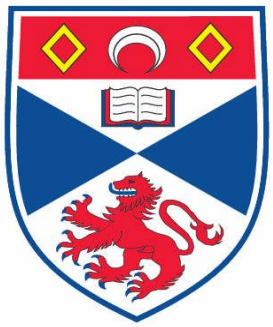
A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2015

DEVELOPMENT OF A FOOT-AND-MOUTH DISEASE VIRUS REPLICON SYSTEM FOR THE STUDY OF RNA REPLICATION

Fiona Tulloch



This thesis is submitted in partial fulfilment for the degree of PhD at the University of St Andrews

2014

Declarations

1. Candidate's declarations:

I, **Fiona Tulloch**, hereby certify that this thesis, which is approximately 51, 900 words in length, has been written by me, that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

Date …… …….. Signature of candidate ……..

I was admitted as a research student in **September 2010** and as a candidate for the degree of **Doctor of Philosophy (PhD) in Molecular Virology**; the higher study for which this is a record was carried out between **2010** and **2014**.

Date …… …….. Signature of candidate ……..

2. Supervisor's declaration:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of **Doctor of Philosophy (PhD) in Molecular Virology** in the University of St Andrews and that the candidate is qualified to submit this portfolio in application for that degree.

Date …… …….. Signature of supervisor ……..

Prof. M.D. Ryan

3. Permission for publication:

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be

made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

PRINTED / ELECTRONIC COPY

Embargo on all or part of print / electronic copies for a period of **one year** on the following grounds:

- Publication would be commercially damaging to the researcher, or to the supervisor, or the University

Supporting statement for printed / electronic embargo request:

Publication would breach the 'prior art' and invalidate any patent application.

Date …….…….. Signature of candidate ……..

Signature of supervisor ……..

Acknowledgements

I am extremely grateful to Martin for his guidance, supervision and overall support during this project. I approach science in a different manner due to your attitude and enthusiasm, and I thank you for that.

Thank you to the MRC for funding my studentship and enabling me to carry out this research.

Thanks to all the collaborators from Dundee, Edinburgh, Leeds and The Pirbright Institute who were involved with this project, and to Professor Peter Simmonds for allowing me to work in his laboratory during my final year - I learned a huge amount of skills and techniques which will help me with my future career in science.

To the Ryan group members, past and present; John Nicholson, Garry Luke, Ekaterina Minskaia and Ashley Pearson for generating constructs used throughout this thesis and/or help carrying out experiments. To all those mentioned above, along with Claire Roulston and Sean Wilson- thank you for all your help, advice, and for all the laughs we have had together. It has been an absolute pleasure to work with you all!

I am grateful to all the friends I made in St Andrews during this time (in particular, Bjoern and Richard): I have gained friends for life in you both.

Many, many thanks to all my 'Perth and Dundee' friends for their constant support throughout my PhD. You have helped to keep me grounded and sane, and understood the lack of my availability at times. In particular, Lisa, Louise, Laura, Gary and Sharon; you made the last few months of this thesis more bearable due to your encouragement.

To my family (I cannot mention you all!); you have been with me the whole way. Thank you for believing in me.

Finally, Craig, without you, I do not think it would have been possible to complete this PhD. You are my calming influence and I have needed you many times during my studies. Thank you for bearing with me through it all. I am eternally grateful for your love, understanding, and support.

Abstract

Foot-and-mouth disease (FMD) is a highly infectious disease of wild and domestic cloven–hoofed animals such as cattle, swine and deer. It is caused by one of the most contagious animal diseases known; FMD virus (FMDV). Since the disease is endemic in many countries, transmission by international travel/ trade presents an on-going potential threat to the UK. Very little is known at the molecular level about how FMDV replicates within host cells. In this study, FMDV replicons have been used to investigate FMDV RNA replication and to improve our understanding of the viral life cycle: a process which will aid in the production of a new generation of live-attenuated vaccine candidate strains. Sequences encoding the capsid coding region of the genome have been replaced with green fluorescent protein (GFP) such that replication can be monitored in live cells *via* GFP fluorescence. This provides a rapid, non-invasive screen for replicative fitness that can be used outwith high disease security facilities. Differences between replicating and non-replicating forms could easily be distinguished, highlighting the potential of this system to screen for attenuated genomes.

The FMDV replicon system was improved through a series of construct modifications until an optimal system was produced. A range of different methods were used to attenuate the replication of these genomes. Of major significance is the finding that increasing dinucleotide frequencies were shown to decrease growth kinetics of Echovirus 7 – as opposed to altering the codon-pair bias - and the application of this finding to construction of further replicon systems (and RNA viruses in general) is described.

Table of Contents

Abbreviations

% (v/v)	percentage concentration (volume per volume)
% (w/v)	percentage concentration (weight/volume)
%	percentage
°C	degrees Celsius
3' UTR	3' untranslated region
3C^{pro}	FMDV 3C proteinase
3D^{pol}	FMDV 3D polymerase
5' UTR	5' untranslated region
aa	amino acid
Abs	absorbance
Aq.GFP	*Aequorea victoria GFP*
BEV 2A^{pro}	Bovine enterovirus 2A proteinase
bp	base pair
BRAV	bovine rhinitis A virus
BRBV	bovine rhinitis B virus
BRV	bovine rhinoviruses
CAT	chloramphenicol acetyl-transferase
cDNA	copy DNA
CHX	cycloheximide
CL	cloverleaf
cm	centimetre
COS-1	african green monkey kidney cells
CPB	codon-pair bias
CPE	cytopathic effect
CpG	cytosine followed by guanine in a RNA sequence
CVB3	coxsackievirus B3
DI	defective interfering
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DTT	dithiothreitol
E.coli	*Escherichia coli*
E7	Echovirus 7

ECL	enhanced chemiluminescence
EDTA	ethylenediaminetetraacetic acid
eEF 1A	eukaryotic elongation factor 1A
eEF2	eukaryotic elongation factor-2
eEF2K	eukaryotic elongation factor-2 kinase
eIF4G	eukaryotic translation initiation factor 4-gamma
ERAV	equine rhinitis A virus
eRF1/3	eukaryotic translation termination (release) factors 1 and 3
FBS	foetal bovine serum
FMD	foot-and-mouth disease
FMDV	foot-and-mouth disease virus
FP	fluorescent protein
g	gram
GFP	green fluorescent protein
GuHCl	guanidinium hydrochloride
HAV	hepatitis A virus
HDV	hepatitis delta virus
HEK 293	human embryonic kidney 293 cell line
HeLa	human cervix carcinoma epithelial cells
HH	hammerhead
HIV-1	human immunodeficiency virus type 1
hnRNP C	heterogeneous nuclear ribonucleoprotein C
hr	hour
HRP	horse-radish peroxidase
HRV	human rhinovirus
IAV	influenza A virus
IgG	immunoglobulin G
IPV	inactivated polio vaccine
IRES	internal ribosome entry site
kDa	kilodalton
LB	lysogeny broth
L^{pro}	leader proteinase
M	molar concentration (moles per litre)
mg	milligram
min	minute
ml	millilitre
mm	millimeter

mM	millimolar
MOI	multiplicity of infection
mRNA	messenger RNA
ng	nanogram
NSP	non-structural protein
nt	nucleotide
OPV	oral polio vaccine
ORF	open-reading frame
p.i.	post-infection
p.t.	post-transfection
PABP	poly(A) binding protein
PAC	puromycin-n-acetyl-transferase
PBS	phosphate-buffered saline
PBS-T	phosphate-buffered saline-tween 20
PCBP	poly(C) binding protein
PCR	polymerase chain reaction
pH	$-\log_{10}[H^+]$
poly(C)	polyribocytidylic acid
PRRSV	porcine reproductive and respiratory syndrome virus
Pt.GFP	*Ptilosarcus gurneyi* GFP
PV	poliovirus
RE	restriction enzyme
RF	release factor
RFP	red fluorescent protein
RIPA	radioimmunoprecipitaton assay buffer
RNA	ribonucleic acid
RNP	ribonucleoprotein
rNTP	ribonucleotide triphosphate
rpm	revolutions per minute
RT-PCR	reverse transcription PCR
RV	rabies virus
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
sec	second
SL1/2	stem-loop 1/2
SV40	simian virus 40
SVDV	swine vesicular disease virus

T7TΦ	T7 Φ terminator
TAE	tris-acetate-EDTA buffer
$TCID_{50}$	50 % tissue culture infective dose
T_m	melting temperature
TMEV	Theiler's murine encephalitis virus
Tris	tris-hydroxymethyl-aminomethane
tRNA	transfer RNA
U	unit of enzyme
UpA	uracil followed by adenine in an RNA sequence
UV	ultraviolet
V	volts
vRNA	viral RNA
VSV	vesicular stomatitis virus
WT	wild-type
µg	microgram
µl	microliter
µm	micrometer
µM	micromolar

Chapter 1: Introduction

1.1 Picornaviruses

1.1.1 Classification and genome structure

The Picornaviruses are a diverse family of pathogens which infect humans and animals. They are responsible for diseases such as poliomyelitis, the common cold, meningitis and foot-and-mouth disease (Lin *et al.*, 2009; Stanway, 1990) They currently belong to 26 genera (Figure 1.1) as classified by the International Committee on Taxonomy of Viruses (ICTV) in 2013 (Knowles, 2013). The classification is based on phylogenetic properties which reflect evolutionary history.

Picornavirus genomes are small icosahedral particles containing a single-stranded, positive-sense RNA, ~7.5-8.5 kb in length. Although there is only a small degree of sequence homology at the nucleic acid level, picornaviruses share a similar genome structure and organisation (Figure 1.2). The RNA genome encodes a highly structured 5' untranslated region (UTR) which contains a covalently attached protein (VPg) to the 5' terminus (Crawford & Baltimore, 1983; Nomoto *et al.*, 1977), a large single open-reading frame (ORF); followed by a structured 3' UTR that is polyadenylated at the 3' terminus. The ORF encodes a single large polyprotein that consists of three main regions; P1 containing the structural proteins, and the P2/P3 regions which encode all the non-structural proteins involved in manipulation of host-cell endomembrane structures (P2), RNA replication (P3), and polyprotein processing ($3C^{pro}$). The polyprotein is processed by virally encoded proteases to produce the mature viral proteins and protein intermediates (Lin *et al.*, 2009; Stanway, 1990). All picornaviruses lack a cap structure and perform cap-independent translation from the internal ribosome entry site (IRES). Despite the similar genome organisation shared, there are a few differences between members of this virus family. The 5' UTRs of cardio- and aphthoviruses contain long polyribocytidylic acid [poly(C)] tracts and are characterised further by the presence of a leader protein encoded 5' of the capsid region. Many picornaviruses also differ in the 2A protein they encode, as well as the number of 3B (VPg) copies they contain (Knowles *et al.*, 2010; Stanway, 1990).

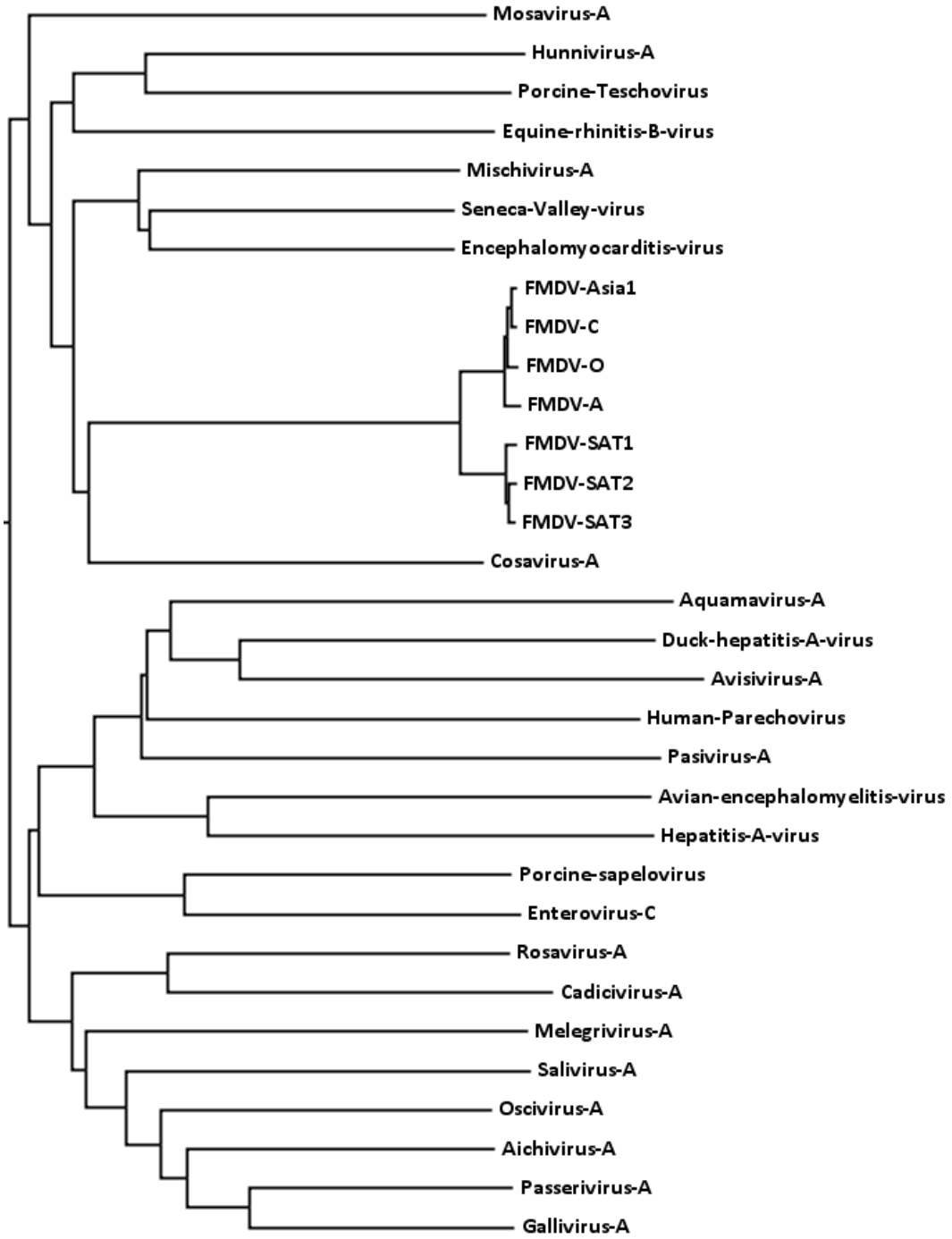


Figure 1.1. Relationship between picornavirus 3D polymerases of the genus type-species of the current 26 genera. Sequences were aligned using MEGA (Tamura *et al.*, 2011) and the tree visualised using FigTree (Rambaut, 2008). Accession numbers of genus type species and FMDV strains are listed within the appendices (A.4).

1.1.1.2 Aphthoviruses

The Aphthoviruses consist of Foot-and-mouth disease virus (FMDV), Bovine rhinitis B virus (BRBV), Equine rhinitis A virus (ERAV) and a newly identified species, Bovine rhinitis A virus (BRAV) (Knowles, 2014; Knowles *et al.*, 2010). FMDV is the genus type species and seven serotypes exist (A, O, C, Asia 1 and South African Territories 1, 2 and 3), with numerous subtypes within each serotype (Knowles *et al.*, 2010). FMDV is the cause of foot-and-mouth disease (FMD), which is a highly infectious disease that affects wild and domestic cloven-hoofed animals. ERAV contains one serotype which is of the same name and causes upper respiratory infections in horses. BRVs have been found in cattle with acute respiratory disease and their role in disease remains uncertain (Knowles, 2014).

1.1.2 Foot-and-mouth disease virus

FMDV is the type species of the *Aphthovirus* genus of the *Picornaviridae* family. The polyprotein translated from the ~8.5 kb viral mRNA genome is processed by FMDV-encoded proteinases, the L proteinase (L^{pro}) and 3C proteinase ($3C^{pro}$), as well as the 2A peptide (Figure 1.2; Grubman *et al.*, 2010). These co- and post-translational cleavages generate 'intermediate' and 'mature' structural and non-structural proteins (see 1.2.3). L^{pro} and $3C^{pro}$ are also involved in the cleavage of a number of host cell proteins.

The 5' UTR of FMDV is the longest within the picornaviridae. It comprises ~ 1,300 nucleotides (nts) or ~ 1/7 of the genome and can be subdivided into at least 5 distinct domains (Belsham, 2005; Grubman *et al.*, 2010) The N-terminal domain of the polyprotein codes for L^{pro} which releases itself from the nascent polyprotein, cleaving at its own C-terminus, then degrades important target host-cell proteins (see 1.1.2.6). As mentioned above, only aphthoviruses and cardioviruses contain a leader proteinase and the two in-frame initiation codons present at the beginning of the polyprotein are a unique feature of the aphthoviruses, as is the proteolytic activity possessed by aphthovirus leader proteinases (see 1.1.2.6; Grubman *et al.*, 2010; Stanway, 1990). Five copies of the capsid proteins (1A-D) assemble to form a pentamer, twelve of which form the particle (see 1.2.1). The capsid protein precursor (P1-2A) is processed by $3C^{pro}$ with the exception of the 1AB cleavage, which occurs with encapsidation of

virus RNA (Belsham, 2005). Like other members of the picornaviruses, the P2 and P3 regions encode replication proteins, however FMDV is unique in encoding three copies of 3B (VPg; see 1.1.2.12; Grubman *et al.*, 2010).

1.1.2.1 5' UTR

The Untranslated Regions (UTRs) account for ~16 % of the FMDV genome. The 5' UTR of FMDV is >1kb in length, which is in stark contrast to the ~50-100 nt encoded by cellular mRNAs. It seems counterintuitive that a virus that can replicate so effectively can afford to devote so much of its genetic content to non-coding sequences, and implies that the RNA structural features in this region must contribute to the efficiency of viral replication. A number of highly structured elements reside within the 5' UTR of FMDV (Figures 1.2, 1.3) and relatively little is known about the detailed functioning of these RNA elements.

1.1.2.2 The S-fragment

The S-Fragment is approximately 360 nucleotides in length and has been predicted to adopt a long, linear hairpin structure interrupted by short unpaired regions (Clarke *et al.*, 1987; Witwer *et al.*, 2001). The function of the S-fragment in unknown, but it is assumed that it has crucial roles in the control of RNA replication. A recent study demonstrated that the S-fragment can form complexes with cellular proteins including RNA helicase A, PABP and PCBP, together with the viral proteins 2C, 3A and $3C^{pro}$ (Lawrence & Rieder, 2009). In addition, it can bind to structural motifs in the 3' UTR (Lawrence & Rieder, 2009; Serrano *et al.*, 2006). These findings suggest that multiple interactions are important for the mechanisms and control of viral replication, but the details of these processes are largely unknown.

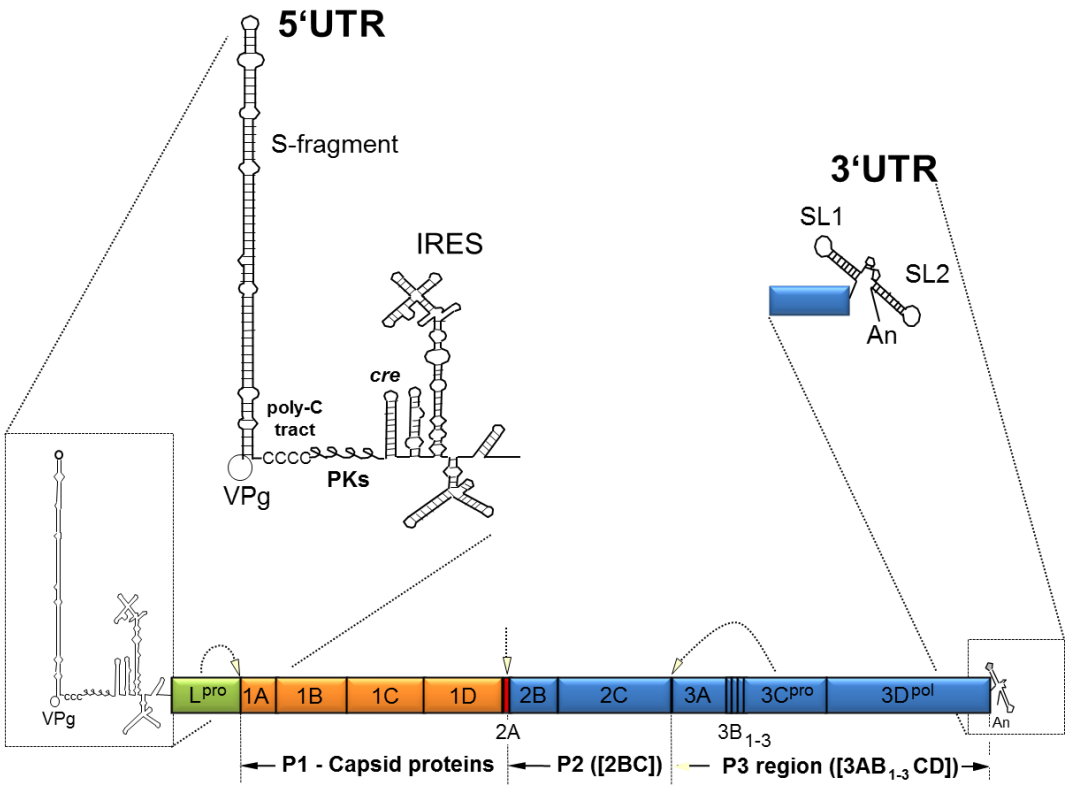


Figure 1.2. Schematic representation of FMDV genome organisation. The secondary structure located within each non-coding region is shown. The large ORF is translated into a polyprotein which is processed by viral proteases. The primary polyprotein cleavages by L^{pro}, 2A and $3C^{pro}$ are shown as dashed arrows with major precursors annotated.

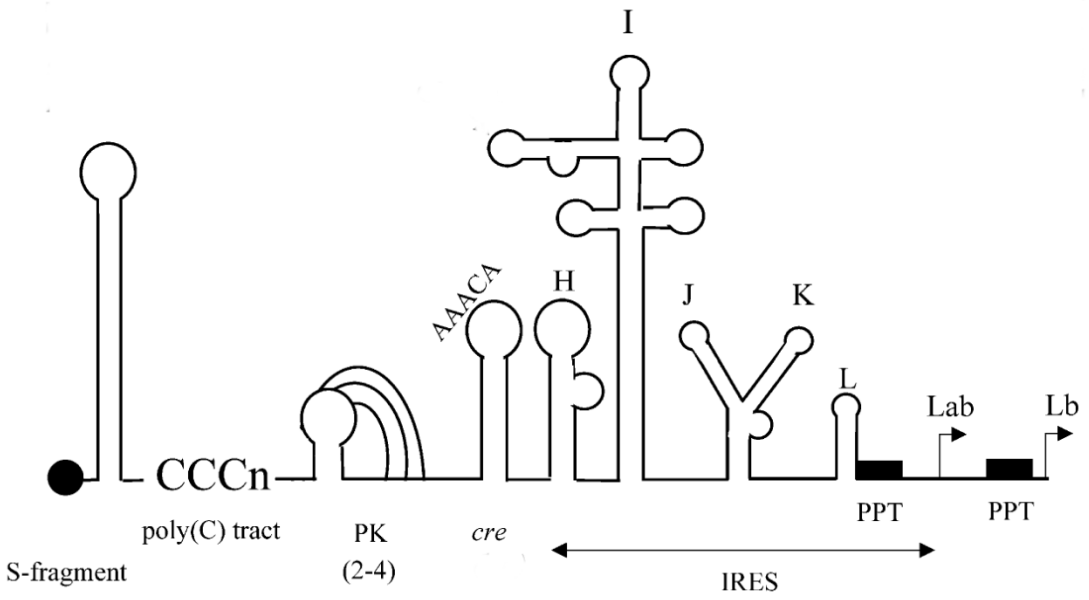


Figure 1.3. Structural elements within the 5' UTR of FMDV. Detailed schematic representation of the structured regions predicted to be present within the 5' UTR. All elements are described within the text. The five domains of the IRES (H-L) and the conserved motif within the *cre* are labelled. PPT, polypyrimidine tract (adapted from Belsham, 2005).

1.1.2.3 The poly(C) tract and pseudoknot domains

The poly(C) tract is a distinctive feature of aphthoviruses and most cardioviruses (excluding Theiler's murine encephalitis virus). In FMDV the poly(C) tract is found immediately 3' of the S-fragment and is followed by a variable number of pseudoknot domains. The pseudoknot domain varies in number in the genomes of different viral isolates, but their biological roles in virus replication are completely unknown (Belsham, 2005; Grubman *et al.*, 2010). Field strains of FMDV typically contain poly(C) tracts of around 150-200 nts in length. The length of the tract can dramatically influence pathogenicity within cardioviruses, but with respect to FMDV, some reports suggest the length is important for virulence (Harris & Brown, 1977; Rieder *et al.*, 1993), and others have suggested otherwise (Giomi *et al.*, 1984). Pathogenicity of mengovirus in mice can be attenuated by the truncation of the poly(C) tract (Duke *et al.*, 1990). It seems that FMDV requires a minimum length of poly(C), the length of the tract increasing during passage. When a tract of just 6 C residues was present in RNA transcripts derived from a cDNA clone, the tract length increased through passage to contain tracts of around 80 C residues. However, when cDNA clones contained only 2 C residues, the tract was not amplified and the virus grew slower in tissue culture. The pathogenicity of the virus however, remained the same when tested in mice (Rieder *et al.*, 1993). The longest poly(C) tract identified was found in persistently infected BHK cells and was 420 nt in length, but this virus was attenuated in cattle and mice (Escarmís *et al.*, 1992). Thus, the function of this region within FMDV, and other members of the picornavirus family, remains unclear.

1.1.2.4 *cis*-acting replication element

Cis-acting replication elements (*cre*) were first discovered when it was shown that HRV-14 genomes lacking the P1 region resulted in inhibition of viral RNA synthesis (McKnight & Lemon, 1996, 1998). This element was required to permit replication as an RNA entity, although it was found within a protein coding region. Picornavirus *cre*s are thought to form stable stem-loop structures, with a conserved motif of -AAACA- located within the terminal loop region. This motif is proposed to act as a template for uridylation of $3B^{VPg}$ - a primer for initiation of RNA replication (Paul *et al.*, 1998). Location of the *cre* within the genome of various picornaviruses has been mapped to different locations; the HRV-14 *cre* is located within 1D (VP1) (McKnight & Lemon, 1998), the PV *cre* within the 2C

coding region (Goodfellow *et al.*, 2000) and within the 5' UTR of FMDV (Mason *et al.*, 2002). FMDV is so far unique in having the *cre* located outwith the coding region. Mutational analyses of the *cre* within FMDV showed that mutations in the -AAACA- motif impaired replication of FMDV replicons, with mutations in the stem region demonstrating the importance of this structure for replication. The location of the *cre* to the 5' UTR is not critical for replication as replicons and viruses encoding a *cre* within the 3' UTR were functional (Mason *et al.*, 2002). Retention of activity following movement of this element within the genome has also been demonstrated for PV (Goodfellow *et al.*, 2000).

1.1.2.5 Internal Ribosome Entry Site (IRES)

The region essential for initiation of protein synthesis is located within the 3'-end of the 5' UTR (Kühn *et al.*, 1990). This element, an internal ribosome entry site (IRES), is similar to the cardiovirus IRES (group 2). The FMDV IRES is ~ 450 nt in length and folds into a complex secondary structure consisting of five domains (Pilipenko *et al.*, 1989). Most of the canonical translation initiation factors necessary for translation of cellular mRNAs are also required by the FMDV IRES (Pilipenko *et al.*, 2000). A polypyrimidine tract is located at the 3'-end of the IRES and is conserved amongst most picornavirus IRES elements (Belsham, 2005). This region interacts with polypyrimidine tract binding protein (PTBP), which has been shown to enhance FMDV IRES activity *in vitro* (Pilipenko *et al.*, 2000). Furthermore, both purified PTBP and IRES-specific trans-acting factor 45 ($ITAF_{45}$) have been shown to stimulate the formation of the 48S initiation complex on the FMDV IRES (Pilipenko *et al.*, 2000). It is thought that the binding of these IRES-interacting proteins helps to facilitate IRES activity through stabilisation of the tertiary structure and enhancement of initiation factor binding (Belsham, 2005; Pilipenko *et al.*, 2000). In FMDV, the initiation of translation occurs at two sites, 84 nt apart, which are conserved across all seven serotypes. This results in the production of two forms of the first protein component of the polyprotein; the L proteinase (described below). The context of this arrangement remains largely unknown.

1.1.2.6 L proteinase

The FMDV L proteinase (L^{pro}) is a papain-like cysteine protease, which cleaves itself from the viral polyprotein (Strebel & Beck, 1986). As well as cleaving itself from the polyprotein, L^{pro} is also involved in the cleavage of the eukaryotic translation initiation factor 4 gamma (eIF4G; Devaney *et al.*, 1988; Kirchweger *et al.*, 1994), resulting in the inhibition of host cap-dependent mRNA translation. FMDV translation is a cap-independent process and therefore does not require eIF4G for successful viral protein synthesis. As a result, host cell protein synthesis is rapidly inhibited without affecting FMDV translation (Belsham, 2005).

L^{pro} contains two in-frame initiation codons, 84 nts apart, which are conserved in all natural FMDV isolates (Carrillo *et al.*, 2005; Sangar *et al.*, 1987) and results in two forms being produced, Lab and Lb (Belsham, 2005). Mutagenesis of either initiation site, leading to the production of only one of these two forms, demonstrated that each form of proteinase behaves similarly in proteolytic activity: there is no effect on eIF4G cleavage or viral polyprotein processing (Cao *et al.*, 1995). Despite the Lb form being the predominant form produced (Cao *et al.*, 1995; Sangar *et al.*, 1987), deletion of the Lb region is not important for virus viability (Piccone *et al.*, 1995). As already mentioned, viruses containing the 84 nt Lab 'spacer' region, but lacking the Lb region, are viable; however, if this spacer region is then removed it is lethal for the virus (Belsham, 2013; Piccone *et al.*, 1995). Additionally, if the Lb start site is mutated within an infectious FMDV cDNA clone, allowing only the Lab form to be produced, virus cannot be rescued. When the reverse is performed, permitting the production of the Lb form only, viable virus can be obtained (Belsham, 2013; Cao *et al.*, 1995). Furthermore, insertion of a 57 nucleotide sequence between the two initiation codons results in an attenuated phenotype in cattle, suggesting this region is important for virulence (Piccone *et al.*, 2010).

L^{pro} has been of particular interest in terms of vaccine development due to differences in replication kinetics observed with viruses encoding the differing forms. Piccone *et al.*, (1995) found that viruses lacking the portion of L directly after the Lb initiation codon replicated less efficiently in BHK-21 cells. This replication was only slightly slower than the wild-type (WT) virus, however, when tested in cattle the 'partially leaderless' virus had greatly reduced pathogenicity and did not spread from the initial site of infection (Brown *et al.*, 1996). When this

virus was later tested in swine, similar results to those observed earlier in cattle were found (Chinsangaram *et al.*, 1998).

There is evidence that L^{pro} is involved in the regulation of genes associated with the innate immune system. Recently it was shown that interferon IFN-β and IFN stimulated gene (ISG) mRNA was significantly induced in cells transfected with leaderless FMDV compared to WT (de los Santos *et al.*, 2006). L^{pro} has also been shown to be involved in the degradation of the p65/RelA subunit of nuclear factor kappaB (NF-κB) (de los Santos *et al.*, 2007, 2009).

1.1.2.7 2A protein

The large polyprotein produced during FMDV translation is primarily cleaved into distinct domains through the action of virally encoded proteinases. However, separation of the capsid protein precursor from the downstream region encoding the replication proteins is achieved through the cleavage action of FMDV 2A. This oligopeptide is only 18 aa long and, unlike the virus-encoded 2A proteinase ($2A^{pro}$) of the entero- and sapeloviruses, lacks protease activity (Martínez-Salas & Ryan, 2010). FMDV 2A-mediated cleavage occurs at its C-terminus, at the novel motif (DxExNPG$^{\downarrow}$P), *via* a novel translational recoding event termed "ribosome skipping" (Donnelly *et al.*, 2001b). When the 2A region (in addition to the N-terminal proline of FMDV 2B) was inserted into an artificial reporter system, such that CAT and β-glucuronidase (GUS) proteins flanked 2A, three major products were observed. These products consisted of uncleaved polyprotein [CAT-2A-GUS], GUS and [CAT-2A], demonstrating that FMDV 2A was able to mediate cleavage in this artificial system in a similar manner to that observed during FMDV polyprotein processing (Ryan & Drew, 1994). The cleavage activity is sensitive to changes N-terminal to 2A, however, replacement of sequences downstream of the N-terminal proline of 2B does not affect activity (Ryan *et al.*, 1991). Using translation systems *in vitro*, a 2- to 5-fold molar excess of protein products encoded upstream of 2A over those downstream is routinely observed following quantification of cleavage products. The proposed model for this activity is the formation of a helix by 2A which interacts with the exit tunnel of the ribosome. The tight turn formed at the C-terminus of this helix, in addition to the interaction with the ribosome, is thought to prevent formation of the peptide bond at the P-site between the incoming prolyl-tRNA in the A-site. Prolyl-tRNA then dissociates from the ribosome and translation is ceased (Donnelly *et al.*, 2001b).

Recent studies have shown the involvement of eukaryotic translation termination (release) factors 1 and 3 (eRF1/3) in 2A mediated cleavage. These factors facilitate the hydrolysis of ester bond leading to release of the nascent peptide (Doronina *et al.*, 2008).

1.1.2.8 2B and 2C proteins

Infection of cells by picornaviruses leads to reorganisation of cellular membranes as sites for RNA replication. The enterovirus 2B protein has been shown to impair Golgi complex trafficking: calcium is released from the ER into the cytoplasm by forming pores within these membranes, reducing the ER luminal Ca^{2+} content (van Kuppeveld *et al.*, 2005; Lin *et al.*, 2009). The exact biological function of the picornavirus 2C protein remains unknown. It is thought to facilitate the rearrangement of membranes, along with 2B, due to the presence of a predicted amphipathic helix in its N-terminus. A conservd ATPase domain is located upstream of this predicted helix, with ATPase activity documented (Sweeney *et al.*, 2010). The presence of a helicase domain has also been predicted, however, helicase activity has never been documented (Gorbalenya & Koonin, 1989; Palmenberg *et al.*, 2010; Sweeney *et al.*, 2010). It was demonstrated that FMDV 3A, 2B, 2C, and 2BC expressed alone in vero cells associated with cellular membranes. The effect of cell trafficking was investigated and it was shown that 2BC disrupted protein transport from the ER to the Golgi body, affecting protein delivery to the cell surface (Moffat *et al.*, 2005).

1.1.2.9 3A protein

The 3A protein of picornaviruses differs in length, with the 153 aa of FMDV 3A being one of the largest described. Point mutations and deletions in FMDV 3A have been linked to altered host specificity, adaptation, attenuation and virulence. Repeated passage of FMDV in chicken embryos generated attenuated strains which contained deletions in the C-terminus of 3A (Dong-Sheng *et al.*, 2011). A similar deletion was also found in the FMDV strain which caused the 1997 outbreak in Taiwan. This outbreak only affected swine and had an abnormally high mortality rate. Beard & Mason, (2000) discovered that this virus contained a 10-amino-acid deletion (aa 93-102) in the C-terminus of 3A. It has been suggested that mutations in the region surrounding the deletion (aa 73-143)

could also be a factor in the modified host range (Knowles *et al.*, 2001). FMDV isolates from pigs carrying deletions in 3A (Type O aa 93-102; Asia-1 aa 133-143) grew well in porcine cells and caused disease in swine, but displayed restricted growth in bovine cells *in vitro* and significant attenuation in bovines *in vivo* (Pacheco *et al.*, 2003). Furthermore, a single mutation within 3A mediates adaptation of FMDV to the guinea pig demonstrating the involvement 3A has in host range (Núñez *et al.*, 2001).

Both FMDV and poliovirus proteins associate with the ER and both have hydrophobic domains that could explain this association, but the N- and C-termini are very different, with FMDV 3A having a C-terminal extension of ~50 aa (Moffat *et al.*, 2005). Poliovirus 3A inhibits protein trafficking from the ER to the Golgi, and it was postulated that FMDV 3A may have a similar role. It was, however, shown that in FMDV, 2BC, and not 3A, has this role (Moffat *et al.*, 2005).

1.1.2.10 3B (VPg) protein

The $3B^{VPg}$ protein is covalently attached to the 5'-end of both positive and negative RNA genomes. This linkage is made through a conserved tyrosine within $3B^{VPg}$ with a phosphate of the terminal uridine present at the 5'-end of the genome. FMDV is unique in comparison to other picornaviruses as it contains three copies of the $3B^{VPg}$ protein in the viral genome (Forss & Schaller, 1982; Forss *et al.*, 1984). Seal aquamavirus A1 and mosavirus contain two $3B^{VPg}$ proteins (Kapoor *et al.*, 2008; Phan *et al.*, 2011; Reuter *et al.*, 2014), with all other members of this family containing one copy of $3B^{VPg}$. All three FMDV $3B^{VPg}$ copies have been found attached to virion RNA within infected cells, but the exact significance of this remains unknown. Viruses produced which lacked the normal number of $3B^{VPg}$ copies, or contained mutated copies, were shown to produce decreased RNA yields and had an effect on infective particle formation (Falk *et al.*, 1992). All three copies of $3B^{VPg}$ are uridylated for priming of RNA transcription by $3D^{pol}$ (see 1.2.2); however $3B^{VPg}{}_3$ shows the highest activity (Nayak *et al.*, 2005).

1.1.2.11 3C and 3D proteins

FMDV $3C^{pro}$ performs most of the polyprotein processing which generates the mature and intermediate protein products (see 1.2.3). It is a member of the

trypsin-like serine proteases, however, it was confirmed during mutagenesis experiments that the active site consisted of a cysteine rather than a serine (Grubman *et al.*, 1995). The predominant function of FMDV $3C^{pro}$ is polyprotein processing, however it has also been shown to modify host cell proteins. Falk *et al.*, (1990) demonstrated cleavage of histone H3 by $3C^{pro}$ during infection. Other cellular targets include the cap-binding components eIF4A and eIF4G. Cleavage of eIF4G occurs at the C-terminus of the site generated by L^{pro} activity (Belsham *et al.*, 2000).

Precursors containing $3C^{pro}$ retain catalytic activity (P3, 3ABC, 3CD). PV $3CD^{pro}$ is involved in processing the capsid protein precursor (Ryan & Flint, 1997) and has a role in RNA replication (see 1.2.2).

The 3D protein is an RNA-dependent polymerase, and like other RNA-dependent polymerases is error prone: ~1-2 nucleotides are misincorporated per genome-copying event (Ward *et al.*, 1988). $3D^{pol}$ is involved in catalysing the synthesis of positive and negative sense RNA during viral replication. Uridylated $3B^{VPg}$ is required as a primer by $3D^{pol}$ to initiate replication of either strand (Paul *et al.*, 1998). The functional domain of $3D^{pol}$ is a GDD motif and is conserved within the picornaviruses, and similar to the YXDD motif conserved amongst other RNA polymerases (Stanway, 1990).

1.1.2.12 3’ UTR and poly(A) tail

The 3’ UTR of FMDV encodes two stem-loop structures (Witwer *et al.*, 2001) followed by a poly(A) of varying length: the length of the poly(A) tract has been implicated to be important for both infectivity of PV RNA (Sarnow, 1989; Spector & Baltimore, 1974) and negative-strand RNA synthesis (Silvestri *et al.*, 2006).

It has been shown that deletion of the 3’ UTR of FMDV abolishes viral replication (Saiz *et al.*, 2001), whereas PV replication can occur when this region is deleted, albeit at a reduced level (Todd *et al.*, 1997). Deletion of SL1 within the 3’ UTR is tolerated, however, deletion of SL2 is lethal for virus viability. Viruses containing the SL1 deletion displayed reduced growth kinetics and were attenuated when inoculated in swine (Rodríguez Pulido *et al.*, 2009). The 3’ UTR interacts with both the S-fragment and the IRES domains within the 5’ UTR (Serrano *et al.*, 2006). Both the SL1 and SL2 stem-loops were required to interact with the IRES, whereas the S-fragment could form contacts with either of these structures.

Additionally, the S-fragment–3' UTR interaction was dependent on the presence of the poly(A) tail. When the 3' UTR was mixed with cellular extracts, co-precipitation occurred with both poly(rC)-binding protein 2 (PCBP2) and poly(A)-binding protein (PABP): both of which are involved in the replication of other members of the picornavirus family (see 1.2.2).

1.2 Viral life-cycle

1.2.1 Virion structure and cell entry

FMDV particles are non-enveloped, icosahedral particles with a diameter of ~25-30 nM. The capsid is comprised of 60 copies of each of the structural proteins 1A-D which form into capsomers (Figure 1.4). Proteins 1B-D (VP2, 3 and 1) fold into a wedge-shaped β-barrel structure. The loops from each β-barrel connect the individual proteins, with some of the loops decorating the virus surface. The 1A protein, however, is located within the capsid and contains a myristyl group attached to its N-terminus (Fry & Stuart, 2010). It is known that this myristolation is essential for capsid assembly within PV (see 1.2.4; Marc *et al.*, 1991). Members of the picornaviruses contain a surface canyon, or pit, on the outer surface of the viral particle; FMDV, however, lacks this surface feature. FMDV capsids are relatively smooth structures with the exception of a long loop belonging to 1D (VP1). This GH loop contains a highly conserved RGD motif which is essential for cell attachment (Fox *et al.*, 1989; Mason *et al.*, 1994). The C-terminal portion of this loop (~140-160 aa) is the main antigenic site, to which most of the immune response is directed (Novella *et al.*, 1993).

FMDV binds to cell surface $\alpha_v\beta$ integrins ($\alpha_v\beta_1$, $\alpha_v\beta_3$, $\alpha_v\beta_6$, $\alpha_v\beta_8$) and enters the cell through clathrin mediated endocytosis (Burman *et al.*, 2006). This is followed by acidification of the endosome which disrupts the particle, releasing the viral RNA into the cytoplasm (Levy *et al.*, 2010).

1.2.2 Genome replication

There have been very few studies which have focused on transcription and replication in FMDV. This process has been studied extensively within enteroviruses; therefore analogies have to be made with FMDV.

Picornaviruses replicate their RNA in close association with host-cellular membranes. This results in major rearrangements of these membranes to create vesicular structures upon which virus genome replication takes place (Bienz *et al.*, 1990; Schlegel *et al.*, 1996). For most members of the picornaviruses these membrane structures cluster around the perinuclear region and occupy most of the cytoplasm (van Kuppeveld *et al.*, 2010). FMDV is one of the few exceptions, displaying distinct membrane vesicles which do not cluster and are present in fewer numbers (Monaghan *et al.*, 2004).

Before replication can begin, the viral RNA must be translated to produce the viral proteins required for subsequent replication. It is postulated that translation of the positive strand has to end before negative strand synthesis can commence. Using PV as a model, it was shown that accumulation of $3CD^{pro}$ repressed translation while PCBP, which interacts with the IRES of PV, enhanced translation in infected cells. It is thought that $3CD^{pro}$ binds to the 5' cloverleaf (CL) structure within the 5' UTR and alters the affinity of PCBP for the IRES, causing a switch from translation to replication (Gamarnik & Andino, 1998).

The first step in viral RNA synthesis is the uridylation of $3B^{VPg}$ by $3D^{pol}$. Uridylated $3B^{VPg}$ ($3B^{VPg}$pUpU) then acts as a primer for initiation of transcription. It is thought that the *cre* acts as the primary site for uridylation of $3B^{VPg}$ by $3D^{pol}$. Using cell-free assays, it was demonstrated for PV that $3B^{VPg}$ could be uridylated by $3D^{pol}$ and this reaction was strongly stimulated by the addition of purified $3CD^{pro}$. Deletion of the *cre* abrogated the synthesis of $3B^{VPg}$pUpU in the presence or absence of $3CD^{pro}$ (Paul *et al.*, 2000). It has, however, been shown that the poly(A) RNA can also act as a site for uridylation (Paul *et al.*, 1998). FMDV does not seem to need the poly(A) as a template, with the *cre* proving critical for uridylation (Nayak *et al.*, 2005). Therefore, it seems differences occur for this reaction between members of the picornavirus family.

Elongation of the negative strand commences once the primer ($3B^{VPg}$pUpU) is translocated to the 3'-end of the positive strand. This process is unknown; however, it is proposed that binding of PABP to the poly(A) tail positions the 3'-end of the positive strand in close proximity to the *cre* to facilitate this process (Grubman & Baxt, 2004). It has been shown for PV that a circular RNP complex forms between proteins bound to both the 5' CL structure and the poly(A) tail. PCBP and $3CD^{pro}$ bind to stem-loops within the cloverleaf (Gamarnik & Andino, 2000; Parsley *et al.*, 1997) and interact with PABP bound to the poly(A) tail

(Herold & Andino, 2001). Deletion of the 5' CL abolishes negative-strand synthesis, and disruption of this structure leads to a decrease in synthesis (Barton *et al.*, 2001; Herold & Andino, 2001). The requirement of this 5' structure for negative-strand synthesis at the 3' end supports the idea of genome circularisation as a necessary step in replication. As described previously, it has been shown that the 3' UTR of FMDV interacts with both PCBP, PABP and the 5' S-fragment (Serrano *et al.*, 2006); therefore, although the precise mechanisms of negative RNA synthesis remain largely unknown, it appears that similarities exist amongst picornaviruses.

Synthesis of negative-strands is thought to produce a double-stranded replicative form (RF) consisting of template and newly-synthesised RNA. The ratio of positive to negative strands is around 50:1 and essentially no 'free' negative strands are found within infected cells (Rozovics & Semler, 2010). A replicative intermediate (RI) is thought to exist, where a single negative strand is the template for the synthesis of numerous positive strands. The details involved in the initiation of positive-strand synthesis are not largely explained, but the cellular RNA binding-protein heterogeneous nuclear ribonucleoprotein C (hnRNP C) and viral proteins 2C and 3C/3CD^{pro} have been implicated as important factors for this process (Ertel *et al.*, 2010), as well as the 3' UTR (Brown *et al.*, 2004) and the 5' CL (Sharma *et al.*, 2005).

Detailed descriptions of genome replication are described in the following book chapters (Kirkegaard & Semler, 2010; Rozovics & Semler, 2010).

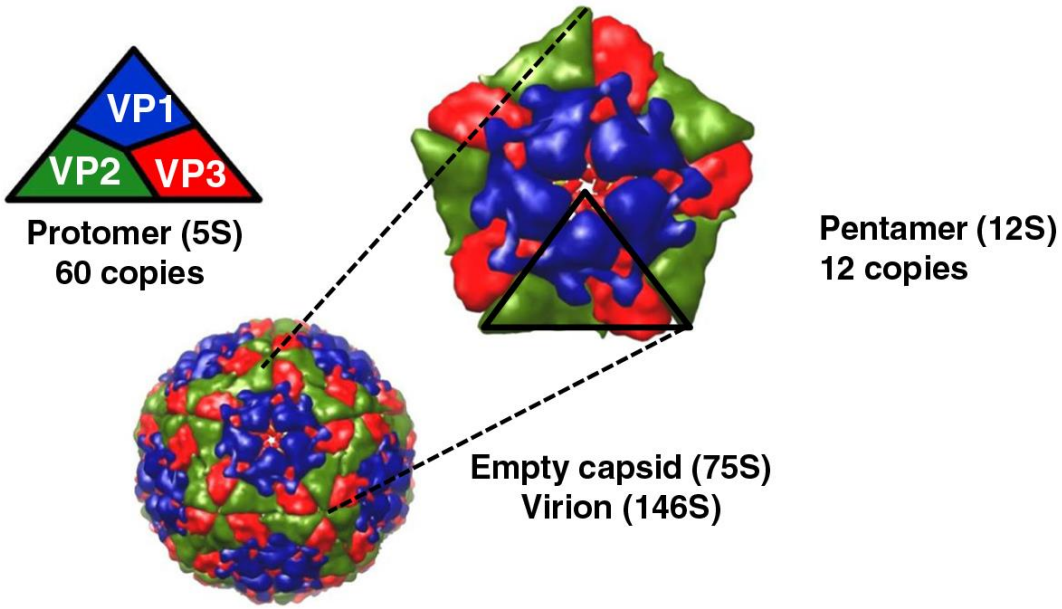


Figure 1.4. FMDV virion structure. The structural proteins 1A (VP4), 1B (VP2), 1C (VP3) and 1D (VP1) arrange into protomeric and pentameric units. These units assemble to form the virion containing 60 copies of each structural protein. 1A is located internally within the capsid and is not shown. Particles lacking an RNA genome have been shown to occur (adapted from Jamal & Belsham, 2013).

1.2.3 Translation and polyprotein processing

As mentioned above, translation of picornavirus RNA is driven by the IRES element. This is in contrast to cellular translation which relies on a cap structure (m^7Gppp) at the 5'-end to initiate translation. Translation therefore proceeds in a cap-independent manner. This is facilitated by proteolysis of specific eukaryotic initiation factors (eIF), favouring IRES-driven translation, and allowing for the use of host cell translation machinery. Initiation of translation *via* the IRES involves recruitment of the translation machinery to an internal region on the viral RNA (Martínez-Salas & Ryan, 2010).

Picornavirus genomes encode a long, single ORF which would be predicted to generate a large polyprotein once translated. This full-length translation product is never observed in cells during infection due to the rapid proteolytic activity of viral encoded proteinases which occur both co- and post-translationally (Figure 1.5). Primary processing consists of co-translational cleavages which separate the capsid precursor from the downstream region encoding the replication proteins. FMDV (and other genera) utilises a novel translational recoding event encoded by the 2A oligopeptide which mediates release at the 2A/2B junction. Other genera within the picornavirus family achieve this through $3C^{pro}$ (hepatoviruses), or $2A^{pro}$ (entero- and sapeloviruses) mediated cleavage (Martínez-Salas & Ryan, 2010; Ryan & Flint, 1997). A second primary processing event occurs at the region between 2C/3A and is performed by $3C^{pro}$ (Ryan & Flint, 1997). An additional processing event involves the release of L^{pro} from the polyprotein *via* auto-catalytic cleavage between its C-terminus and the N-terminus of 1A. These primary cleavages result in the formation of protein precursors which undergo further processing. Generation of "intermediate" and "mature" forms is mediated by $3C^{pro}$.

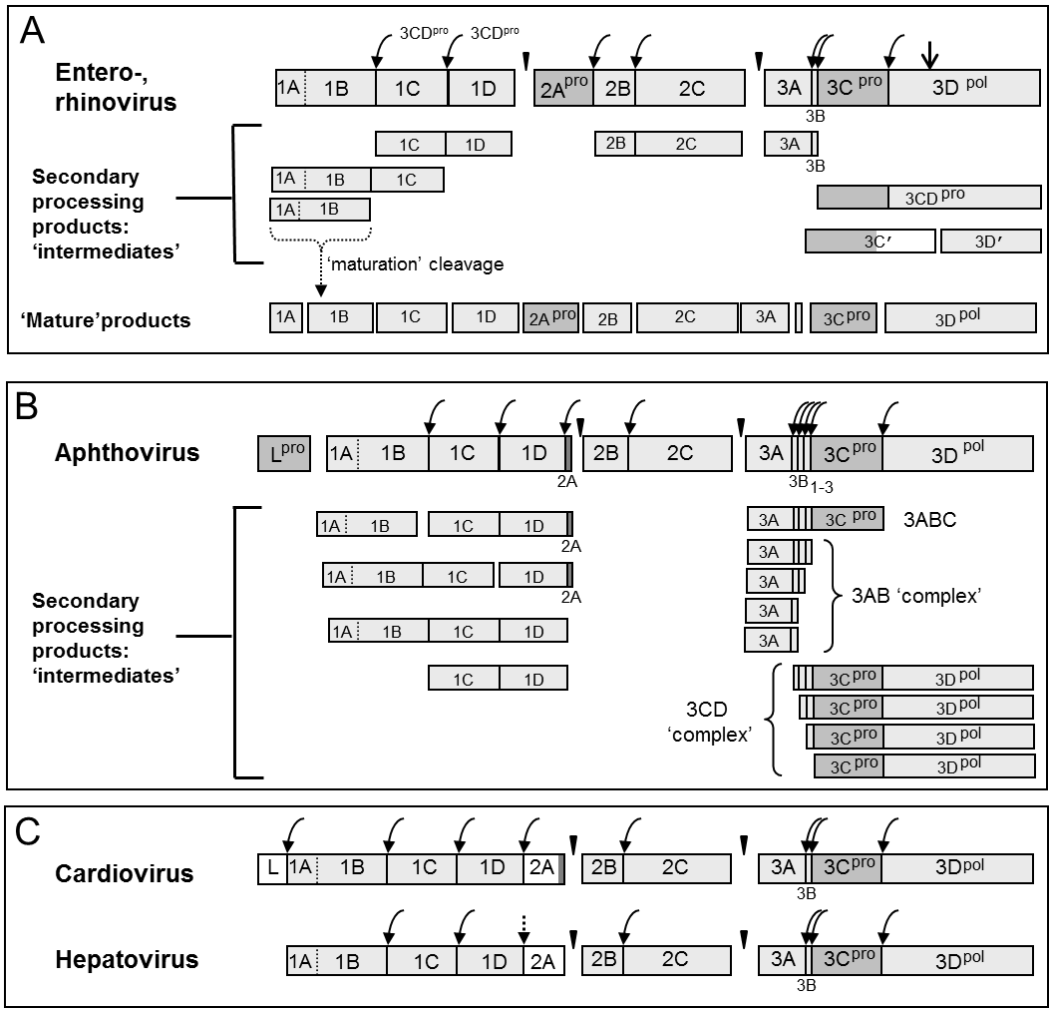


Figure 1.5. Polyprotein processing within the picornaviruses. Primary (black triangles) and secondary cleavages mediated by $3C^{pro}$ (rounded arrows) are shown. (A) The P1 precursor is processed by $3CD^{pro}$, rather than $3C^{pro}$, amongst enteroviruses. (B) Aphthoviruses mediate [P1-2A] processing (into 1AB, 1C, 1D and 2A) by both 3C and 3CD proteinases. The multiple copies of $3B^{VPg}$ within the P3 precursor makes processing a highly complex event as several alternative processing events occur, giving rise to many protein forms. (C) Viruses which lack L or 2A proteinases are processed by $3C^{pro}$ or $3CD^{pro}$. A host cell proteinase is responsible for cleavage of the 1D/2A site within hepatoviruses (arrow with dotted line; image adapted from Martínez-Salas & Ryan, 2010).

1.2.4 Encapsidation and maturation

The processes involved in encapsidation and the maturation 'cleavage' remain largely unknown. Two proposed models of encapsidation exist. The first model implies that empty capsids form (by the assembly of pentamers into the capsid structure; Figure 1.4) and the viral RNA is inserted following this assembly. The second model proposes an interaction between the capsid pentamers and the viral RNA, forming a provirion (a viral particle which contains an uncleaved 1AB (VP0) protein; (Grubman & Baxt, 2004). The maturation cleavage of 1AB (VP0) is the final stage in encapsidation. It is believed that this cleavage is autocatalytic, and is usually observed during encapsidation of viral RNA. This step is thought to involve the activation of water molecules, which leads to nucleophilic attack at the scissile bond. For FMDV, a conserved histidine (2145H within 1B [VP2]) located close to the VP0 cleavage site is responsible for this cleavage reaction (Curry *et al.*, 1997). This 1AB 'cleavage' event is required for particle stability, as interactions are made between regions of the pentamers following release of both 1A and 1B (VP4 and VP2). Additionally, prevention of 1AB (VP0) cleavage *via* mutagenesis results in a loss of infectivity, demonstrating that this event is important for the generation of infectious virus (Knipe *et al.*, 1997).

Figure 1.6 shows a summary of the life cycle which has been described within section 1.2.

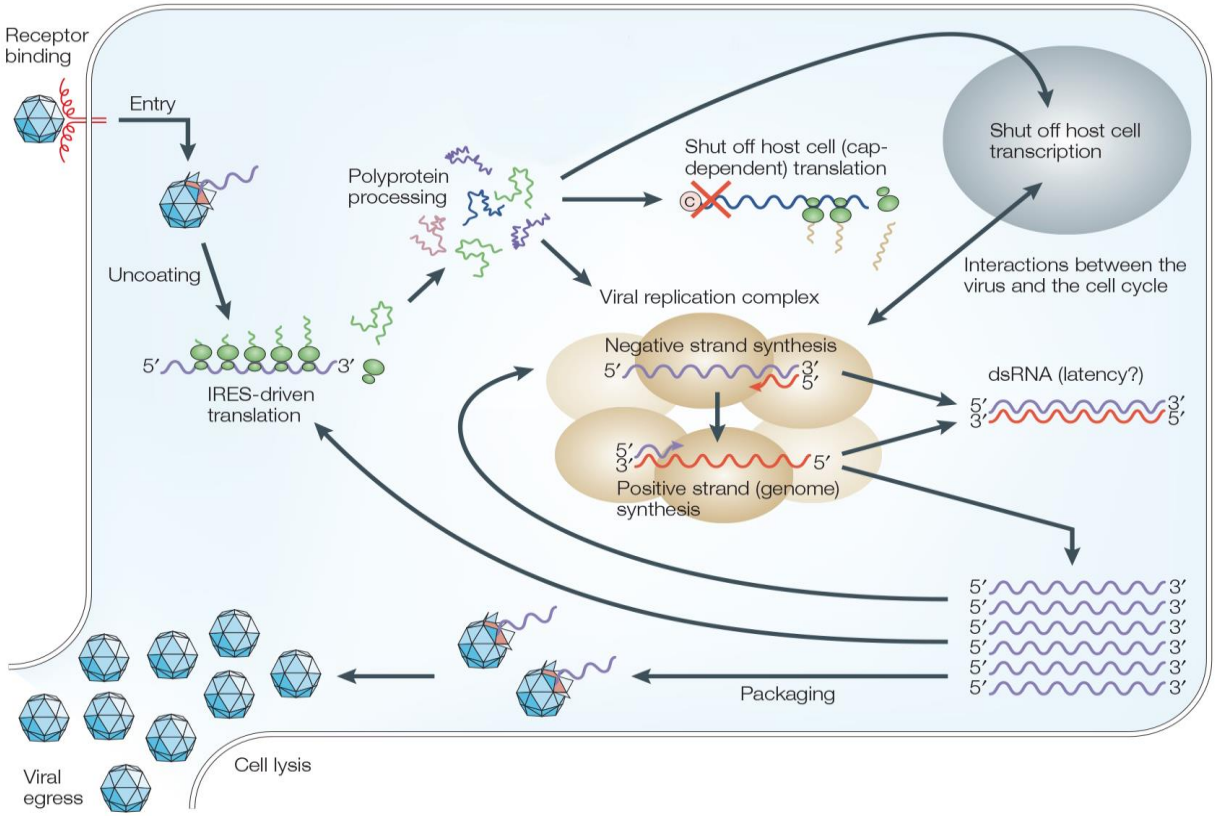


Figure 1.6. Picornavirus life cycle. Major events which occur during viral infection are summarised. Detailed descriptions are provided within the text (image adapted from Whitton *et al.*, 2005)

1.3 Foot-and-mouth disease

FMD is one of the most important animal diseases known and currently affects many parts of the world. The disease affects cloven-hoofed animals such as cattle, sheep, pigs, goats and ~70 species of wildlife including African buffalo and deer (Grubman & Baxt, 2004; Jamal & Belsham, 2013). In European strains of domestic animals, FMDV rapidly replicates and spreads, with signs of disease appearing after 2-3 days. In buffalo, however, strains SAT 1, 2 and 3 cause inapparent, persistent, infections. Clinical features of the acute form of the disease include fever, viremia, excessive salivation and vesicular lesions which form on the oral mucosa, nose and feet. The virus is excreted into bodily fluids such as semen, urine and in the milk of dairy cattle as well as in faeces. FMD is rarely fatal in adult animals; however it can adversely affect the heart function of young animals and may cause high mortality (Grubman & Baxt, 2004). The clinical signs of FMD are similar to other diseases, such as SVDV and VSV, and therefore it is necessary for diagnosis to be made using laboratory methods. Current diagnostics include the "gold standard" virus neutralisation test (VNT), ELISA, qRT-PCR and nucleotide sequencing. Determination of the serotype involved in outbreaks is crucial for the implementation of proper control strategies (Jamal & Belsham, 2013).

FMD is highly infectious and transmission occurs through many different routes. The most common means of transmission are direct contact between infected and susceptible animals, through distribution/disposal of contaminated food products or by indirect contact through contaminated personnel, vehicles and aerosols (wind-borne). Illegal activities have also contributed to spread of the disease by the importation of infected meat, illegal movement of animals and the feeding of non-heat treated swill to pigs (Sutmoller *et al.*, 2003).

Following the acute stage of infection FMDV can cause persistent infection in ruminants, which has no clinical signs i.e. the carrier state. Carrier animals are defined as animals which are virus positive for 28 days or more post infection (Grubman & Baxt, 2004; Sutmoller *et al.*, 2003). These animals show no clinical symptoms of the disease and can remain in this state for a long period of time (2-3 yrs in domestic cattle; Jamal & Belsham, 2013). The carrier state is characterised by isolation of live virus from the pharynx and can also occur in vaccinated animals. The site of this persistent viral replication is proposed to be the light zone of germinal centres (GC). Tissues (tonsils, lymph nodes and soft

palate) isolated from cattle 38 days post-infection were shown to harbour FMDV genomes and were restricted to the GCs (Juleff *et al.*, 2008). The role of carrier animals in the spread of virus in the field is still largely unknown. It is possible that these animals may facilitate in the spread of disease, but this has only been shown for African buffalo during an outbreak, and has not been demonstrated in experiments with cattle (Grubman & Baxt, 2004; Jamal & Belsham, 2013). The mechanisms underlying persistent infection are not well understood, but it has been proposed that the immune status of the animal has a direct role (Grubman & Baxt, 2004). Due to the occurrence of persistent infection, it is difficult to differentiate between infected and vaccinated animals (DIVA). This causes problems with trade, as it is important to be able to distinguish between previously infected and those that have been vaccinated. Vaccination causes the production of antibodies raised against the coat protein only. When animals are infected by field isolates of the virus, replication produces both structural and non-structural proteins (NSP), therefore it is possible to discriminate between the antibodies induced by vaccinated and those induced by infected animals. However, repeated vaccination using vaccine that has not been purified, and therefore contains contaminating NSPs, can affect serological tests undertaken (Jamal & Belsham, 2013; Sutmoller *et al.*, 2003).

As mentioned previously, FMDV exists as seven different serologically distinct serotypes, and various subtypes occur within each serotype. These serotypes are distributed geographically with serotypes O, A and C having the widest distribution. The SAT 1-3 viruses are usually limited to sub-Saharan Africa and Asia-1 viruses are normally restricted to Asia; these viruses have, however, caused limited outbreaks in other regions. There is a tendency for the virus to recur in the same geographical area, despite the propensity for it to spread to surrounding areas and cause small outbreaks. It is thought this is due to ecological isolation which may reflect patterns of animal movement and trade, and has resulted in the creation of seven regional pools (Figure 1.7). There are more than 100 countries affected by FMD worldwide. The distribution of the disease reflects economic development, with more developed countries having disease-free status. The occurrence of FMD poses a potential threat to countries who have eradicated the disease.

Financial losses caused by an FMD outbreak are high due to loss of meat, milk and livestock. Eradication and control costs, as well as indirect costs associated with tourism and trade restrictions, contribute to these losses (Grubman & Baxt,

2004). The recent 2001 outbreak in the UK was estimated to cause losses of 12.3-13.8 billion USD. This outbreak also resulted in the culling of 6.5 million animals. Large outbreaks in recent years have occurred in Taiwan, Japan and South Korea all contributing to huge economic consequences (Jamal & Belsham, 2013).

1.3.1 Disease control

Until 1991, continental Europe (except Denmark) systematically vaccinated against FMDV. Britain and Ireland, however, did not undertake this strategy and as a result these countries had certain trading advantages over other European nations. The European Union therefore decided to abolish general vaccination to harmonise trading positions (Sutmoller *et al.*, 2003). Benefits to stopping routine vaccination include reduction in costs associated with vaccine administration, less trade restrictions and problems associated with vaccinating animals such as immune surveillance, dispersed herds and persistent infections in vaccinated animals. Countries currently control FMD outbreaks by killing and destroying all infected animals and inhibition of animal movement, with ring vaccination employed during extreme cases.

1.3.1.1 Vaccines

Of the picornavirus family members, only three agents have commercial vaccines available; PV, HAV and FMDV. Initial experiments demonstrating protective immunity against these viruses involved the use of immune serum (FMDV) or immunoglobulins (PV and HAV). These procedures were, however, impractical on a large scale. Current picornavirus vaccines are produced through the growth of virus in tissue culture under stringent bio-containment conditions – dramatically increasing the cost of vaccine production and storage. The final product must be free from cell culture contaminants or infectious agents. These days, inactivation (innocuity) and testing (efficacy) is better regulated and such contaminants are less likely to occur. The escape of virulent virus from vaccine production facilities, however, still remains a large problem with current production strategies (Doel, 2003).

Attenuated vaccines consist of viruses which replicate to a minimal level without inducing clinical symptoms of the disease. As replication imitates natural

infection, a high dose is unnecessary as amplification occurs *via* the altered agent. Desirable advantages include the immune response elicited following vaccination as full activation of the relevant immune components occurs. The main risk associated with attenuated vaccines is the potential for reversion to virulence, with uncontrolled replication or persistence also factors to be considered. Inactivated preparations contain viral particles which are unable to replicate. The does required is much higher than that of attenuated vaccines due to the antigen load needed to produce an adequate immune response (Rowlands & Minor, 2010).

1.3.1.1.1 PV vaccines

The live and inactivated vaccines produced against poliovirus are the most well documented of those mentioned thus far. Like FMDV, PV exists as distinct serotypes; 1, 2 and 3, and again, like FMDV, protection against one serotype does not confer protection against others. Vaccine preparations therefore contain a strain from each serotype.

The inactivated polio vaccine (IPV) developed by Salk was made by inactivating PV harvested from infected tissue culture cells with low levels of formalin for prolonged periods of time. This was performed to ensure the virus was killed without destroying its antigenicity, as PV particles are easily disrupted through chemical stress such as heating, UV etc. This vaccine preparation is still used at present and has been very effective (Rowlands & Minor, 2010). As mentioned previously, there are risks associated with chemically inactivating large amounts of pathogenic virus. This was documented for IPV following successful vaccination throughout the USA. Cases of poliomyelitis were reported and it was soon realised that these cases arose from insufficient chemical inactivation of the vaccine. Termed the 'Cutter Incident' after the manufacturer involved, it lead to 94 cases of poliomyelitis and ~160 contact cases (Nathanson & Landmuir, 1995).

The attenuated oral polio vaccine (OPV) developed by Sabin was produced by passaging viruses in tissue culture cells at high or low MOI and suboptimal temperatures. This vaccine was used as the basis for the polio eradication program and helped to stop this disease from remaining a public health problem. Vaccine-derived cases of poliomyelitis have occurred using this vaccine, and it is now well documented that this vaccine is able to revert to neurovirulence. For example, for all three serotypes, the main neurovirulent attenuating mutation is a

single point mutation which lies in domain V of the IRES within the 5' UTR. It has been shown that this mutation leads to a reduction in translation efficiency causing an attenuated phenotype (Ochs *et al.*, 2003; Rowlands & Minor, 2010).

1.3.1.1.2 FMDV vaccines

There have been several tried, and failed, attempts to generate a live-attenuated vaccine for FMD. The genetic stability of the attenuated viruses was not sufficient enough to be used within the field without the risk of reversion to virulence.

The current vaccine for FMD exists as an inactivated vaccine preparation. The first functional inactivated vaccines involved chemical inactivation with formaldehyde. Virus was isolated from vesicular fluid extracted from lesions on the tongues of purposely infected cattle. The first *in vitro* production method involved work by Frenkel who infected suspensions which were produced from tongue harvested from healthy cattle (Rowlands & Minor, 2010). This finding demonstrated the potential for large scale production of FMD vaccine, however the breakthrough for FMD vaccine production was made by Mowat & Chapman, (1962) who revolutionised production methods through the growth of virus in BHK-21 cells. Initially, cells were grown in monolayers; however this was soon replaced by growth in suspension allowing vaccine production involving large-scale tissue culture fermenters (Telling & Elsworth, 1965). The vaccine production process is summarised in Figure 1.8 (Doel, 2003). As the process involves chemical inactivation of pathogenic virus, vaccine production takes place within a high security biocontainment facility. Formaldehyde has been replaced by ethylene imines due to formaldehyde inactivation often proving incomplete (as demonstrated by the Cutter Incident). BHK-21 suspension cells grown in large stainless steel vessels are infected with virus and non-structural proteins/cell debris are removed using centrifugation or filtration. Binary ethylenemine (BEI) is mixed with the virus for the inactivation process. The inactivated antigen is then purified using ultrafiltration and chromatography, which is stored in the gaseous phase of liquid nitrogen units until required (Doel, 2003).

The multiplicity of FMDV serotypes and subtypes add an element of complexity to FMD vaccine production, with vaccine preparations being tailored to specific world regions/countries (virus pools, Figure 1.7). Adjuvants are required within FMD vaccine preparations as they are essential for sufficient potency. Aluminium hydroxide and saponin are widely used for ruminants with oil based adjuvants

being preferred for vaccination of pigs. Other drawbacks associated with the current vaccine include; poor induction of long-term immunity (especially within pigs) with repeated boosters required, short shelf life and a sufficient cold chain is required (Rodríguez & Grubman, 2009).

A rapid humoral response is produced in both infected and vaccinated animals. Antibodies specific to FMDV protect animals against re-infection, or infection in the case of vaccinated animals. This protection is normally correlated with a high level of neutralising antibodies and the literature suggests that both macrophage and antibody activity may be required (Mccullough *et al.*, 1992).

The role of cellular immunity in protection from FMD is still a matter of some debate. T cell responses, $CD4^+$ and $CD8^+$, have been observed in either infected or vaccinated animals and a correlation between antibody and T cell responses has been reported (Grubman & Baxt, 2004), however, more work is required to determine the importance of cellular immunity in protection.

Due to the disadvantages of the current commercial vaccine available for FMD, alternative strategies are required to improve the security-associated risks (e.g. the 2007 UK outbreak), as well as broadening the immune response to include cell-mediated responses.

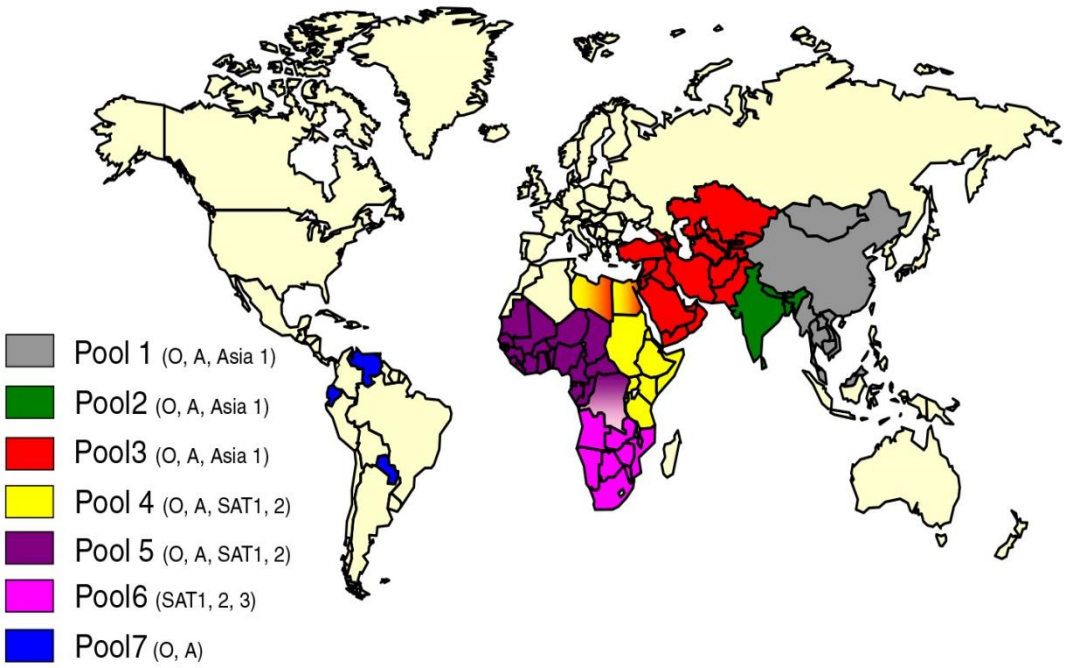


Figure 1.7. FMDV virus pools and their geographical distribution. The most widely distributed serotype is O and is found in 6 out of the 7 pools. Serotypes Asia-1 and SAT 1-3 are the most geographically limited. FMDV-free countries are shown in off-white (adapted from Jamal & Belsham, 2013).

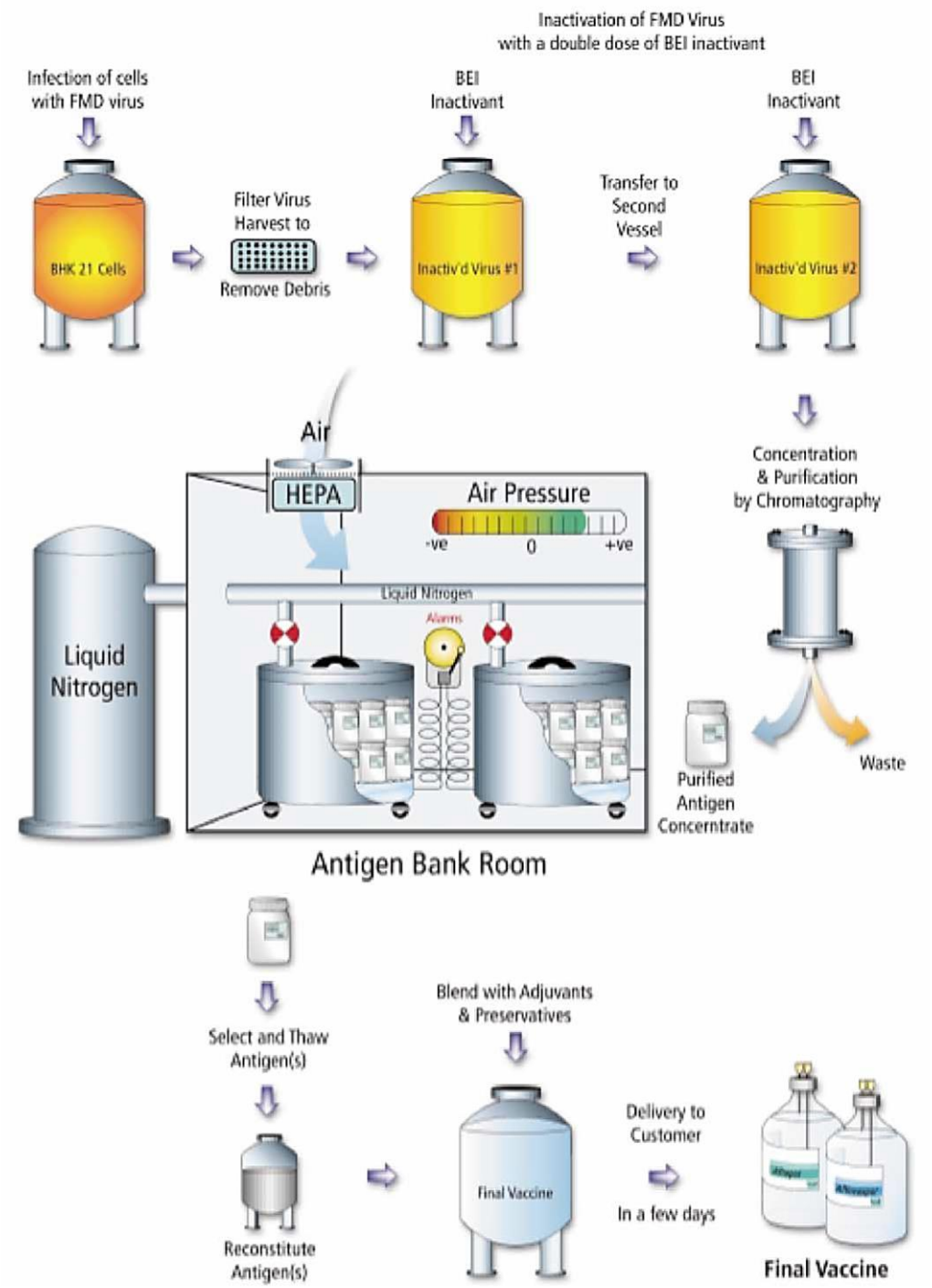


Figure 1.8. FMD vaccine production. Highly pathogenic virus is grown in large fermenters, chemically inactivated, and purified to remove adventitious agents (figure taken from Doel, 2003).

1.4 Reverse genetics systems

Reverse genetics systems are important as they serve as tools to modify regions of the viral genome whereby the effects of such modifications can then be studied. Manipulation of the genome can be performed in a strategic manner to create pre-determined alterations and investigate the effects of such changes. This helps to gain insight into viral functions and the use of derivatives that differ from WT for vaccine development and characterisation of the life cycle. They also decrease the need for stringent, high-containment facilities, where genome alterations make resulting virus that is less (or completely) un-infectious, and therefore safer to handle outwith BSL3/4 labs.

1.4.1 Reverse genetics and synthesis of viral RNA

The first reverse genetics system was invented in the 1970s by Charles Wiseman and colleagues (Taniguchi *et al.*, 1978). Using reverse transcriptase, Wiseman and colleagues made a double-stranded cDNA from purified phage Qβ RNA. Once this cDNA was inserted into a plasmid it produced Qβ phage when transformed into bacteria (Taniguchi *et al.*, 1978). Viral RNA can now be prepared readily for a wide range of RNA viruses. Generation of reverse genetics systems has been more difficult for negative strand RNA viruses due to reasons such as; the need for precise 5’ and 3’-ends for replication and genome packaging, the viral polymerase being essential for transcribing both the genomic and antigenomic RNA, and due to genomes forming RNP complexes. Reverse genetics systems for negative strand RNA viruses can comprise full-length cDNA plasmids (unsegmented genomes) or multiple plasmids (segmented genomes) which contain promoters such as phage T7 RNA polymerase or endogenous promoters, RNA polymerase I or II. Plasmid DNA is transfected into cells producing viral particles. For those systems which use T7-based promoters, T7 RNA polymerase must be supplied *in trans* through plasmid-based expression, cell lines constitutively expressing T7 RNA polymerase, or through superinfection with a vaccinia virus vector encoding the T7 polymerase. Some systems also require the cDNA plasmids to be transfected along with expression plasmids for the viral proteins involved in forming the ribonucleoprotein (RNP) complex (reviewed in Hoenen *et al.*, 2011; Pekosz *et al.*, 1999).

Rescue of positive sense RNA viruses from cDNA plasmids has proven more straight forward than those from the negative sense. Viral transcripts are made

using the corresponding bacterial polymerase and these transcripts are transfected into cells. This cDNA can be genetically manipulated and the phenotype observed from rescuing the corresponding infectious virus.

The first publication demonstrating a eukaryotic RNA virus reverse genetics system was from Racaniello and Baltimore (Racaniello & Baltimore, 1981), who obtained authentic PV from a cDNA plasmid encoding the PV genome. When this plasmid was transfected into HeLa cells PV was produced at very low yields. This system was improved a few years later *via* the insertion of simian virus 40 (SV40) transcription and replication signals which increased virus yield by 18-250-fold depending on the cell type used (Semler *et al.*, 1984). The use of bacterial polymerases such as *Salmonella enterica* serovar Typhimurium phage SP6 or phage T7 RNA polymerase to transcribe viral cDNA from plasmids was a further advance in this field for a range of picornaviruses such as Coxsackie B3, Cardioviruses, Human Rhinovirus 14, Hepatitis A virus and FMDV (Cohen *et al.*, 1987; Duke & Palmenberg, 1989; Kandolf & Hofschneider, 1985; Mizutani & Colonno, 1985; Zibert *et al.*, 1990). A study by Kaplan *et al*., involved transcription of PV RNA in either direction using the SP6 RNA polymerase. These transcripts were not initially infectious (Kaplan *et al.*, 1985). Infectious transcripts could only be achieved by employing a strategy that involved PV replicase (now known as $3D^{pol}$), which used negative strands transcribed by SP6 RNA polymerase as template for positive strand synthesis. When these transcripts were transfected into HeLa cells the yield was very poor in comparison to isolated PV RNA. In the same year, another study was published which used T7 RNA polymerase to generate PV transcripts in both orientations (van der Werf *et al.*, 1986). Positive-sense transcripts were much more infectious than the plasmid alone; however they were only ~0.1% as infectious as purified viral RNA (van der Werf *et al.*, 1986). Transcripts produced this way contained non-viral nucleotides at both termini of the PV transcript RNA introduced by the bacterial polymerase (Clark, 1988; Milligan *et al.*, 1987) and through cloning. Removal of these extra sequences; leaving 2 guanine residues after the T7 promoter and no more than seven nucleotides (nts) past the poly(A), resulted in transcripts with increased infectivity: 5 % that of viral RNA (van der Werf *et al.*, 1986). Removal of large 5' sequences therefore helped to increase infectivity: SP6 negative strand PV transcripts made by Kaplan and colleagues contained 190 and 82 nts at the 5' and 3'-ends respectively. This explains the low infectivity observed upon transfection of these transcript RNAs.

1.4.1.1 Improvement of reverse genetics systems using autocatalytic ribozymes

As mentioned in the previous section, the effect of nonviral nucleotides at the ends of transcript RNA can decrease infectivity of viral transcripts. These nucleotide extensions arise from additions by bacterial polymerases (Clark, 1988; Milligan *et al.*, 1987) and through cloning techniques used to insert the viral cDNA into the vector backbone. The importance of removing extensions from the 5'-end of positive strand RNA genomes has been demonstrated for Sindbis and Semliki Forest virus (Liljeström *et al.*, 1991; Rice *et al.*, 1987) and the picornaviruses CVB3, Echovirus 5 and PV (Herold & Andino, 2000; Israelsson *et al.*, 2014; Klump *et al.*, 1990). When a self-cleaving hammerhead (HH) ribozyme (Uhlenbeck, 1987) was inserted between the T7 promoter and the PV 5' UTR, transcripts produced had authentic ends and replicated more efficiently in 293 cells with no lag in replication (Herold & Andino, 2000). This was a vast improvement on the infectivity and fitness (compared to isolated viral RNA) achieved by van der Werf *et al*., (1986) when PV transcripts contained two guanine residues at the 5' terminus. These transcripts were also able to replicate efficiently within *Xenopus laevis* oocytes which had not been achieved previously.

Extensions to the 3'-end of positive strand RNA viruses such as Kunjin and Sindbis virus can have a profound affect on replication/infectivity (Dubensky *et al.*, 1996; Khromykh *et al.*, 2003). Picornaviruses, however, can tolerate 3' additions, unlike extensions to the 5'-end of the genome. Sarnow (1989) studied the effect of 3' extensions on the infectivity of PV transcript RNA. The addition of 17 cytosine residues (after the poly(A) tail) decreased infectivity, while a reduction to 4 residues increased the infectivity to 10% of virion RNA. A precise 3'-end is critical for rescue systems coding for negative sense RNA viruses as this contains important replication signals, as well as the importance of the "rule of six" where the N protein must cover exactly 6 nt: if this is not followed, the RNA is poorly replicated or not at all. An exact 3'-end is also thought to be important for accessibility by the polymerase for RNA synthesis *via* "open ends" (Ghanem *et al.*, 2012; Pattnaik *et al.*, 1992). Introduction of the hepatitis delta virus (HDV) ribozyme (Perrotta & Been, 1991) into a rabies virus (RV) reverse genetics system improved virus rescue 10-fold, with a 100-fold increase achieved by combination of both HH and HDV ribozymes, demonstrating the considerable effect ribozymes can make to reverse genetics systems.

1.4.1.2 The use of reporter proteins in mini-genome/replicon systems

Mini-genome or mini-replicon systems can also be used to study characteristics of the viral life cycle. These include a reporter protein such as GFP, luciferase or chloramphenicol acetyltransferase (CAT) in place of, or in frame with, viral coding sequences (Albariño *et al.*, 2013; McInerney *et al.*, 2000; Mueller & Wimmer, 1998). Mutational analysis of viral genome regions can be monitored by reporter protein analysis to determine their effects. Attenuation for vaccine development and the effect of antivirals on viral replication can also be monitored through the use of such systems.

Using fluorescently-tagged viruses as tools to further characterise the viral life cycle has been investigated for various RNA viruses. Viral proteins can be tagged and monitored over time, or the virus particles themselves can be fluorescent. Attempts to do this using FMDV (Seago *et al.*, 2013), PV (Mueller & Wimmer, 1998) and CVB3 (Tong *et al.*, 2011) have proven to be quite unsuccessful. In both studies the GFP (and *Renilla* luciferase in the case of FMDV) sequence was rapidly deleted after serial passage (~P4), and replication was impaired by the insertion of GFP into the viral genome sequence. Packaging limits of the viral capsid and stability of the RNA genome following GFP insertion were thought to be the attributing factors to the results seen by both groups. Tagging negative sense RNA viruses has been more successful with Marburg (Albariño *et al.*, 2013) and Influenza viruses (Manicassamy *et al.*, 2010) both producing recombinant tagged virus. Thus, as tagging of positive sense RNA viruses such as PV and FMDV has proven to be a challenge, mini-replicon systems are a suitable alternative.

Experiments investigating defective interfering (DI) particles of PV demonstrated that packaged virus genomes lacking the capsid coding region could infect and replicate within cells, but could not synthesise capsid proteins (Kajigaya *et al.*, 1985; Lundquist *et al.*, 1979; Nomoto *et al.*, 1979). Sequence analysis of these DI genomes showed that the correct reading frame for protein synthesis was retained (Kuge *et al.*, 1986). DI genomes were packaged by capsid proteins from co-infecting wild-type viruses which supplied these proteins *in trans*. Following these findings, PV sub-genomic replicons were constructed which contained in-frame deletions within the capsid coding region. Transcript RNA synthesised from these replicons were shown to replicate once transfected into HeLa cells (Kaplan & Racaniello, 1988). This evidence allowed researchers to insert reporter

proteins in place of the capsid region to study aspects of replication using various assays. Such mini-replicon systems have helped to determine important elements of the picornavirus life cycle. Luciferase, CAT and GFP replicons have been used in experiments which lead to the discovery of the *cre* within HRV and PV, and elements of PV and FMDV replication (Andino *et al.*, 1993; Forrest *et al.*, 2014; Goodfellow *et al.*, 2000; Herold & Andino, 2000; McInerney *et al.*, 2000; McKnight & Lemon, 1996; Percy *et al.*, 1992).

The generation of full-length cDNA copies of a viral genome does not always mean successful viral replication or rescue. A prime example of this is the difficulties encountered with Hepatitis C Virus (HCV). Tissue-culture propagation of isolates and cloned viral RNA proved unsuccessful. Full-length clones were generated, but the replication efficiency was extremely low; detection of viral replication was only possible using RT-PCR (Blight & Norgard, 1996) – no viruses could be rescued. The use of animal models to propagate virus was also limited: chimpanzees are the only animal where reproducible results have been obtained (Bartenschlager & Lohmann, 2000). The creation and development of a HCV subgenomic replicon which could replicate within Huh-7 cells was a breakthrough in this field (Lohmann *et al.*, 1999). This replicon encoded the selectable marker neomycin phospho-transferase (Neo), and following infection of cells with transcript RNA, cells resistant to G418 were found to contain replicons which had replicated to high levels. A novel isolate of HCV (JFH-1) which supports efficient replication and virus production is now used to study HCV; however, until this discovery, replicons played a very important part in HCV research (Bartenschlager, 2006).

Thus, using replicon systems permits viral replication to be studied using various assays and have proven critical in certain cases - such as HCV. The use of fluorescent proteins as a reporter for replication has certain advantages in that replication can be studied in live cells using sensitive microscopy techniques: a substantial improvement over methods which rely upon assays of cell extracts.

1.5 Aims of the Project

The overall aim of this study was to generate a FMDV replicon system which utilised *Aequorea victoria* GFP as a reporter of replication, such that replication could be monitored indirectly *via* measurement of GFP fluorescence in live cells. The capsid coding region of the genome is to be replaced by a bi-functional puromycin-acetyl transferase-GFP fusion protein ([GFP-PAC]), to allow for the selection of transfected cells as well as a marker for replication. This GFP-PAC replicon could be used to study various aspects of the FMDV life cycle without the need for high-containment facilities. Differences in replicating and non-replicating forms will be investigated during this project to confirm the use of the system for measuring attenuated genomes as potential vaccine candidates.

Chapter 2: Materials and Methods.

2.1 Materials

2.1.1 Oligonucleotides

All primers used during this study were purchased from Integrated DNA Technologies (IDT) and are listed in Table A.1.

2.1.2 Gene blocks

A gene block containing genes for Bovine Enterovirus 2A and *Ptilosarcus gurneyi* GFP was obtained from Dundee Cell Products (DCP). Echovirus 7 region 1 and 2 gene blocks were purchased from Eurofins.

2.1.3 Plasmids

All replicons generated were made by altering the pGFP-PAC replicon (described in Chapter 3), which was created by modifying the original FMDV CAT Replicon (pT7rep, McInerney *et al.*, 2000). The pSP70 and pSP72 plasmids were obtained from Promega and used in the construction of the pPt.GFP and pBEV2A replicons, respectively. The pcDNA 3.1(+) plasmid was purchased from Invitrogen and used to generate all replicons described in Chapters 3-6.

2.1.4 Mammalian cells

All cell lines were obtained from either the European Collection of Cell Cultures (ECACC) or the American Type Culture Collection (ATCC).

2.1.4.1 Cell lines used in this study

BHK-21 (clone 13): Baby hamster (Syrian) kidney cell line.

RD: Human rhabdomyosarcoma cells (provided by Professor Peter Simmonds, University of Edinburgh).

BHK-21_RFPNuc: Baby hamster kidney cells stably expressing Red FP (RFP) in the nucleus (generated by Mr John Nicholson, University of St Andrews).

2.1.5 Antibodies

Primary antibodies: The polyclonal anti-2A antibody was purchased from Dundee Cell Products. The anti-eIF4G antibody was kindly provided by Professor Lisa Roberts, University of Surrey, and the anti-FMDV $3D^{pol}$ was provided by Professor Nicola Stonehouse, University of Leeds. Monoclonal antibodies specific to β-tubulin and GFP were purchased from Invitrogen and Roche, respectively.

Secondary antibodies: Anti-mouse IgG and anti-rabbit IgG HRP-conjugated antibodies were purchased from Invitrogen.

2.1.6 Viruses used in this study

E7: Echovirus 7 (isolate Wallace (accession number AF465516) was provided by Prof. Peter Simmonds, Roslin Institute, University of Edinburgh.

Mutant E7 viruses were generated which contained two modified regions of the E7 WT genome to study dinucleotide frequencies and CPB on viral replication. Region one was located in the capsid coding region at 1878-3119 and region two at 5403-6462, which is situated within P3 in the genome.

culcu: A CpG/UpA low mutant, with all CpG dinucleotides and as many UpA dinucleotides removed.

PlP: Codon order was permuted while retaining protein coding and native dinucleotides.

Max-U: Maximised CPB score, with dinucleotide frequency similar to WT E7.

Min-E: Minimum CPB score, with dinucleotide frequency identical to WT E7.

Min-U: CPB score identical to Min-E, with a slightly elevated dinucleotide frequency.

Min-H: CPB score identical to Min-E, with a maximised dinucleotide frequency.

2.2 Methods

2.2.1 Computer analysis

2.2.1.1 DNA sequencing and sequence alignments

Plasmid DNA generated *via* cloning was analysed by sequence analysis using the SeqMan Pro software (Blattner, 2014). Sanger sequencing data was aligned with a reference sequence (compiled using sequences listed in A.2-3) to ensure the correct order and composition of the DNA sequence.
All other sequence alignments were generated using DNAMAN bioinformatics software (Lynnon Biosoft, 2001).

2.2.1.2 RNA secondary structure prediction

Models of the RNA structure from reference sequences were generated using mfold v.3.6 (Zuker, 2003) and RNADRAW (Matzura & Wennborg, 1996).

2.2.1.3 Protein structure prediction

Amino acid sequences were compiled using DNAMAN and the protein structure predicted using the Phyre^2 software (Kelley & Sternberg, 2009). PDB files generated were manipulated using PyMOL (Schrodinger, 2010).

2.2.2 Polymerase chain reaction

The polymerase chain reaction (PCR) was used to amplify specific DNA fragments for cloning purposes. A typical PCR reaction was carried out in a 50 µl volume containing; 1 X Go *Taq* PCR Master Mix (1.5 mM $MgCl_2$), 0.2 mM each dNTP (Promega), 0.2 µM of each primer (Table A.1), template DNA (10-50 ng

plasmid DNA), 1.25 units (U) of *Taq* DNA Polymerase (Promega) and nuclease-free water up to 50 µl.

Typical reactions were performed using a Veriti 96-well thermal cycler (Applied Biosystems) with the following cycling parameters: initial denaturation at 95 °C for 2 minutes (min) and 35 cycles of; denaturing at 95 °C for 30 seconds (sec), annealing at 60 °C for 30 sec, extension at 72 °C for 1 min/kb of DNA template, and a final extension at 72 °C for 5 min. The annealing step was typically carried out at 60 °C, but was dependent on the melting temperature (T_m) of each primer. Samples were then held at 4 °C indefinitely. PCR reactions were checked on an agarose gel and the product purified for further use (see 2.2.4).

2.2.2.1 Mutagenesis PCR

Mutagenesis PCR was performed in a total volume of 100 µl using *Pfu* DNA polymerase (Promega). Reactions contained the following; 1 X *Pfu* DNA polymerase buffer (2 mM $MgSO_4$), 0.2 mM each dNTP (Promega), 0.1 µM of each primer (Table A.1), template DNA (10-50 ng) and 1.25 U of *Pfu* DNA polymerase. Cycling parameters consisted of; initial denaturation at 95 °C for 2 min and 20 cycles of; denaturing at 95 °C for 30 sec, annealing at 60 °C for 45 sec, extension at 72 °C for 2 min/kb of DNA, and a final extension at 72 °C for 7 min. Samples were then held at 4 °C indefinitely. PCR reactions were treated with 1 U of DpnI (Promega) to remove the plasmid DNA template and a 5 µl aliquot run on an agarose gel to check correct amplification. Five µl of each PCR reaction was combined and transformed as described (see 2.2.6).

2.2.2.2 Combined reverse transcription PCR

Reverse transcription PCR (RT-PCR) was used to amplify cDNA from viral RNA, followed by PCR amplification in a one-step reaction. This was performed using a Superscript III One-Step RT-PCR System with Platinum *Taq* DNA Polymerase kit (Invitrogen). Reactions contained; 10 µl 2 X reaction mix, 6 µl RNA, 0.5 µM each gene-specific primer (Region 1, Table A.1), 0.8 µl Superscript III RT/Platinum Taq Mix and made up to 20 µl with nuclease-free water. Cycle parameters include a cDNA synthesis step followed by PCR amplification. These consisted of: 1 cycle at 43 °C for 1 hr and 20 cycles of; 53 °C for 1 min, 55 °C 1 min and 1 cycle at 70 °C for 15 min. PCR amplification was as follows: initial denaturation at 94 °C for 2

min and 40 cycles of; denaturing at 94 °C for 30 sec, annealing at 50 °C for 30 sec, extension at 68 °C for 1 min/kb of template DNA, and a final extension at 68 °C for 5 min. Samples were held at 4 °C indefinitely. Products were checked on an agarose gel to verify successful amplification, before further use in determining competition assay virus pairs (see 2.2.20).

2.2.3 Restriction enzyme digests

Restriction enzyme (RE) digests of DNA plasmids were performed under conditions specified by the supplier (Promega; New England Biolabs). Generally, 2-5 µg of DNA was digested in the appropriate enzyme buffer, with 1 U of enzyme per µg of DNA, in a volume of 50-100 µl for 1 hr at 37 °C.

2.2.4 Agarose gel purification of DNA fragments

RE digest products were run on a 0.8-1 % (w/v) agarose (Bioline) gel in 1 X TAE buffer (40 mM Tris, 1 mM EDTA). Ethidium bromide was added to a final concentration of 0.4 µg/µl. Samples were mixed with the appropriate volume of 6 X loading buffer (Promega) and ran at 100-120 V in 1 X TAE for ~1 hr. A known-size 1 kb DNA marker (Thermo Scientific) was run along with the samples to determine the band of interest. DNA bands were visualised with UV light, excised with a clean scalpel, placed into a sterile microfuge tube and weighed. The DNA was extracted with a Gel/PCR Purification kit (Biomiga) according to manufacturer's instructions.

2.2.5 Ligation of DNA fragments

Vector and insert DNA was prepared as described above in sections 2.2.3 and 2.2.4. Blunt-end DNA fragments were dephosphorylated with calf intestinal alkaline phosphatase (Promega) prior to ligation to prevent self-ligation of the vector DNA. Varying molar concentrations of vector:insert DNA; 1:1, 1:3 and 1:5, were typically used in a final volume of 10 µl. One unit of T4 DNA ligase was used in each ligation reaction with the 10 X buffer supplied (Promega) and nuclease-free water. The reaction was incubated overnight at 4 °C and transformed into competent *E.coli* cells as described (see 2.2.6).

2.2.6 Transformation of plasmid DNA

Competent JM109 *E.coli* cells (Promega) were used to transform plasmid DNA, ligation reactions and mutagenesis PCR products. Typically, 1-2 µl of plasmid DNA, the entire ligation reaction, or 10 µl of a mutagenesis PCR reaction, was added to 50 µl of thawed competent cells and incubated on ice for 30 min. Samples were heat shocked in a 42 °C water bath for 50-60 sec before being immediately replaced on ice for 1 min. LB (lysogeny-broth) medium (10 g/l tryptone, 5 g/l yeast extract, 10 mM NaCl, pH 7) was added to a total volume of 1 ml and incubated at 37 °C for 1 hr in a shaking incubator. Following incubation, the sample was centrifuged at 6,000 rpm for 30 sec and the bacterial pellet resuspended in 100 µl of LB. This cell suspension was plated on solid LB-agar supplemented with ampicillin (100 µg/ml). The plates were inverted and incubated at 37 °C for 12-16 hrs. Colonies were isolated and mini-cultures prepared.

2.2.7 Plasmid DNA preparation

Plasmid DNA extraction was performed according to the E.Z.N.A. Plasmid Mini Kit II protocol, as per manufacturer's instructions (Omega bio-tek). Briefly, 10 ml of LB broth containing 100 µg/ml ampicillin was inoculated with a single colony and incubated in a shaking incubator overnight at 37 °C. The overnight culture of E.coli was pelleted by centrifugation at 3,000 rpm for 10 min and the supernatant discarded. The pellet was resuspended in 500 µl of Solution I (containing Tris-HCl/RNase A, pH 7.5), lysed with 500 µl Solution II (containing NaOH/SDS, pH 13.2) and neutralised with 700 µl Solution III (containing GdmCl/CH_3CO_2H, pH 13.2). After centrifugation at 13,000 rpm for 10 min, the supernatant was removed and passed through the silicon HiBind Mini Column. The plasmid DNA was precipitated with 500 µl of Buffer HB (containing GdmCl/C_3H_8O, pH 7) and washed twice with 700 µl DNA wash buffer (containing Tris-HCl/NaCl/70 % ETOH, pH 7.5). The DNA was eluted in 50 µl nuclease-free water, and the concentration recorded using a NanoDrop 1000 Spectrophotometer (see 2.2.9, Thermo Fisher). Potential clones were verified by sequencing (Dundee Sequencing Service) with the appropriate primer followed by computational alignment (see 2.2.1.1), and stored at – 20 °C until required.

2.2.8 Colony screening

Following isolation and purification of plasmid DNA, samples were screened to identify potential positive clones. RE digests were carried out using appropriate enzymes to distinguish newly synthesised clones. Reactions were set-up in a total volume of 10 µl and contained 1 µl of plasmid DNA, 0.5 U of restriction enzyme(s), enzyme buffer and nuclease-free water. Samples were incubated at 37 °C for 30 min and resolved on an agarose gel (see 2.2.4).

2.2.9 UV spectrophotometry

The purity and yield of plasmid DNA was checked using a NanoDrop 1000 Spectrophotometer (Thermo Fisher). DNA concentration was quantified by measuring Abs_{260} (since nucleotides absorb at this wavelength), and DNA purity checked using the Abs_{260}/Abs_{280} ratio, with ratios greater than 1.8 accepted as pure.

2.2.10 RNA transcription

Generally, constructs were linearised with HpaI or AscI (New England Biolabs) for 1 hr at 37 °C. Linearised DNA was purified using a Gel/PCR Purification kit (Biomiga). Each transcription reaction was carried out in a volume of 100 µl containing: 20 µl 5 X transcription buffer (Promega), 10.5 µl 100 mM DTT (Promega), 10 µl linearised DNA (~500 ng), 2.5 µl RNasin Ribonuclease Inhibitor (40 U/µl, Promega), 2.5 µl T7 RNA polymerase (20 U/µl, Promega), 40 µl 4 x 10 mM rNTP's (Promega) and 15 µl nuclease-free water. Reactions were incubated at 37 °C for 2 hr. Following this incubation, 5 µl of RQ DNaseI (1U/µl Promega) was added to remove template DNA, and reactions incubated for a further 20 min at 37 °C. Quality and yield was analysed using 0.8 % agarose gel and by spectrometry (NanoDrop 1000 Spectrophotometer, Thermo Fisher).

2.2.11 Mammalian cell culture

2.2.11.1 Cell maintenance

BHK-21 cells were grown in Dulbecco's modified Eagle medium (DMEM) containing 10 % (v/v) foetal bovine serum (FBS). Human rhabdomyosarcoma

(RD) cells were grown in DMEM with 10 % FBS and supplemented with penicillin (100 U/ml) and streptomycin (100 µg/ml). Cell monolayers were maintained in 75 cm^2 or 175 cm^2 tissue culture flasks (Greiner; Nunc), and routinely passaged with trypsin/EDTA at ~90 % confluency - determined by the growth rate of the cells. All cells were incubated at 37 °C in 5 % CO_2.

2.2.11.2 Cell stock freezing and resuscitation

Adherent cells (~90 % confluent) were incubated in 1 X trypsin/EDTA to detach cells from the plastic support and then resuspended in DMEM/10 % FBS, followed by centrifugation at 2,000 rpm for 5 min. Cell pellets were resuspended in DMEM supplemented with 20 % FBS and 10 % DMSO and aliquoted into labelled cryovials. Aliquots were firstly stored in an isopropanol chamber at -80 °C overnight before transferring to liquid nitrogen. Frozen cells were resuscitated by thawing at 37 °C before centrifuging at 2,000 rpm for 5 min. Pellets were resuspended in 1 ml DMEM/10 % FBS and left overnight at 37 °C/5 % CO_2. Media was replaced the following day to remove residual DMSO.

2.2.11.3 Cell transfection

The day before transfection, adherent cells were washed with PBS, incubated in 1X trypsin/EDTA, and detached cells diluted in DMEM/10 % FBS before seeding into 12-well, 6-well, or 60 mm plates at the appropriate cell seeding density. Replicon transcript RNA (1-4 µg) was transfected into cell monolayers the following day (80-90 % confluent) using Lipofectamine 2000 transfection reagent (Invitrogen). Briefly, RNA was diluted in 50-200 µl Opti-MEM (Gibco) and added to a Lipofectamine-Opti-MEM mixture consisting of; Lipofectamine (3-10 µl) and Opti-MEM (to a final volume of 50-200 µl). This RNA-Lipofectamin-Opti-MEM mixture was left for 5 min at room temperature, and then added drop-wise to cells. The plate was rocked gently back and forward to mix and incubated at 37 °C/5 % CO_2 for the time period required.

2.2.11.4 Cell treatment with Cycloheximide

To inhibit cell protein synthesis, cells were treated 24 hr post-transfection (unless otherwise stated) with cycloheximide (to a final concentration of 100 µg/ml). The

required volume of cycloheximide was added to 100 µl Opti-MEM and added drop-wise to cells to ensure uniform distribution. Plates were incubated at 37 °C/5 % CO_2 for the duration of the experiment.

2.2.12 IncuCyte microscopy

Images of transfected cells were captured at intervals between 0-24 hr post-transfection (unless otherwise stated) using an IncuCyte ZOOM kinetic imaging system (Essen BioScience) housed within an incubator maintained at 37 °C/5 % CO_2. Images were captured from 9 regions/well in a 12-well plate using the 10x objective. GFP positive cell counts and GFP intensities were measured using the IncuCyte image processing software. Values from all 9 regions of each well were pooled and averaged across 4 replicates.

2.2.13 SDS polyacrylamide (SDS-PAGE) analysis of proteins

Samples were prepared by placing the plates on ice, removing the media, washing with 1 ml ice-cold PBS and scraping cells from the plate using a bent pipette tip. This cell suspension was transferred to a microfuge tube and centrifuged at 2,000 rpm, at 4 °C, for 5 min. The supernatant was removed and the pellet resuspended in RIPA lysis buffer (100 µl of 50 mM Tris-HCl, 150 mM NaCl, 1 % (v/v) Nonidet P-40, 0.5 % Deoxycholic acid, 0.1 % (w/v) SDS, pH 7.4). Samples were left on ice for 25-30 min with occasional vortexing. This was followed by centrifugation at 13,000 rpm for 15 min, at 4 °C and removal of the supernatant to a new microfuge tube. To check sample yield, small aliquots (10 µl) were run on 12 % gels and stained with Coomassie blue (2.5 g/l Coomassie Brilliant Blue R250, 50 % (v/v) methanol, 10 % (v/v) acetic acid) for 1 hr. Gels were de-stained (40 % (v/v) methanol, 10 % (v/v) acetic acid) overnight and suitable protein yield in each sample determined visually. Samples were mixed with 5 X SDS- sample buffer (25 % (v/v) β-mercaptoethanol, 10 % (w/v) SDS, 50 % (v/v) glycerol, 0.5 % (w/v) bromophenol blue) and boiled at 95 °C for 5 min, prior to gel-loading. Sodium dodecyl sulphate-polyacrylamide gels, with 5 % stacking gels, were cast using ingredients and methods described by Sambrook *et al.*, (1989). Precast 4-20 % polyacrylamide gradient gels (Expedeon) were used during certain experiments to separate samples containing many proteins of interest, of varying size. Gels were run using a KuroGel Verti Electrophoresis unit

(VWR) in 1 X SDS-PAGE Tris-Glycine (1.92 M Glycine, 0.25 M Tris, 0.1 % (w/v) SDS), or 1 X Tris-Tricine (Expedeon) running buffer, at 100-140 V for 1.5-2.5 hr.

2.2.14 Western blotting

Proteins were resolved by SDS-PAGE (see 2.2.13) and transferred to a nitrocellulose membrane (Invitrogen) at 20 V for 10 min in an iBlot transfer system (Invitrogen); as described by the manufacturer. Membranes were blocked for 1 hr at room temperature in 5 % PBS-T (PBS, 0.1 % Tween 20, 5 % non-fat milk). The membrane was probed with the appropriate primary antibody (1:1000-2000 dilution in 5 % PBS-T) overnight at 4°C. The following day the membrane was washed three times for 5 min in PBS-T (PBS, 0.1 % Tween 20), then probed with HRP- conjugated anti-rabbit or anti-mouse (Invitrogen, 1:2000) secondary antibodies in 5 % PBS-T for 1 hr at room temperature. Finally, the membrane was washed a further three times for 5 min in PBS-T, and antibody binding detected using the EZ-ECL HRP chemiluminescence detection kit (Biological Industries).

2.2.15 *In vitro* transcription and translation

Protein synthesis was measured from replicon plasmids or viral transcript RNA by using a coupled or uncoupled system, respectively. [^{35}S] methionine-labelled proteins were synthesised from plasmid DNA (100 ng) using a TNT® Quick Coupled Transcription/Translation System (Promega) following manufacturer's instructions. Reactions were incubated at 30 °C for 90 min and analysed by SDS-PAGE (see 2.2.13).

Transcript RNA was used to programme nuclease-treated Rabbit Reticulocyte Lysates (Promega) supplemented with HeLa cell S10 cytoplasmic extracts (DCP). Reactions were set-up as follows; rabbit reticulocyte lysate (7 µl), viral RNA (0.25 µg-2 µg), amino acid mix (minus methionine; 0.5 µl 1 mM), [^{35}S] methionine (1,200 180 Ci/mmol; 0.5 µl), 10 U RNasin and 2.25 µl HeLa cell extract in a total volume of 12.5 µl. Samples were incubated at 30 °C for 3 hr and analysed by SDS-PAGE (4-20 % Tris-Glycine, Expedeon, see 2.2.13). Gels were stained with coomassie blue for 1 hr and de-stained overnight before fixing (10 % (v/v) acetic acid, 25 % (v/v) isopropanol) for 1 hr. Gels were then dried onto filter

paper for 15-20 min and exposed to film (Thermo Scientific) for 1-4 days at -70 °C.

2.2.16 Preparation of virus stocks

Stocks of the CDLR and cu|cu viruses were provided by Prof. P. Simmonds (Roslin Institute, University of Edinburgh). All other virus stocks were prepared by transfection of transcript RNA into confluent monolayers of RD cells (1 x 25 cm^2 flask). Plasmids were linearised with NotI and transcript RNA synthesised using T7 RNA polymerase (see 2.2.10). RNA (100 ng) was transfected into cells using Lipofectamine 2000 (Invitrogen, see 2.2.8.3). Cells were incubated at 37 °C/5 % CO_2 and monitored for the presence of CPE. This cell lysate was used to generate passage 1 stocks by re-infecting RD cells (1 x 75 cm^2 flask). Once CPE had reached ~70 % (~24 hr), the supernatant was harvested and centrifuged at 2,000 rpm for 5 min to remove cell debris. Aliquots were made and frozen at -80 °C. Titres were determined by $TCID_{50}$ titration (see 2.2.17).

2.2.17 Virus titrations

RD cells were seeded to ~60 % density in 96-well plates. Virus preparations were serially diluted 10-fold in the cell medium (200 µl final volume, DMEM/10 % FBS) of the 96-well plate. Plates were incubated at 37 °C/5 % CO_2 for ~4 days until CPE had formed. The media was removed and monolayers fixed for 1 hr in 10 % (v/v) formal saline. Cells were then stained with crystal violet (0.1 % (w/v) crystal violet, 10 % (v/v) ethanol) for 1 hr to visualise CPE. This was followed by washing with H_2O and scoring of plates for the presence or absence of CPE. The 50 % end-point dilution ($TCID_{50}$) was calculated using the method described by Reed and Muench (1938). The assay was performed in quadruplicate for each virus.

2.2.18 Virus infections

Cell monolayers were washed with PBS prior to inoculation to remove traces of serum. Monolayers were infected with virus diluted in serum-free DMEM (or DMEM only for mock transfections) at the required multiplicity of infection (MOI) for 1 hr at 37 °C. The inoculum was removed and the cells washed before applying fresh media. Cells were incubated at 37 °C/5 % CO_2 until harvested.

2.2.19 Replication phenotype

To determine the viral titre over multiple-step growth cycles, RD cell monolayers in triplicate in 24-well plates were infected at an MOI of 0.01 for 1 hr as described above (see 2.2.18). Samples were withdrawn at given time points (6, 24 and 42 hr post-infection) and the viral titre determined by $TCID_{50}$ assay (see 2.2.17).

2.2.20 Competition assays

Equal titres of virus pairs (combined MOI of 0.01) were applied to RD cells in 25 cm^2 flasks. Following the development of CPE, supernatants were removed and 300 µl applied to fresh cell monolayers. This procedure was continued for up to 10 passages. Viral RNA was extracted from the supernatant (described below) and region 1 amplified by combined reverse transcription- PCR (see 2.2.2.2). The results were determined by RE digestion of the region amplified.

2.2.21 Extraction of viral RNA from cell supernatant

RNA was isolated from viral supernatant collected during competition and multiple-step growth assays using an RNeasy Mini Kit (Qiagen). Briefly, 100 µl of supernatant was mixed thoroughly with RLT buffer (350 µl). This solution was mixed with ethanol (250 µl; 100 %) and transferred to an RNeasy mini spin column before centrifuging at 12,000 rpm for 15 sec. The column was washed with RPE buffer (500 µl) and centrifuged at 12,000 rpm for 15 sec. This step was repeated, but with a longer centrifugation step of 2 min. The column was dried by centrifugation at full speed for 1 min. The RNA was eluted by the addition of nuclease-free H_2O (50 µl) to the column membrane, followed by a final centrifugation at 12,000 rpm for 1 min. Samples were stored at -80 °C until required.

Chapter 3: FMDV Replicons Encoding Green Fluorescent Protein are Replication Competent.

3.1 Introduction

The original foot-and-mouth disease virus (FMDV) replicon was based on the genome of FMDV type-O O1/Kaufbeuren/FRG/66 (Forss *et al.*, 1984), and was created by deletion of sequences encoding the L proteinase (L^{pro}) and the majority of the capsid coding region. The reporter gene CAT was inserted such that FMDV RNA replication could be monitored *via* CAT assays of cell extracts (pT7rep, McInerney *et al*., 2000, Figure 3.1). The use of such replicons allows replication of highly pathogenic viruses without the need for high-containment facilities, allows the characterisation of viral functions and to screen for attenuated forms for vaccine development.

The aim of this section was to create a new FMDV replicon by replacing CAT from pT7rep with a GFP-PAC fusion protein: the pGFP-PAC replicon (Figure 3.1). This replicon allows FMDV replication to be monitored *via* GFP fluorescence in a non-invasive manner, in real time, outwith high-containment facilities. The presence of Puromycin-N acetyl-transferase (PAC) also permits persistent infection to be investigated through selection of FMDV-replicon transfected cells with puromycin.

3.2 Construct design

The existing FMDV replicon system (pT7rep, Figure 3.1) was modified by the replacement of sequences encoding chloramphenicol acetyl-transferase (CAT) with those encoding a functional L proteinase (L^{pro}) linked to a bi-functional fluorescent/antibiotic resistance fusion protein (enhanced green fluorescent protein/puromycin resistance; pGFP-PAC; Figure 3.1). Sequences encoding the N-terminal 18 aa of the capsid protein 1A were introduced between L^{pro} and GFP-PAC to ensure correct proteolytic processing by L^{pro} at its C-terminus. Similarly, 40 aa C-terminal residues of the capsid protein 1D were introduced upstream of 2A to ensure accurate processing (‘ribosome skipping’) at the 2A/2B site.

A function of the proposed system was to distinguish between replicating and replication-incompetent forms; therefore an attenuated form of pGFP-PAC was generated. As discussed in the Introduction (1.1.2.6), viruses lacking L^{pro} were attenuated in cell culture and in swine (Chinsangaram *et al*., 1998; Piccone *et al*., 1995). Therefore a 'leaderless' replicon completely lacking the L gene, pLL-GFP-PAC, was constructed. The GFP-PAC region was amplified *via* PCR (primers listed in Table A.1) and ligated into the pGFP-PAC replicon which was digested with PsiI and XmaI to remove L (Figure 3.1, A.2). Transcript RNA transfected into cells can act as mRNA, therefore it was important to create a replicon to differentiate fluorescence generated by the input RNA from those genomic forms which are replication competent. A replication-incompetent form was created through a large deletion in the viral polymerase, $3D^{pol}$, by restriction digestion with MluI: pGFP-PAC-Δ3D (Figure 3.1, A.2).

3.3 Restoration of L proteinase activity

The major host cell target of L^{pro} is the translation initiation factor eIF4G (Devaney *et al.*, 1988). To confirm L^{pro} activity had been restored in the pGFP-PAC replicon, western blots were carried out to detect eIF4G-cleavage by L.BHK-21 cell monolayers were transfected with 3 µg of replicon RNA and harvested at the time points indicated. Proteins were separated by SDS-PAGE and probed using anti-eIF4G and anti-β-tubulin antibodies. Figure 3.2A shows eIF4G was cleaved as early as 1 hr post-transfection (p.t.) in pGFP-PAC-transfected BHK-21 cells. From 6 hr p.t. eIF4G began to re-accumulate and the cleavage product disappeared thereafter. It should be noted that transfection of cell monolayers leads to ~40-50 % cell-transfection efficiency; therefore these data are due to adherent, non-transfected cells. Cells transfected with leaderless, pLL-GFP-PAC, transcript RNA did not show the same kinetics of eIF4G degradation (Figure 3.2B), however some degradation occurred which is consistent with evidence that $3C^{pro}$ can also cleave this initiation factor (Belsham *et al.*, 2000).

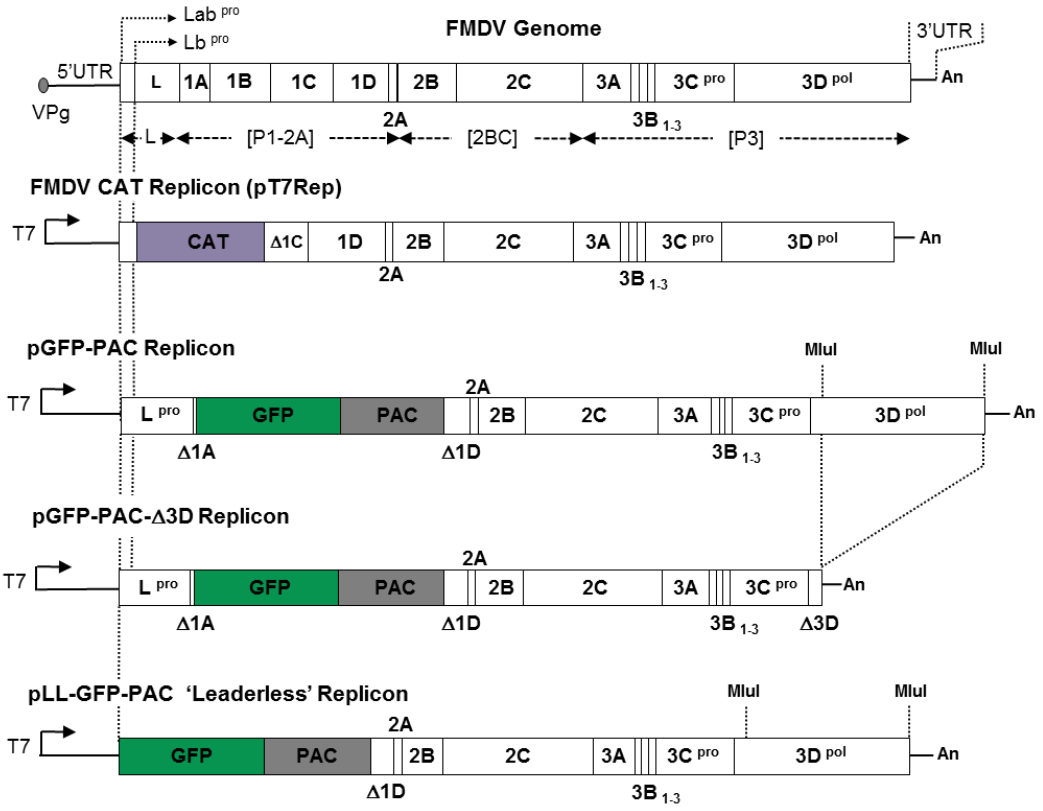


Figure 3.1. Replicon Constructs. The structure of the FMDV genome is shown together with replicon plasmid constructs. Polyprotein domains are shown as boxed areas, together with the 'primary' processing products L^{pro} (Lab^{pro} and Lb^{pro} forms), [P1-2A], [2BC] and P3 ([$3AB_{1-3}CD$]). The original CAT replicon (pT7Rep;Ellard *et al.*, 1999; McInerney *et al.*, 2000) was modified to re-insert the L proteinase sequences and the CAT reporter replaced with a GFP-PAC fusion protein (pGFP-PAC, assembled by Dr G.Luke). This plasmid was modified to create a replication incompetent form by deletion of the 3D polymerase (pGFP-PAC-Δ3D). A replication attenuated form was created by deletion of the L proteinase (pLL-GFP-PAC; similar to that described by Piccone *et al.*, 1995).

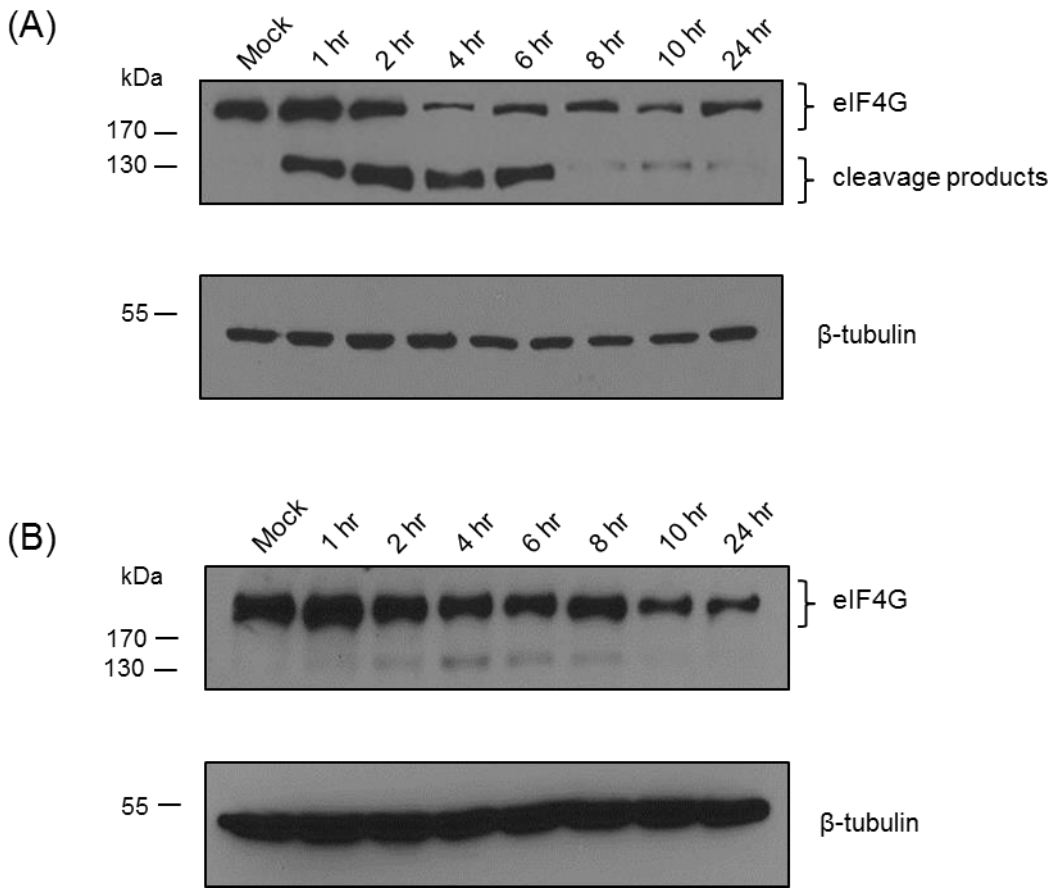


Figure 3.2. Cleavage of eIF4G in cells transfected with replicon RNA. BHK cell extracts from pGFP-PAC (A) or pLL-GFP-PAC (B) replicon-transfected cells (1-24 hr as indicated) were separated by SDS-PAGE, transferred to nitrocellulose membranes, and analysed by western blotting with anti-eIF4G or β-tubulin as a control. Uncleaved eIF4G and its cleavage products are shown on the right with molecular weight markers on the left.

3.4 Quantitation of FMDV replicon-derived GFP fluorescence

RNA replication was quantified indirectly by quantifying GFP fluorescence. Plasmids were linearised with HpaI and transcribed *in vitro* using T7 RNA polymerase (Promega). BHK-21 cells were seeded into 12-well plates 24 hr prior to transfection. Monolayers were transfected with 1 µg of replicon RNA and fluorescence monitored using an IncuCyte ZOOM fluorescent microscope. Images were taken every 2 hrs for a total of 24 hrs, with nine images captured per well of a 12-well plate. GFP fluorescence could be detected in replicon-transfected cells ~2 hr p.t. (Figure 3.3, 3.4), with the peak signal achieved ~6 hr p.t. (Figure 3.4). This is directly comparable to vRNA replication detected *via* RT-qPCR (Chang *et al.*, 2009; Gu *et al.*, 2007). The GFP signal started to decline ~8-10 hr p.t. (Figure 3.4), with cells displaying a rounded-up morphology compared to the characteristic elongated, fibroblastic form of BHK-21 cells (Figure 3.3); typical of the cytopathic effect (CPE) observed during viral infection. The life cycle of FMDV is a rapid process with replication lasting around 8-10 hrs. At this point in the cycle cells lyse releasing particles, however, this process does not occur when using the current replicon system. From 8-10 hr onwards cells are effectively dead, hence any measurements recorded after this time are that of dead cells and this is demonstrated by the rapid decline in GFP fluorescence after this time point (Figure 3.4).

Since replicon RNA can act as mRNA, it was expected that pGFP-PAC-Δ3D would produce fluorescence. Although it cannot replicate due to deletions in the viral polymerase, L^{pro} is still active and able to shut-off host cell translation, which in turn increases IRES-driven translation. As predicted, cells transfected with RNA transcripts produced from this construct generated GFP fluorescence, albeit at a lower level (~2-fold) in comparison to the replication-competent pGFP-PAC (Figure 3.4).

The pLL-GFP-PAC replicon showed a lag in the detection of fluorescence, with the GFP signal arising ~4 hr p.t. (Figure 3.3, 3.4). Unexpectedly, the total fluorescent signal for cells transfected with this construct was much higher than the pGFP-PAC 'WT' replicon (~2-fold, Figure 3.4B). It was hypothesised that deletion of L^{pro} would lead to a reduction in fluorescent signal as the replicon RNA would have to compete with cellular mRNA for the host-cell translation machinery, and therefore viral protein synthesis would decrease. Due to the results obtained with the pLL-GFP-PAC replicon, cell extracts were prepared and

probed with anti-GFP antibodies. Western blotting of these extracts showed that in pGFP-PAC replicon transfected cells the [Δ1A-GFP-PAC-Δ1D-2A] processing product was cleaved, but not in cells transfected with the leaderless form (Figure 3.5A). Examination of the L^{pro} /1A, eIF4G cleavage sites (Kirchweger *et al.*, 1994; Strebel & Beck, 1986) and the GFP-PAC sequence (Figure 3.5B) suggested a potential L^{pro} cleavage site within GFP at Lys^{126} –Gly^{127} , which maps to the C-terminus of β-sheet 6 (Ormö *et al.*, 1996; Yang *et al.*, 1996), and would produce a product corresponding to the molecular mass of the major cleavage product (~42 kDa, Figure 3.6A). The cleavage of [Δ1A-GFP-PAC-Δ1D-2A] was much slower than eIF4G, with the cleavage product appearing ~4 hr p.t. (Figure 3.5A), in contrast to eIF4G which was mostly cleaved by then (Figure 3.2A). The plasmid pJC3-PAC (unpublished; encodes GFP-PAC, FMDV 2A and mCherry FP) and was used to show the size of the fusion product [GFP-PAC-2A]. As $3C^{pro}$ can also cleave eIF4G, potential $3C^{pro}$ sites were examined and none were detected within GFP or PAC.

Due to both the western blotting and IncuCyte data obtained from analysis of pLL-GFP-PAC replicon-transfected cells, the average GFP intensity was calculated at various time points over the course of the experiment using the IncuCyte. Analysis of transfected cells at 2, 6 and 12 hr p.t. showed that the WT pGFP-PAC replicon increased up to ~1.4-fold from 2-6 hr (Figure 3.6), steadily decreasing from 12 hr onwards (data not shown). In conjunction with the data described above, the average intensity of pLL-GFP-PAC transfected cells was lower at 2 hr p.t., but increased ~7-fold from 2-6 hr. As expected, there was no change in pGFP-PAC-Δ3D average GFP fluorescence. This data demonstrates two points. Firstly, cell death induced by both replication and through cleavage of eIF4G impacts on measurements obtained from 12 hr onwards. Secondly, cleavage of GFP by L^{pro} affects the GFP intensity derived from the 'WT' pGFP-PAC replicon placing constraints on experiments involving analysis of GFP fluorescence: the true signal intensity is not being determined. Both of these points should be addressed prior to commencing studies into aspects of the viral life cycle.

3.5 Conclusions

- FMDV replicons encoding a GFP-PAC fusion protein were replication competent and could be used to measure RNA replication indirectly *via* GFP fluorescence.

- A fully functional L^{pro} was reinstated into the existing replicon system and could cleave its major host cell target, eIF4G.

- The fluorescent signal generated from replicating/non-replicating forms could be distinguished, despite some cleavage of GFP by L^{pro}. This system will facilitate studies to screen for attenuated genomes which can be rescued by colleagues at the Pirbright Institute as potential live-attenuated vaccine candidates.

- Measurement of replication indirectly *via* GFP fluorescence can allow for characterisation of the viral life cycle through screening studies which can incorporate mutagenesis, deletions, insertions *etc*., with rescue of infectious copies verifying experimental observations.

- This screen provides a rapid, facile method of quantifying FMDV replication in comparison to nucleic-acid based methods, and removes the need for high-containment facilities to measure FMDV RNA replication.

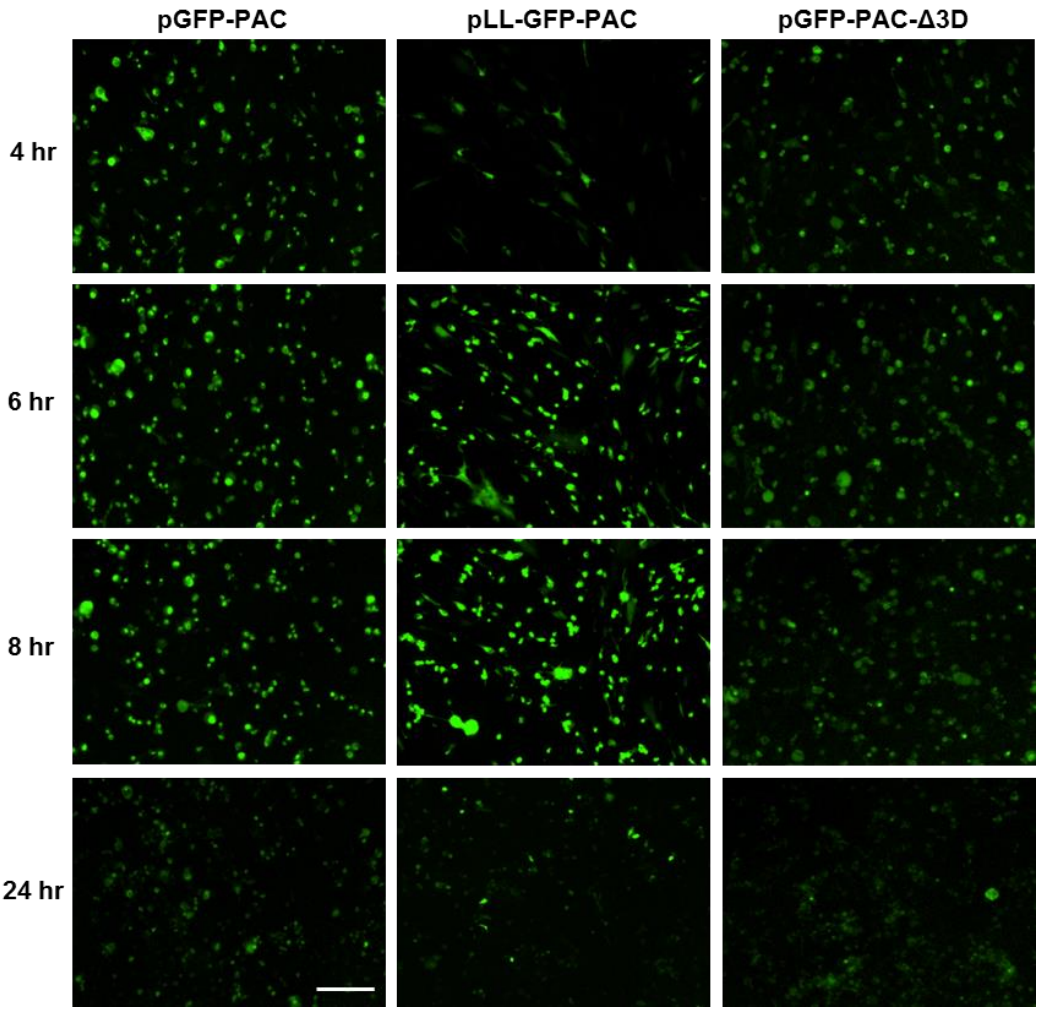


Figure 3.3. GFP expression in FMDV replicon-transfected BHK-21 cells. Transcript RNA from the pGFP-PAC, pLL-GFP-PAC and pGFP-PAC-Δ3D replicons was introduced into cell monolayers, and fluorescent images captured at 2 hr intervals over a 24 hr period using the IncuCyte ZOOM imaging system. A representative of the nine images captured is shown at 4, 6, 8 and 24 hr post-transfection. Data was obtained from 3 independent transfections, with 4 replicates per transfection. Scale bar represents 100 μm.

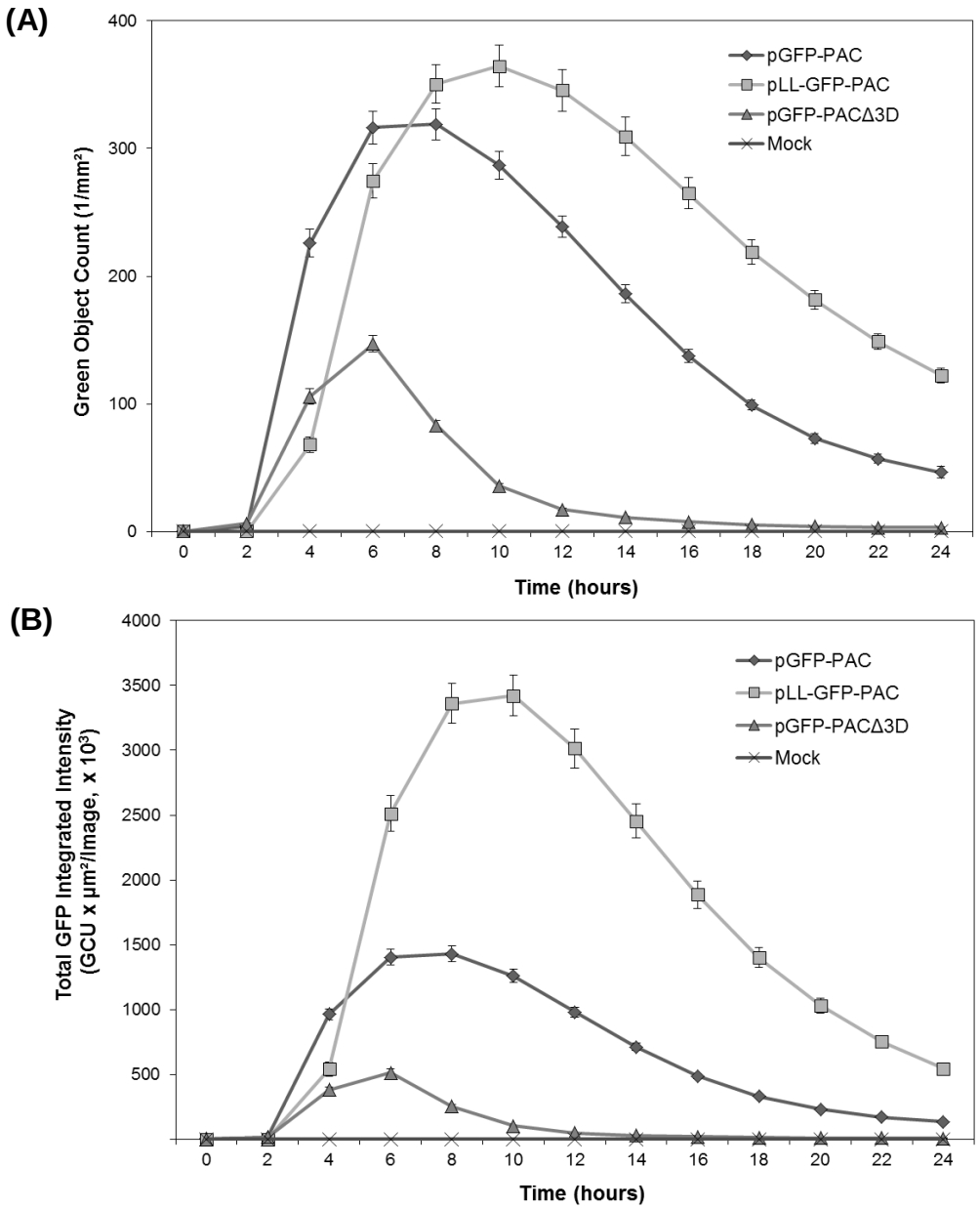


Figure 3.4. Time course of FMDV replicon-derived GFP fluorescence. Data from mock-transfected BHK-21 cells are shown, together with cells transfected with transcript RNA derived from the pGFP-PAC replicon, 'leaderless' replicon pLL-GFP-PAC and the polymerase deletion pGFP-PAC-Δ3D. At the time points indicated images were captured and the GFP fluorescence quantified for each replicon construct over a 24 hr period: data shown as the green object count/mm^2 (A) or as the total integrated GFP fluorescence intensity (B). Data points/error bars shown are derived from 3 independent transfections, with 4 replicates for each transfection.

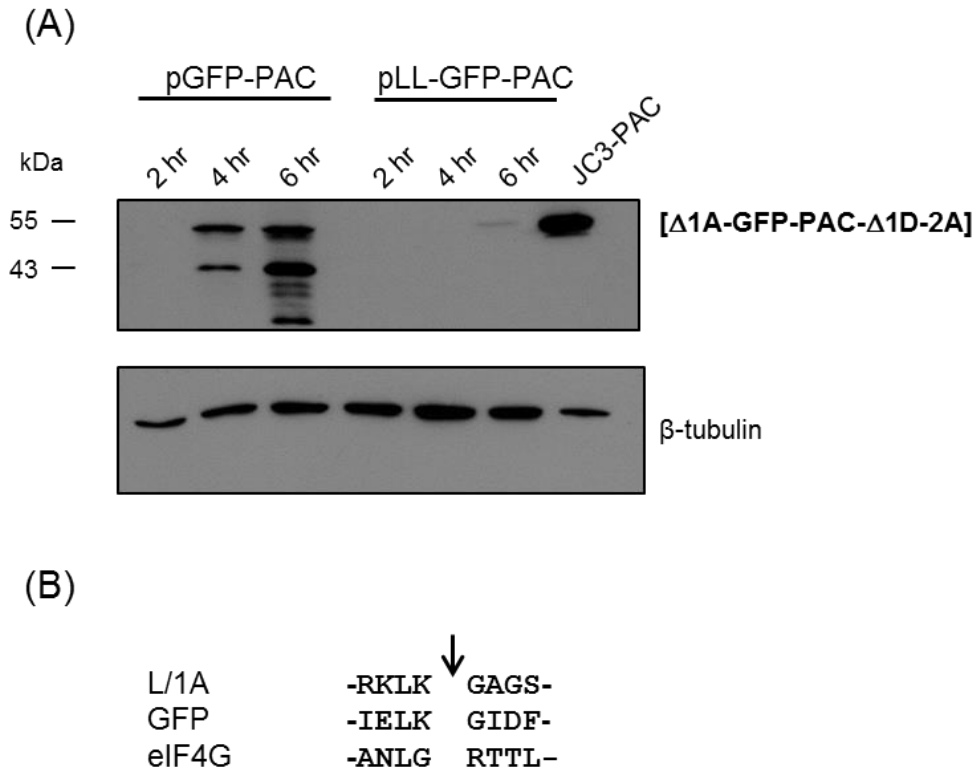


Figure 3.5. Cleavage of GFP in cells transfected with replicon RNA. Extracts were prepared from pGFP-PAC or pLL-GFP-PAC (A) replicon-transfected BHK-21 cells at the time points indicated. Extracts were separated by 12 % SDS-PAGE, transferred to nitrocellulose membranes, and analysed by western blotting with anti-GFP and anti-β-tubulin antibodies. The sequence flanking the L/1A, eIF4G and predicted GFP-L proteinase cleavage sites are shown (B).

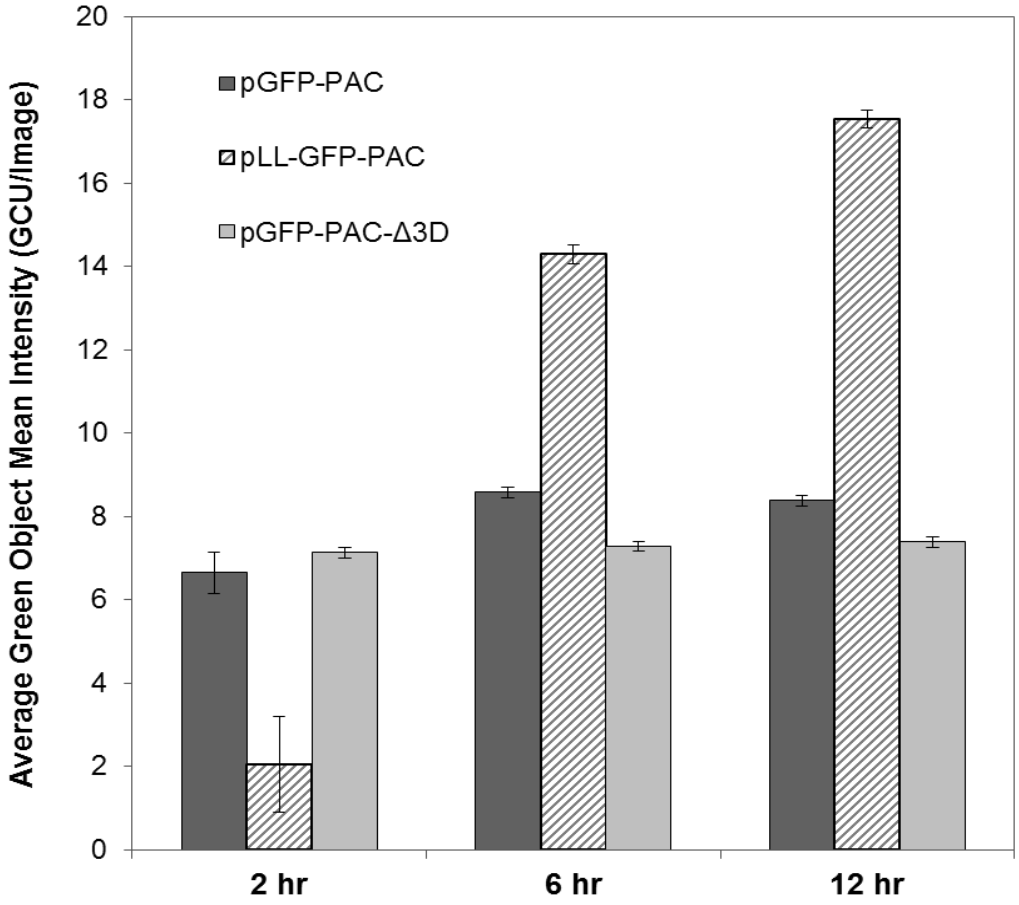


Figure 3.5. Average mean intensity of GFP-positive replicon-transfected BHK cells. At the time points indicated images (9/well) were captured and the average mean intensity quantified for each replicon construct. Data shows the average mean intensity at 2, 6 and 12 hr p.t. and is derived from 3 independent transfections, with 4 replicates for each transfection.

3.6 Discussion

3.6.1 Detection of FMDV replication *via* GFP fluorescence

Replicon RNA can act as mRNA and, therefore, will be translated following entry into the cell. Fluorescence generated by input RNA must be differentiated from fluorescence generated by active replication. The replication-incompetent pGFP-PAC-Δ3D replicon allows for this to be distinguished: replicon RNA cannot be synthesised by the viral polymerase as it contains a large deletion- thus, signal intensities are derived from initial input RNA only. Detection of GFP fluorescence from replication-competent replicon RNA, however, includes (i) translation of the input RNA, (ii) the synthesis of –ve strand and (iii) subsequent +ve strand synthesis, followed by further protein translation. The GFP fluorophore also requires time for post-translational modifications and maturation of the protein structure. Therefore, one might expect some additional lag in the detection of replicon RNA replication *via* GFP fluorescence in comparison to nucleic acid-based detection methods. GFP fluorescence was, however, detected ~2 hr p.t. which is directly comparable to strand-specific RT-qPCR (Gu *et al.*, 2007). The pLL-GFP-PAC replicon showed a lag in replication of ~2 hr that continued until 6-8 hr p.t. The number of GFP positive cells then increased up to ~1.5- fold above pGFP-PAC for the remainder of the experiment. This trend could be attributed to the stronger signal attained by the pLL-GFP-PAC replicon, resulting in a larger proportion of transfected cells being detected by the IncuCyte microscope in comparison to pGFP-PAC. The observation that $3C^{pro}$ can also cleave eIF4G (Belsham *et al.*, 2000) explains the kinetics observed in pLL-GFP-PAC transfected cells from 4 hr onwards. Once cells are unable to undergo cap-dependent translation, the machinery can be used solely by FMDV replicon RNA, therefore increasing the rate of translation and detection of GFP.

As both eIF4G and replication induce cell death, measurements taken following a single growth cycle i.e. 8- 10 hr are essentially of dead cells. Induction of cell death can affect analysis due to cells rounding-up and detaching from the plastic support. This does not pose problems when measuring differences between replicating and non-replicating forms to screen for attenuated genomes; however, it may cause complications when studying aspects of the viral life-cycle. It could

be more beneficial to plot data corresponding to the first 6-10 hr only, allowing for the kinetics during a single growth curve to be captured.

Interestingly, although the pLL-GFP-PAC replicon demonstrated a lag in replication, active replication was detected; an observation not reported by Piccone *et al.*, (1995). In this study, when BHK-21 cells were transfected with transcript RNA lacking L^{pro}, virus could not be rescued. Viruses containing the 84 nt "spacer" region (described in Introduction 1.1.2.6) between the two L^{pro} initiation codons grew, however, to slightly lower titres and produced a reduced plaque phenotype when compared to WT (Piccone *et al.*, 1995). This virus was also attenuated when tested in cattle and swine (Brown *et al.*, 1996; Chinsangaram *et al.*, 1998). A replicon containing the 84 nt spacer region, pLa-GFP-PAC, displays kinetics that are comparable to pGFP-PAC at 2 hr p.t. – no lag in GFP detection was observed, however, like the pLL-GFP-PAC replicon the signal intensities were higher (data not shown). Furthermore, insertion of L^{pro} or Lb^{pro} into plasmid pJC3-PAC (pL^{pro}-PAC or pLb^{pro}-PAC), further implies that GFP is cleaved by this protease, with GFP fluorescence substantially reduced when compared to plasmids containing either the spacer region of L^{pro} or Δ1A (data not shown). mCherry expression generated by the pL^{pro}-PAC plasmid was 10-fold higher than Aq.GFP expression, implying that Aq.GFP was being cleaved by the proteinase.

Insertion of the spacer region into pJC3-PAC, such that the GFP now bears the N-terminal 28 amino acids of L^{pro}, showed ΔL-GFP localised to the perinuclear region of transfected BHK-21 cells (data not shown). A different localisation pattern was observed with both the pL^{pro}-PAC or pLb^{pro}-PAC plasmids, with GFP remaining in the cytoplasm (data not shown). The presence and utilisation of the two conserved AUG's in L^{pro} is still largely unexplained, as is the conserved RNA secondary structure found in this region (Witwer *et al.*, 2001). These findings suggest the spacer region may serve as a localisation signal, ensuring the Lab form is delivered to specific areas of the cell to carry out activities essential to the viral life cycle. This is supported by sequence analysis performed by Carrillo *et al.*, (2005) on isolates of FMDV, and found the methionines within this region to be invariant, suggesting that both isoforms of L^{pro} are important for aspects of FMDV biology. These results pose very interesting and future experiments are planned to investigate these findings.

3.6.2 Cleavage of GFP by L^{pro}

Unexpectedly, following fluorescence and western blot analyses it was apparent that the *Aequorea* GFP reporter protein showed some cleavage by L^{pro}, leading to a ~2-fold reduction in signal derived from pGFP-PAC. Despite this unforeseen anomaly, attenuation could still be measured, and the strong signal obtained from the leaderless form demonstrated the potential for this system. Experiments undertaken to either mutate the possible L^{pro} cleavage site within GFP, or to replace GFP with another fluorescent protein that is not cleaved by L^{pro} are described in Chapter 5.

The next two chapters describe further development of this system to increase replication efficiency to mimic a viral infection as closely as possible, within the constraints of the reporter system. Decreasing time and costs through modification of the replicon is also investigated, as well as mutagenesis of the potential L^{pro}-GFP cleavage site and the use of other fluorescent reporter proteins that are not cleaved by L^{pro} (Chapters 4-5).

Chapter 4: An Improved FMDV Replicon System Encoding Self-Cleaving Ribozymes.

4.1 Introduction

Synthetic transcripts produced from viral cDNA vectors can contain extra non-viral nucleotides introduced during cloning, from vector sequences, through linearization of plasmids creating overhangs and from additions made by bacterial polymerases (Clark, 1988; Milligan *et al.*, 1987). During PV RNA replication, the viral polypeptide $3B^{VPg}$ is uridylated onto the 5'-end of both negative and positive sense transcripts and is thought to be an important step during initiation of RNA synthesis (Paul *et al.*, 1998). It has previously been reported that PV RNA transcripts with authentic 5'-ends replicated more effectively in cell culture when compared to transcripts containing non-viral sequences (Herold & Andino, 2000). Using a hammerhead ribozyme to produce exact 5'-ends, Herold and Andino (2000) demonstrated that transcripts replicated with better kinetics and there was no lag before replication.

A precise 3'-end doesn't seem to be as imperative for picornavirus replication; however the presence of extra nucleotides has been shown to reduce infectivity of viral RNA transcripts. Sarnow (1989) showed that long extensions of 17 or more nucleotides decreased infectivity of PV transcripts, with extensions of 4 nucleotides showing a 5-fold increase in comparison to virion RNA. Ghanem *et al*., (2012) demonstrated a 10-fold increase in rabies virus rescue when using an improved version of the HDV ribozyme to generate an exact 3'-end.

The synthesis of RNA transcripts by the use of bacterial polymerases is a costly and time-consuming process. The use of the major late T7 Φ terminator (T7TΦ) sequence (Carter *et al.*, 1981) at the 3'-end of viral sequences can remove the need for linearisation of cDNA plasmids in preparation for transcription, or reduce the addition of non-viral sequences onto the ends of transcripts that have been linearised.

Transfection of cDNA plasmids encoding infectious PV RNA can produce viral particles, albeit to a low yield (Racaniello & Baltimore, 1981; Semler *et al.*, 1984), and also eliminates the requirement for synthesising transcript RNA.

Thus, this section describes improvement of the existing pGFP-PAC FMDV replicon (Tulloch *et al.*, 2014a), by introducing self-cleaving ribozymes into the 5'- and 3' UTRs. The 5'-end of pGFP-PAC contains three extra guanine residues which are introduced by T7 RNA polymerase following transcription initiation at the T7 promoter (Ikeda & Richardson, 1986; Imburgio *et al.*, 2000; Martin *et al.*, 1988, A.2). The 3'-end contains consecutive cytosine residues (C_{16}) following the poly(A) tail (A.2) that were introduced *via* cloning (Zibert *et al.*, 1990). Linearization with HpaI prior to RNA transcription also results in the addition of non-viral sequences (GTT, A.2), leading to transcripts containing nineteen non-template nucleotides added onto the poly(A) sequence.

Further modifications include incorporation of the human cytomegalovirus (CMV) promoter (Boshart *et al.*, 1985), with the intention to transfect replicon DNA into cells; driving transcription of replicon RNA *via* the promoter, diminishing the need for RNA transcription kits (A.3). A T7TΦ sequence was also inserted at the 3'-end of the viral cDNA to either eliminate the necessity for linearisation prior to transcription, or to aid in the termination of T7 RNA polymerase following linearization (A.3). In combination, these modifications could reduce the time taken to prepare transcript RNA, lower costs by limiting the need for expensive reagents, and increase the efficiency of the existing system by improving replicon kinetics to closely represent a viral infection.

4.2 Construct design

Insertion of the CMV promoter was achieved by cloning replicon cDNA from pGFP-PAC into the commercially available pcDNA 3.1 (+) (Invitrogen). This clone was modified further to include cDNA copies of each ribozyme which were provided as gene blocks (Dundee Cell Products). The *cis*-active Hammerhead (HH) ribozyme (Herold & Andino, 2000; Uhlenbeck, 1987) was cloned between the T7 promoter and 5' terminal sequences of the UTR, with the *cis*-active Hepatitis delta ribozyme (HDV), followed by a T7TΦ sequence, inserted between the poly(A) tail and the vector backbone, creating the pRbz replicon (Figure 4.1). To examine the activity of each ribozyme independently from one another and to ensure both were fully functional, constructs were generated that contained either the HH ribozyme (pHH_Rbz) or the HDV ribozyme (pHDV_Rbz) (Figure 4.1). This was achieved through RE digestion of the pGFP-PAC and pRbz replicons with SpeI and EcoRI, followed by ligation of the corresponding insert into the

appropriate vector (A.2-A.3 for sequence maps and RE sites). Finally, replicons containing a modification to the HDV ribozyme were made in order to increase the processing activity, as a previous report demonstrated that the cleavage activity of the HDV "core" sequence was poor in comparison and viral rescue and mini-genome expression was increased when constructs contained this altered HDV sequence (Ghanem *et al.*, 2012). A sequence consisting of 5 nucleotides (AGCCA; Ghanem *et al.*, 2012; Perrotta & Been, 1998) was added to the 3'-end of the HDV ribozyme in both the pRbz and pHDV_Rbz replicons, generating pRbz_5nts and pHDV_Rbz_5nts, respectively (Figure 4.1).

4.3 The effect of ribozymes at the 5' and 3'-ends on replicon-derived replication

To examine the effect of exact ends on replication, BHK-21 cells were transfected with 1 µg of transcript RNA and replication measured indirectly *via* GFP fluorescence using an IncuCyte ZOOM fluorescent microscope, as previously described (Tulloch *et al.*, 2014a). GFP fluorescence increased almost immediately for the pRbz_5nts and pHH_Rbz replicons, whereas in comparison the pGFP-PAC replicon showed a delay of ~2 hr (Figure 4.2, 4.3B). At 2 hr p.t. there was a 20-fold increase in GFP fluorescence from constructs containing the 5' HH ribozyme compared to all others tested (Figure 4.3B). This increase in fluorescence continued throughout the course of the experiment resulting in an increase in overall expression of up to 1.5-fold (Figure 4.3B). Constructs containing the HDV ribozyme alone exhibited replication kinetics that were similar to pGFP-PAC, with the pHDV_Rbz_5nts replicon exhibiting ~1.5-fold less fluorescence than the pHDV_Rbz replicon (Figure 4.3B) This was unexpected since previous reports demonstrated that the 5 nt extension improved processing of the ribozyme (Ghanem *et al.*, 2012; Perrotta & Been, 1998). Nevertheless, the combination of both ribozymes increased the initial GFP signal at 2 hr p.t. to 20-fold above that of the existing pGFP-PAC replicon. GFP counts showed identical trends to total GFP fluorescence, with a 16-fold increase in the number of GFP positive cells at 2 hr p.t. in comparison to constructs lacking the 5' HH ribozyme (data not shown). An increase in the number of GFP cells of up to 1.5-fold continued throughout the experiment (data not shown).

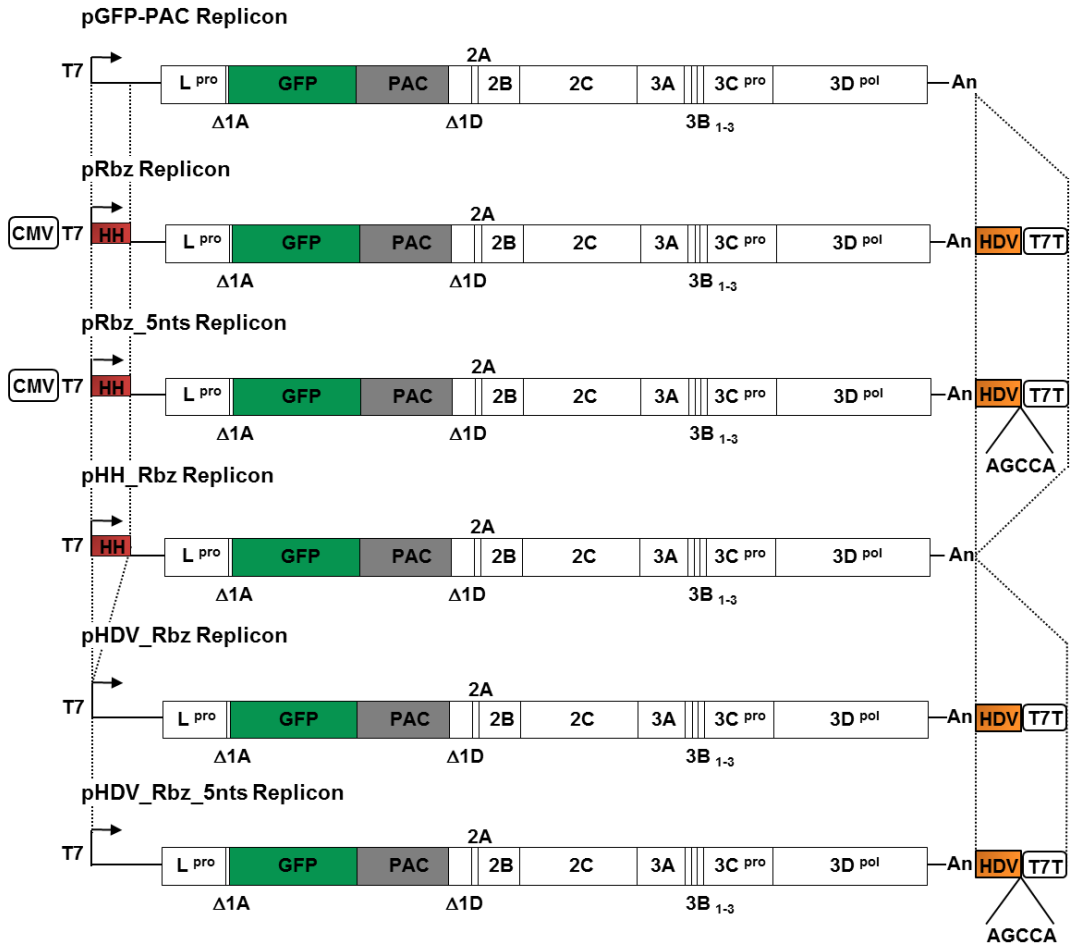


Figure 4.1 Improved FMDV Replicon System. An improved version of the FMDV replicon (pGFP-PAC, Tulloch *et al.*, 2014a) was created, the pRbz replicon, containing both T7 and CMV promoters, self-cleaving ribozymes and a T7TΦ sequence. To test the activity of each ribozyme independently from each other, replicons containing either the 5' Hammerhead (HH) or the 3' Hepatitis Delta Virus (HDV) ribozyme were constructed; pHH_Rbz and pHDV_Rbz replicons. A 5nt sequence, AGCCA, was inserted into the 3'-end of the HDV ribozyme to optimise the cleavage activity. This generated the pRbz_5nts and pHDV_5nts replicons.

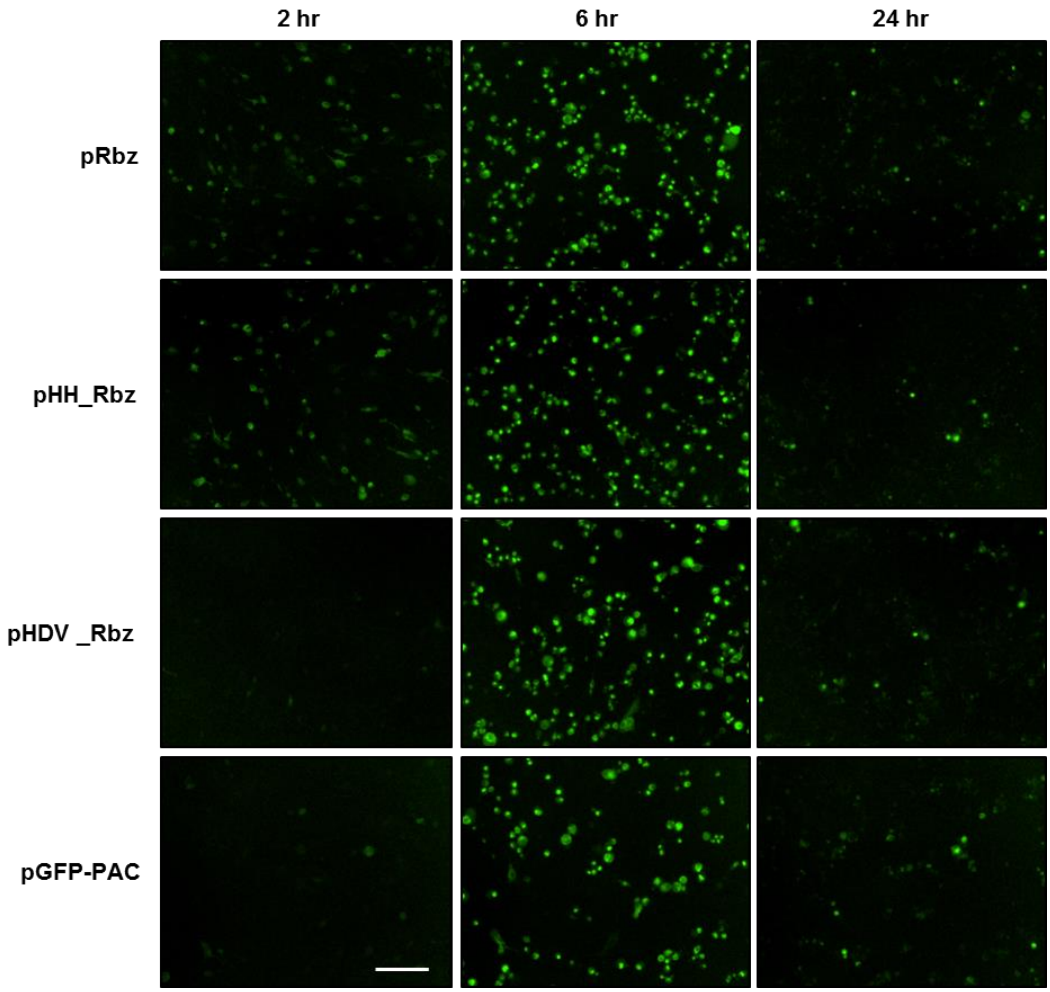


Figure 4.2 GFP expression generated by FMDV replicons containing self-cleaving ribozymes. GFP expression from replicons containing both (pRbz replicon) or each ribozyme alone (pHH or pHDV_Rbz replicons), were compared to the unmodified pGFP-PAC replicon. Following transfection of BHK-21 cells with 1 µg transcript RNA from each construct, fluorescent images were captured at 2 hr intervals over a 24 hr period using the IncuCyte ZOOM imaging system. Data is representative of one image captured (from nine) at 2, 6 and 24 hr post-transfection, and was obtained from 3 independent transfections, with 4 replicates per transfection. Scale bar represents 100 µm.

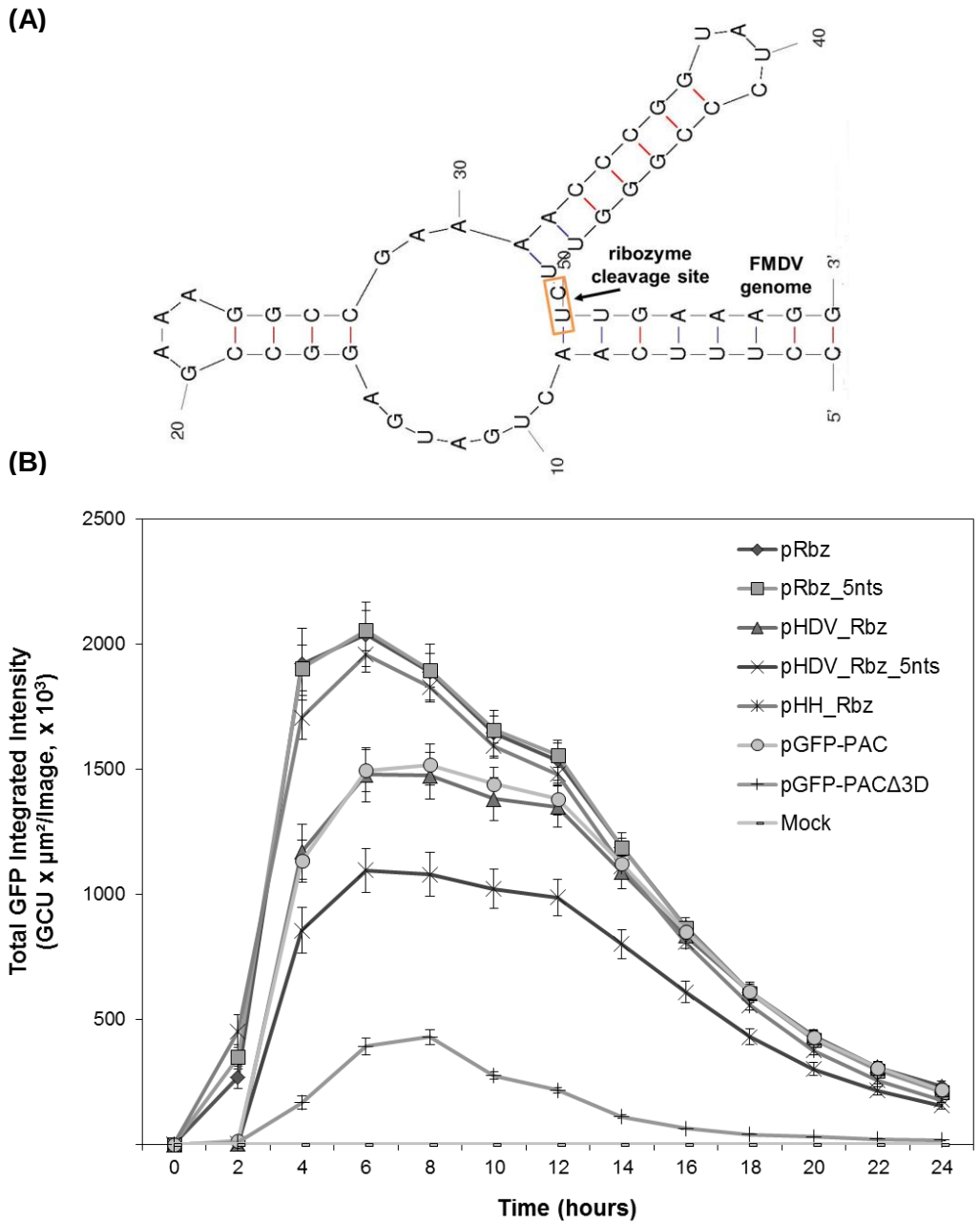


Figure 4.3. The effect of ribozymes on FMDV replicon-derived GFP expression. (A) Predicted structure of the HH ribozyme connected to the 5'-end of the FMDV genome (generated using mfold; Zuker, 2003). Numbering begins at the first nucleotide of the HH ribozyme C-1. Orange box and arrow indicate cleavage site. (B) GFP expression from BHK-21 cells transfected with transcript RNA derived from mock, pRbz, pRbz_5nts, pHDV_Rbz, pHDV_Rbz_5nts, pHH_Rbz, pGFP-PAC and pGFP-PAC-Δ3D replicons. Images were captured at 2 hr intervals over a 24 hr period and the GFP fluorescence quantified for each replicon construct: data shown as the total integrated GFP intensity. Data points/error bars are derived from 3 independent transfections, with 4 replicates for each transfection.

4.4 CMV promoter-driven transcription of replicon RNA

To test the efficiency of promoter-driven transcription of replicon RNA, plasmid DNA from both the pRbz and pRbz_5nt replicons (the only replicons containing a fully-functional CMV promoter) was transfected into BHK-21 cells and GFP fluorescence measured. Fluorescence was not detected until 6 hr p.t. with the peak being reached ~12 hr later at 18 hr p.t. (Figure 4.4): in contrast to replication kinetics displayed by linear transcript RNA (Figure 4.3B). The total GFP fluorescence differed by a ~20-fold decrease at the peak of replication for the pRbz replicon (18 hr p.t., Figure 4.4), when compared to cells transfected with transcript RNA from this construct (6 hr p.t., Figure 4.3B). Analysis of GFP fluorescence generated by transcript RNA showed the signal arising at 2 hr p.t., whereas upon transfection of replicon plasmid DNA there was a lag in detection of 4 hr, with the signal being ~2.5-fold lower on average at this early time point (Figure 4.3B, Figure 4.4). There was ~1.5-fold difference in the total fluorescence derived from the pRbz and pRbz_5nt replicons, with the pRbz replicon producing a higher signal. The number of GFP positive cells displayed a similar trend to the fluorescence data obtained, with pRbz giving rise to a higher GFP cell count (data not shown). This correlates with the previous fluorescence data shown in Figure 4.3B, with replicon RNA containing the HDV_5nt modification replicating less efficiently. In contrast, transfection of replicon RNA containing both ribozymes (pRbz and pRbz_5nt) did not show such a difference (Figure 4.3B).

4.5 Replication kinetics of replicon RNA synthesised using the T7 Φ terminator

A T7TΦ sequence was inserted directly after the HDV ribozyme in order to increase the efficiency of T7 RNA polymerase transcription. To test the effectiveness of the T7TΦ, plasmids were linearised using SspI (located ~3.2 kb upstream of the T7TΦ sequence, A.3) and transcript RNA synthesised using T7 RNA polymerase. The pGFP-PAC and pHH_Rbz replicons, which do not contain a T7TΦ, were linearised with both HpaI and SspI to function as controls. Reaction products were run on a non-denaturing 0.8 % (w/v) agarose gel to check integrity and yield. Transcript RNA from the HpaI-linearised pGFP-PAC and pHH_Rbz replicons displayed a single, sharp band, whereas RNA synthesised from templates containing the T7TΦ and linearised with SspI exhibited three bands, one of which co-migrated with RNA from both the pGFP-

PAC and pHH_Rbz replicons (Figure 4.5A). RNA synthesised from SspI-linearised pGFP-PAC and pHH_Rbz also contained multiple transcripts, with two bands present (Figure 4.5A). These results suggested that both the SspI enzyme and the T7TΦ were responsible for the generation of multiple transcripts. These bands could have arisen from i) inefficient termination at the T7TΦ and read through to the end of the template DNA, ii) successful termination and iii) cleavage by the HDV ribozyme leading to transcripts of the correct size (compare HpaI transcripts which are the correct length, Figure 4.5A). Irrespective of these results, transcript RNAs generated displayed replication kinetics comparable to those shown in Figure 4.3B and did not seem to affect replication efficiency (data not shown).

Although replication did not seem to be affected by the presence of multiple transcripts, other RE sites were used to generate transcripts to establish whether distinct bands could be generated for future use. The pRbz_5nts replicon was linearised with AscI, BstBI and SspI (A.3) for comparison. Transcripts produced by both the AscI and BstBI-linearised template were sharp, distinct bands, in contrast to the multiple transcripts seen when SspI-linearised template was used (data not shown). The AscI site is situated directly after the 3' terminus of the HDV ribozyme, leading to the removal of the T7TΦ following linearisation; however data obtained suggests the T7TΦ does not improve termination efficiency and therefore is not required. Thus, it was decided that all replicons would be linearised using AscI during future experiments.

To compare the use of a linear template when synthesising transcript RNA with that of circular plasmid DNA containing a T7TΦ, RNA was made using AscI linearised pRbz_5nts as a template, together with plasmids from the pRbz, pRbz_5nts, pHDV_Rbz and pHDV_Rbz_5nts replicons. RNA synthesised from circular plasmid DNA appeared as a smear when resolved on a 0.8 % (w/v) agarose gel, with RNA made from a linear template displaying a bright, distinct band (data not shown). GFP expression from BHK-21 cells transfected with RNA made from the linear template displayed kinetics as expected, whereas RNA produced from circular plasmids gave GFP intensities ~70-fold lower at 2 hr p.t. and ~10-fold lower thereafter on average (Figure 4.5B).

4.6 Conclusions

- Insertion of a HH ribozyme to the 5'-end of a FMDV cDNA increased GFP detection by 20-fold initially, with an overall increase of ~1.5-fold throughout the 24 hr time-course. Constructs containing the 3' HDV ribozyme exhibited kinetics similar to the existing pGFP-PAC replicon, however when combined with the HH ribozyme exhibited kinetics almost identical to the pHH_Rbz replicon which contained the HH ribozyme alone.

- DNA transfection of replicon plasmids displayed poor GFP fluorescence when compared to transcript RNA. There was a lag in GFP detection of ~4 hr and the peak of replication was not reached until 18 hr p.t.: a 12 hr delay in comparison to transcript RNA (6 hr p.t.).

- Use of the T7TΦ to generate transcript RNA from plasmids resulted in RNA smearing and a large reduction of fluorescent signal initially of ~70-fold, with a ~10-fold decrease throughout.

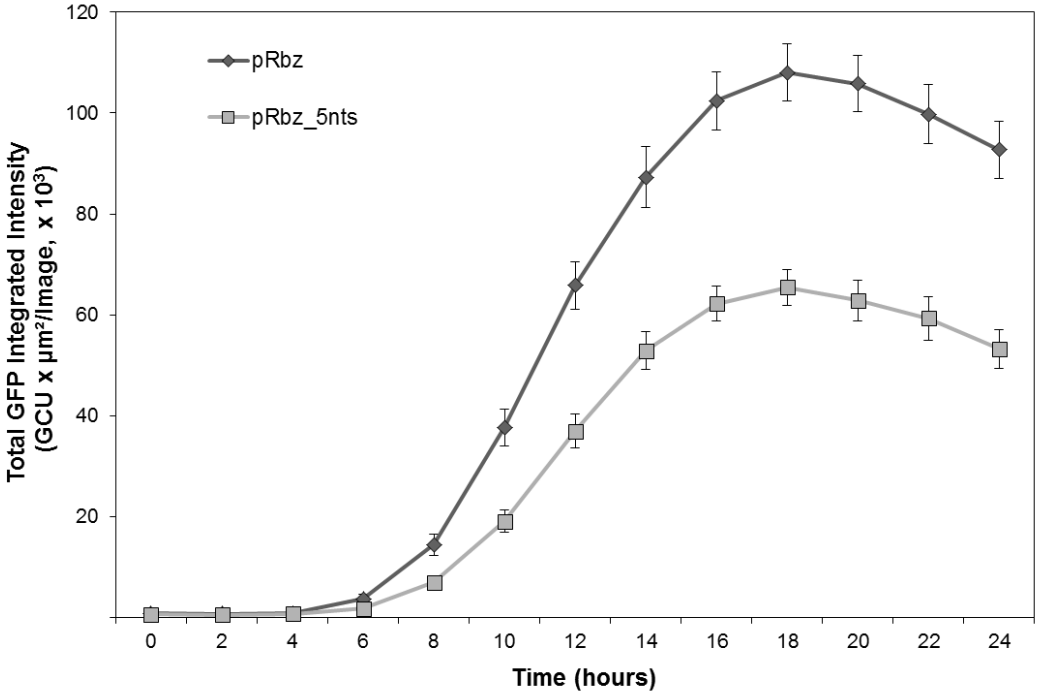


Figure 4.4. GFP expression derived from transfection of FMDV replicon plasmid DNA containing ribozymes. BHK-21 cells were mock transfected, together with cells transfected with plasmid DNA from the pRbz, and pRbz_5nts replicon. At the time points indicated images were captured over a 24 hr period and the GFP fluorescence quantified for each replicon construct: data shown as the total GFP integrated intensity. Data points/error bars are derived from 2 independent transfections, with 3 replicates for each transfection.

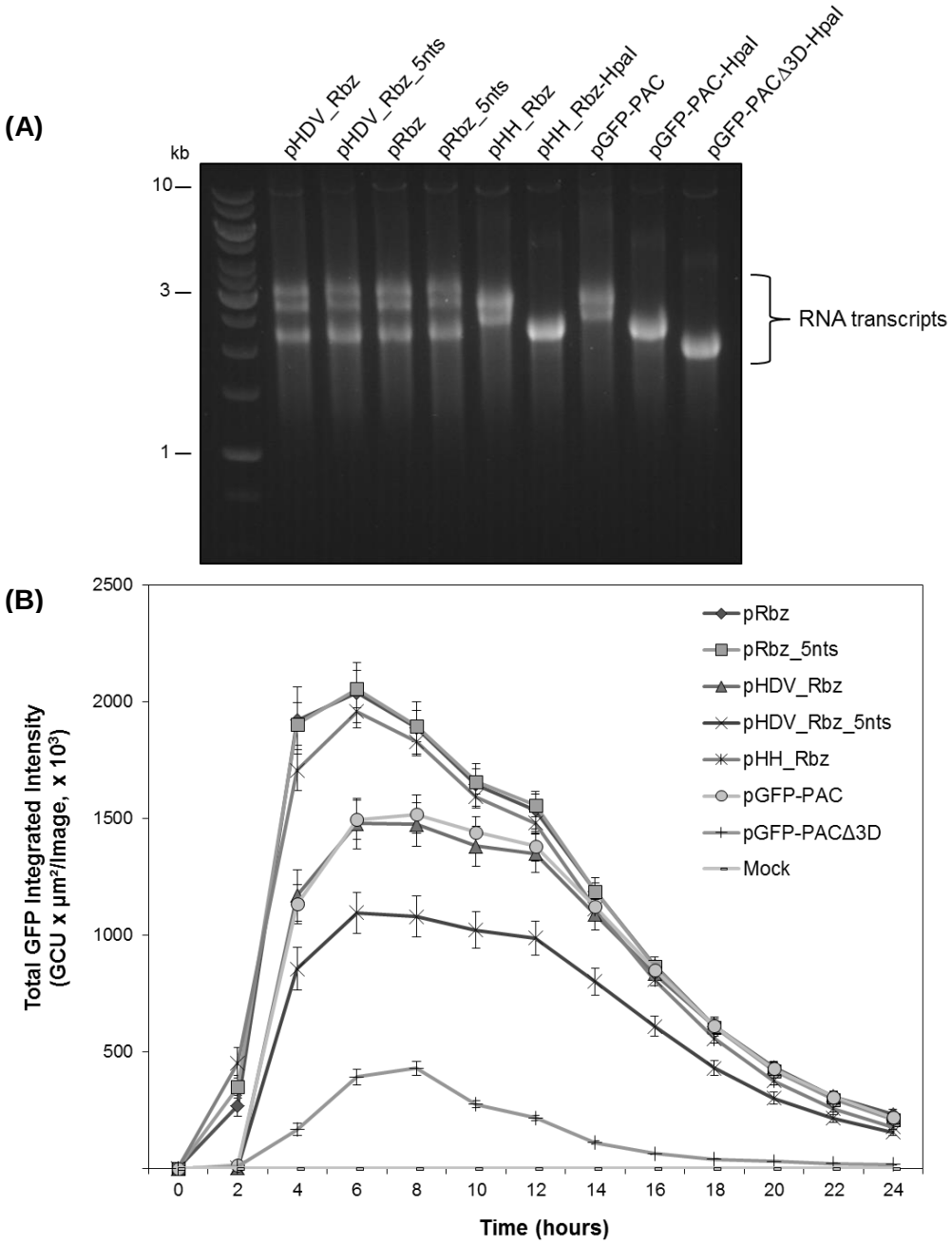


Figure 4.5. Replicon RNA synthesised using the T7TΦ. (A) RNA transcripts prepared from template DNA linearised with SspI or HpaI (as indicated). (B) GFP expression from mock-transfected BHK-21 cells, together with cells transfected with transcript RNA derived from the pRbz, pRbz_5nts, pHDV_Rbz and pHDV_Rbz_5nts replicons- created using the T7TΦ. The pRbz_5nts replicon was linearised prior to T7 transcription (as indicated), acting as a control. At the time points shown images were captured and the GFP fluorescence quantified for each replicon construct over a 24 hr period: data represents the total integrated GFP intensity. Data points/error bars are derived from 2 independent transfections, with 4 replicates for each transfection.

4.7 Discussion

4.7.1 Addition of self-cleaving ribozymes to the 5' and 3'-ends

Replicons containing *cis*-active ribozymes at the 5' and 3'-ends of a FMDV cDNA were made *via* cloning in order to improve replication kinetics. Enhanced GFP fluorescence was observed for constructs containing the HH ribozyme alone, or in combination with the 3' HDV ribozyme, when compared to the pGFP-PAC replicon, which does not contain self-cleaving ribozymes. In contrast, pGFP-PAC showed delayed kinetics, suggesting the extra sequences at the 5'-end were contributing to the observed lag in replication. This is in agreement with the work of others who have shown that removal of non-viral guanosines at the 5'-end of picornavirus genomes improves replication kinetics (Herold & Andino, 2000; Israelsson *et al.*, 2014). The overall trend in replication was similar for all constructs tested; the peak of fluorescence occurred at 6 hr p.t. and declined thereafter. The trend only differed dramatically within the first 2 hr p.t., which is in accordance with results from other groups who demonstrated that RNA transcripts accumulated to similar levels, following an initial lag, when the 5'-end contained extra guanosine residues (Herold & Andino, 2000; Sarnow, 1989). It is known that the viral peptide $3B^{VPg}$ is uridylated during initiation of replication, and this ($3B^{VPg}$-pUpU) acts as a primer for RNA synthesis (Paul *et al.*, 1998). Therefore, it was expected that removal of non-viral guanosines would increase the synthesis of positive-strand RNA, as the negative-strand template would have authentic ends ready for attachment of uridylated $3B^{VPg}$ to the 5' terminus. The lag in replication observed with pGFP-PAC could indicate the time taken to remove these extra sequences, as the GFP fluorescence increases exponentially from 2 hr p.t. - the time fluorescence is initially detected for this construct. It is also important to note that the delay is comparable to the replication-incompetent pGFP-PACΔ-3D replicon, which also contains extra guanosines. Therefore, these sequences must delay replication. This is also in agreement with findings from Herold and Andino (2000). PV luciferase replicons containing an active or inactive HH ribozyme and a form without the HH ribozyme, displayed luciferase activity comparable to GFP fluorescence derived from constructs listed in the present study. Initial luciferase activity from the PV replicon containing the active HH ribozyme increased exponentially immediately, whereas PV replicons without authentic ends displayed a sigmoid curve, with a lag of ~1.5 hr, before increasing

exponentially at 2 hr. Activity detected during the first 2 hr from these replicons was directly comparable to both untreated cells, and cells treated with guanidinium hydrochloride (GuHCl): a potent inhibitor of poliovirus replication. The authors suggested this activity was due to translation of the input RNA, which is in contrast to the GFP fluorescence obtained from the pGFP-PAC and replication-incompetent form during a similar time-frame in this study. Little-to-no fluorescence was detected from 0-2 hr, with fluorescence increasing thereafter. The replication curve observed for pGFP-PAC-Δ3D displayed a similar trend as pGFP-PAC, albeit at much lower levels, whereas the luciferase activity from inhibited PV replicons remained at a constant level from 2 hr onwards. This could reflect the difference in assays used, as GFP fluorescence is detected in live cells in real-time, whereas extracts are harvested, prepared, and then analysed during a luciferase assay. Regardless of this slight discrepancy, the data reported by Herold and Andino (2000) and observations made during this study suggest the extra guanosines are removed by an unknown mechanism, or, initiation of positive-strand synthesis can occur internally within the 3'-end of the negative-strand, but at a lower efficiency.

A HDV *cis*-active ribozyme was introduced at the 3'-end of the FMDV replicon genome directly after the poly(A) tail, replacing a string of non-viral cytosine residues (C_{16}) to generate a precise 3'-end. Sarnow (1989) found non-viral extensions to the 3'-end of PV transcripts reduced infectivity, and when these extensions were reduced to around 4 nucleotides in length infectivity of transcript RNA increased. GFP fluorescence derived from the pHDV_Rbz replicon displayed similar kinetics as the pGFP-PAC replicon and had no effect on GFP fluoresence. This highlights the importance of a precise 5'-end for efficient replication and suggests an exact 3'-end is not essential. Enterovirus 71 cDNA plasmids encoding differing lengths of poly(A) and 3' terminal sequences did not have any effect on the infectivity of transcripts in cell culture when the poly(A) tract was of sufficient length (Lazouskaya *et al.*, 2014). Thus, it seems there is more plasticity at the 3'-end of picornavirus genomes, but authenticity of the 5'-end is more important.

The insertion of a 5 nt sequence to increase the cleavage efficiency of the HDV ribozyme did not increase GFP fluorescence in either the pRbz_5nts or the pHDV_Rbz_5nts replicons, despite previous studies reporting enhanced cleavage activity with this modification (Ghanem *et al.*, 2012; Perrotta & Been, 1998). These 5 nt are well conserved amongst clinical HDV isolates and were

shown to increase the cleavage efficiency under denaturing and physiological salt conditions (Perrotta & Been, 1998). Ghanem *et al*., (2012) reported a 10-fold increase in rescue events and reporter gene expression from mini-genome cDNAs containing this 5 nt addition to the HDV sequence. A correct 3'-end is more critical during replication of negative sense RNA viruses as RNAs with 3' overhanging ends are not replicated (Ghanem *et al.*, 2012; Hoenen *et al.*, 2011) and could therefore explain the differences observed with our replicon system. It has also been demonstrated that flanking sequences can inhibit cleavage activity of the HDV ribozyme with the authors suggesting this could be due to these sequences interfering with structure formation (Perrotta & Been, 1990). There is also the possibility that the stem-loop structures within the 3' UTR could be affecting the correct formation of the HDV ribozyme. The HH ribozyme is also located at the terminus of a structured region and cleavage activity seemed unaffected, however the HDV ribozyme is highly structured in comparison to the HH ribozyme and this could explain the differences observed. Despite no measurable difference detected in replication by the presence of the HDV ribozyme within the FMDV replicon during fluorescent reporter studies, perhaps this ribozyme will play a better role during rescue events following insertion of the capsid sequences by collaborators at Pirbright by increasing infectivity, as seen by Sarnow (1989) with PV transcripts.

4.7.2 Replication derived from plasmid DNA transfection

Transfection of cDNA plasmids encoding viral sequences into susceptible cells resulted in the recovery of virus (Racaniello & Baltimore, 1981; Semler *et al.*, 1984). It was considered that GFP expression could be monitored following transfection of FMDV replicon DNA, removing the need for costly and time-consuming RNA transcription. Transfection of replicon DNA resulted in an initial decrease in GFP expression of ~2.5-fold and a ~20-fold decrease at the peak of replication when compared to cells transfected with transcript RNA. The peak of replication occurred at 18 hr p.t., 12 hrs later than observed with transcript RNA and the GFP counts obtained were also considerably lower suggesting a decrease in transfection efficiency. Semler *et al*., (1984) observed an 18 to 250-fold increase in PV rescue when SV40 transcription and replication signals (promoter, enhancer and origin of replication sequences) were inserted into cDNA plasmids encoding PV. This increase was cell dependent and the largest increase was found when COS-I cells were used. These cells express the SV40

T antigen which is known to increase expression of plasmids containing the SV40 origin of replication. The vector backbone used during transfection experiments contains both the CMV promoter and SV40 signals, therefore it was expected that transcription would proceed as normal. It is possible that expression is cell-line dependent, as reported by Semler *et al.,* (1984). A study by Liu *et al.*, (1997), compared the transcription efficiencies of several enhancer/promoter elements in different cell lines *via* CAT assays of cell extracts, and found a large variation in plasmid expression depending on the cell line used. In agreement with Semler *et al*., (1984), HEK 293 and COS cells provided the largest expression from plasmids containing the SV40 or CMV elements, with BHK-21 cells demonstrating lower expression on average (Liu *et al.*, 1997). Perhaps GFP expression could be increased using these cell lines, however initial testing of transfection efficiencies in differing cell lines resulted in low GFP expression when replicon RNA was introduced into HEK 293 cells, with BHK-21 cells providing the best expression and transfection efficiency (data not shown). It seems the best option is the synthesis of RNA transcripts, as this provided the best expression and transfection efficiency with replication kinetics similar to viral growth curves, which is the desired outcome of using this system.

4.7.3 The use of a T7 Φ terminator sequence for RNA synthesis

It was hoped that use of a T7TΦ would aid in termination of T7 RNA polymerase during synthesis of RNA transcripts and eliminate the need for linearisation, which in turn would reduce costs and save time. Following linearisation of constructs with HpaI (no T7TΦ sequence) or SspI (located ~3.2 kb downstream of the T7TΦ sequence), RNA was synthesised and transcripts analysed by non-denaturing agarose gels. RNA synthesised from SspI-linearised template DNA contained multiple transcripts, whereas lanes corresponding to transcripts made from HpaI-linearised contained single bands. This seemed to be due to both the T7TΦ and the use of SspI, as transcripts made from SspI-linearised templates which did not contain T7TΦ also contained multiple bands. Synthesis of extraneous RNA transcripts following linearization with enzymes which generate 3’ overhanging ends has been reported, so called “snap-back” transcription where T7 RNA polymerase can turn to use the 3’ protruding end as a second template (Schenborn & Mierendorf, 1985); however all enzymes used during this study to linearise templates contained blunt or 5’ protruding ends and therefore this cannot explain the observed results. RNA synthesised using circular DNA as

template appeared smeared following examination on an agarose gel, and replicated poorly with a ~10-fold decrease observed during the course of the experiment. This could be due to the circular template affecting the transcription reaction, however the use of supercoiled plasmid DNA as a template by T7 RNA polymerase has been described (Du *et al.*, 2012; Jeng *et al.*, 1990), and no major effect on transcription termination was found. T7TΦ sequences fall into two classes. Class I terminators, like T7TΦ, are similar to the intrinsic terminators of *E.coli* RNA polymerase and form stem-loop structures in the nascent RNA, causing the polymerase to disengage from the template strand, halting transcription. Class II terminators do not involve secondary structures but share a common sequence 7 bp long within the template DNA which causes transcription termination (He *et al.*, 1998). It is known that sequences up- and downstream of the T7TΦ sequence are important for efficiency of termination , as well as stable hairpin formation (Jeng *et al.*, 1990, 1992). It is possible that the large amount of secondary structure upstream of the terminator within the 3' UTR of FMDV and the HDV ribozyme is inhibiting secondary structure formation and affecting termination when either linear or circular templates are used. It has also been demonstrated that RNA lacking secondary structure at the 3'-end can be accepted as a second template by T7 RNA polymerase and extended (Triana-Alonso *et al.*, 1995). SspI-linearised templates contain ~3.2 kb of vector sequence between the T7TΦ and SspI site and therefore may encourage formation of these aberrant transcripts. Due to these results, it was decided that future experiments would involve linearisation of the template DNA with AscI to ensure transcription terminated at the 3'-end of the HDV ribozyme.

The next chapter describes further modification to the replicon system through site-directed mutagenesis of the potential L^{pro}-GFP cleavage site (described in Chapter 3) or replacement of GFP with an alternative reporter protein.

Chapter 5: Fluorescent Reporter Proteins.

5.1 Introduction

Aequorea victoria GFP (Aq.GFP) and its many derivatives have been used widely as a visual marker in a large variety of organisms for protein localisation and gene expression studies (Chalfie, 1995; Chudakov *et al.*, 2010; Tsien, 1998). Aq.GFP was firstly discovered by Shimomura *et al.*, (1962), however, when the gene was cloned by Prasher *et al.*, (1992) and expressed in other organisms such as *E.coli* (Inouye & Tsuji, 1994) this led to a huge advance in the field of molecular biology.

Since the discovery of GFP, many more GFP-like proteins have been discovered in bioluminescent species with a large spectral diversity and improved characteristics (Chudakov *et al.*, 2010; Rizzo *et al.*, 2009). GFPs with enhanced fluorescent properties have been isolated from Anthozoan species such as the sea pansy *Renilla mulleri* and the sea pen *Ptilosarcus gurneyi* (Peelle *et al.*, 2001). These GFPs share ~25 % aa sequence identity with Aq.GFP and were found to be more fluorescent when expressed in mammalian cells (Peelle *et al.*, 2001).

Attempts to shift the fluorescent properties of Aq.GFP beyond yellow to red proved to be very unsuccessful. Therefore, researchers sought red FPs from animals such as corals and sea anemones (Rizzo *et al.*, 2009). All red FPs discovered thus far have required a large amount of optimisation in comparison to other FPs. The tendency to oligomerise and form tetramers, as well as poor maturation and fluorescent properties, has hindered their use as reporter proteins (Chudakov *et al.*, 2010; Rizzo *et al.*, 2009; Shaner *et al.*, 2004). The monomeric cherry FP (mCherry) arose from mRFP1, a derivative of DsRed FP from *Discosoma striata*. Mutations to mRFP1, including the addition of C-and N-terminal aa sequences from GFP, improved the maturation speed, photostability and tolerance of C- and N-terminal tags (Shaner *et al.*, 2004). As with *Renilla* and *Ptilosarcus* FPs, mCherry originates from Anthozoan species and should therefore have low sequence similarity with the Hydrozoan Aq.GFP.

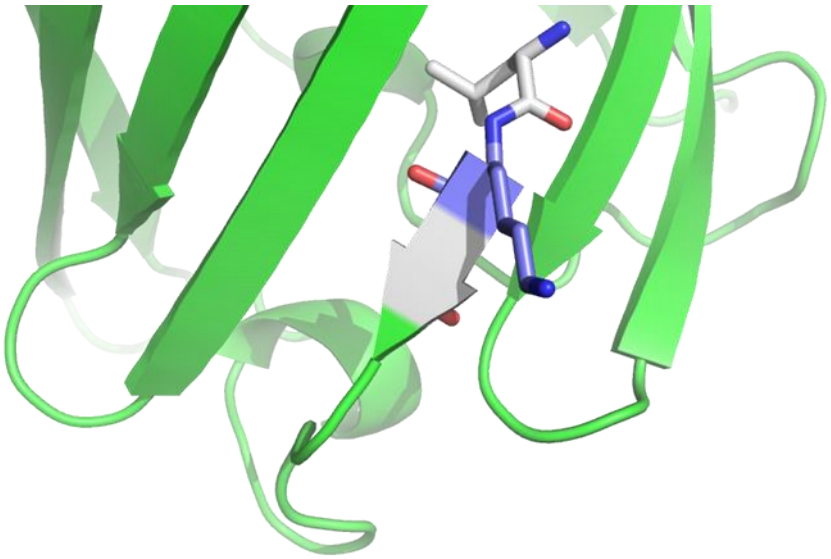
During initial studies in Chapter 3 it was evident that Aq.GFP was undergoing cleavage and evidence suggested this involved FMDV L^{pro}. This affected quantitative studies to distinguish between replicating and non-replicating forms

through the measurement of GFP fluorescence. Due to this unforeseen cleavage of Aq.GFP, it was necessary to mutate the putative cleavage site, or, to replace Aq.GFP with a suitable alternative. Therefore, the aim of this chapter was to firstly carry out mutagenesis to prevent GFP cleavage by L^{pro} and, since this proved to be unsuccessful, to replace GFP with mCherry FP (Shaner *et al.*, 2004), or, the reportedly brighter *Ptilosarcus gurneyi* GFP (Pt.GFP; Peelle *et al.*, 2001), both of which might well be suitable substitutes due to a lower level of sequence identity with Aq.GFP.

5.2 Mutagenesis of the potential L^{pro} cleavage site within Aequorea GFP

Analysis of GFP fluorescence derived from the pGFP-PAC WT replicon in comparison to the pLL-GFP-PAC replicon lacking L^{pro} led to the observation of a 2-fold decrease in GFP signal. Further investigation *via* western blotting of cell extracts with anti-GFP antibodies and subsequent examination of the GFP-PAC sequence, strongly suggested a potential L^{pro} cleavage site within GFP (Tulloch *et al*., 2014a, Figure 5.1). An alignment of the L^{pro}/1A cleavage site and the predicted L^{pro} cleavage site within Aq.GFP are shown in Figure 5.1B, with the potential residues involved highlighted on the Aq.GFP crystal structure (Ormö et al. 1996, Figure 5.1A). Bioinformatics analysis suggested that cleavage at this site within the GFP-PAC fusion protein would produce a cleavage product of ~42 kDa (data not shown). Western blot analysis demonstrated the major cleavage product was indeed ~42 kDa, and corresponded well to Lys^{127}-Gly^{128} within β-sheet 6 of Aq.GFP (Ormö *et al.*, 1996; Yang *et al.*, 1996). Therefore, Lys 127 was mutated to either Leu, Pro, His or Arg using site-directed mutagenesis of the pRbz replicon with primers spanning the cleavage site containing the motif CNC (Table A.1), where 'N' is any nucleotide (Figure 5.1B). Following sequence analysis of potential positive clones, two mutants were generated comprising both a proline and histidine in place of Lys127. They were designated as the pGFPmut_K127P and pGFPmut_K127H replicons, respectively. A polymerase deletion form of the pRbz replicon containing self-cleaving ribozymes was also generated by RE digestion with MluI, to ensure that subsequent analysis of replicons was comparable.

(A)



(B)

L/1A -RKLK GAGS-
GFP -IELK GIDF-

ATCGAGCTGAAGGGCATCGACTTC
I E L K G I D F

CNC = Leu (L), Pro (P), His (H), Arg (R).

Figure 5.1. Mutagenesis of possible L^{pro} cleavage site within *Aequorea* GFP. (A) Predicted L^{pro} cleavage site on β-sheet 6 of Aq.GFP at residue Lys127 (blue), between Leu 126 and Gly 128 (grey). Residues are represented as ball-and-stick. Image generated using PyMOL (Schrodinger, 2010). (B) The predicted cleavage site is shown with the proposed strategy to mutate K127 to L, P, H or R amino acids in order to abolish cleavage activity.

5.3 GFP expression derived from replicons bearing a mutation within *Aequorea* GFP to prevent cleavage by FMDV L^{pro}

To investigate whether Aq.GFP fluorescence was affected by mutagenesis of Lys 127, Aq.GFP fluorescence of the pRbz, pGFPmut_K127P, pGFPmut_K127H and pRbzΔ3D replicons was monitored over a 24 hr period using the IncuCyte ZOOM fluorescent microscope. Replicon plasmid DNAs were linearised with AscI, RNA synthesised using T7 RNA polymerase and 1 µg of transcript RNA transfected into BHK-21 cells. Virtually no fluorescence could be observed in cells transfected with both the pGFPmut_K127P and pGFPmut_K127H constructs (Figure 5.2, 5.3). The pRbz replicon displayed kinetics as expected. Rather unexpectedly, the pRbzΔ3D replicon also exhibited no fluorescence, in contrast to the existing replication-incompetent form, pGFP-PAC-Δ3D, which exhibits Aq.GFP fluorescence 2-fold lower than the replication-competent pGFP-PAC. Transcript RNA derived from these replication-incompetent forms differ only in the presence or absence of ribozymes at the 5' and 3'-ends. Data described in Chapter 4 shows that replicons containing the 5' HH ribozyme replicate considerably better than those lacking this ribozyme, and hence contain non-viral guanosines at their 5'-end. Therefore, this data implied removal of these residues resulted in either no or very little translation. The cell morphology in both Aq.GFP-mutants and the polymerase deletion was characteristic of the rounded-up morphology associated with CPE during viral infection (Figure 5.2) and suggested active replication had occurred; however all contained an intact L^{pro}, hence this observation could have been solely due to L^{pro} – which induced cell death through inhibition of host cell translation (Devaney *et al.*, 1988). Aq.GFP cell counts displayed trends identical to that observed for the total Aq.GFP fluorescence (data not shown).

5.4 Cleavage of *Aequorea* GFP and synthesis of viral proteins during replication

Due to the findings presented above, western blotting of cell extracts prepared from replicon-transfected cells were analysed with anti-GFP antibodies to check whether Aq.GFP cleavage occurred following mutagenesis of the possible cleavage site, and to assess the production of viral proteins during replication to ensure replicons were actively replicating. While mutagenesis of Lys 127 seemed to decrease the amount of Aq.GFP cleaved, it did not abolish cleavage activity

completely (Figure 5.4). FMDV 3D and 2A (data not shown) expression confirmed that the pGFPmut_K127P mutant was actively replicating and suggests the lack of fluorescence was due to the mutation within Aq.GFP.

5.5 Replacement of *Aequorea* GFP with *Ptilosarcus* GFP and mCherry fluorescent proteins

Due to mutagenesis of the putative L^{pro}-GFP cleavage site proving unsuccessful, a suitable alternative was investigated. The replicon system was designed to study FMDV replication to screen for attenuated forms as potential vaccine candidates, and for a better understanding of aspects of the viral life cycle, such as persistent infection. In order to facilitate such studies into persistence, a strategy had been devised to replace Aq.GFP with a brighter FP before this became a necessity due to unexpected cleavage by L^{pro}. A brighter FP should aid in the detection of low-level replication and increase the detection limits of experiments undertaken. As the IncuCyte ZOOM imaging system can detect both green and red fluorescence, mCherry FP was also investigated as a suitable replacement.

Almost all FPs derived from Aq.GFP will contain the potential L^{pro} cleavage site due to identical sequence composition, except for mutations to the chromophore which yield different spectral properties (Chudakov *et al.*, 2010; Rizzo *et al.*, 2009). As expected, amino acid sequence alignment of cyan and yellow FPs with the parental GFP confirmed the presence of this site and any FPs originating from Aq.GFP were ruled out (Figure 5.5). Pt.GFP and mCherry FPs (Peelle *et al.*, 2001; Shaner *et al.*, 2004) displayed lower sequence similarity with the potential L^{pro} cleavage site than Aq.GFP, and were therefore explored as substitutes.

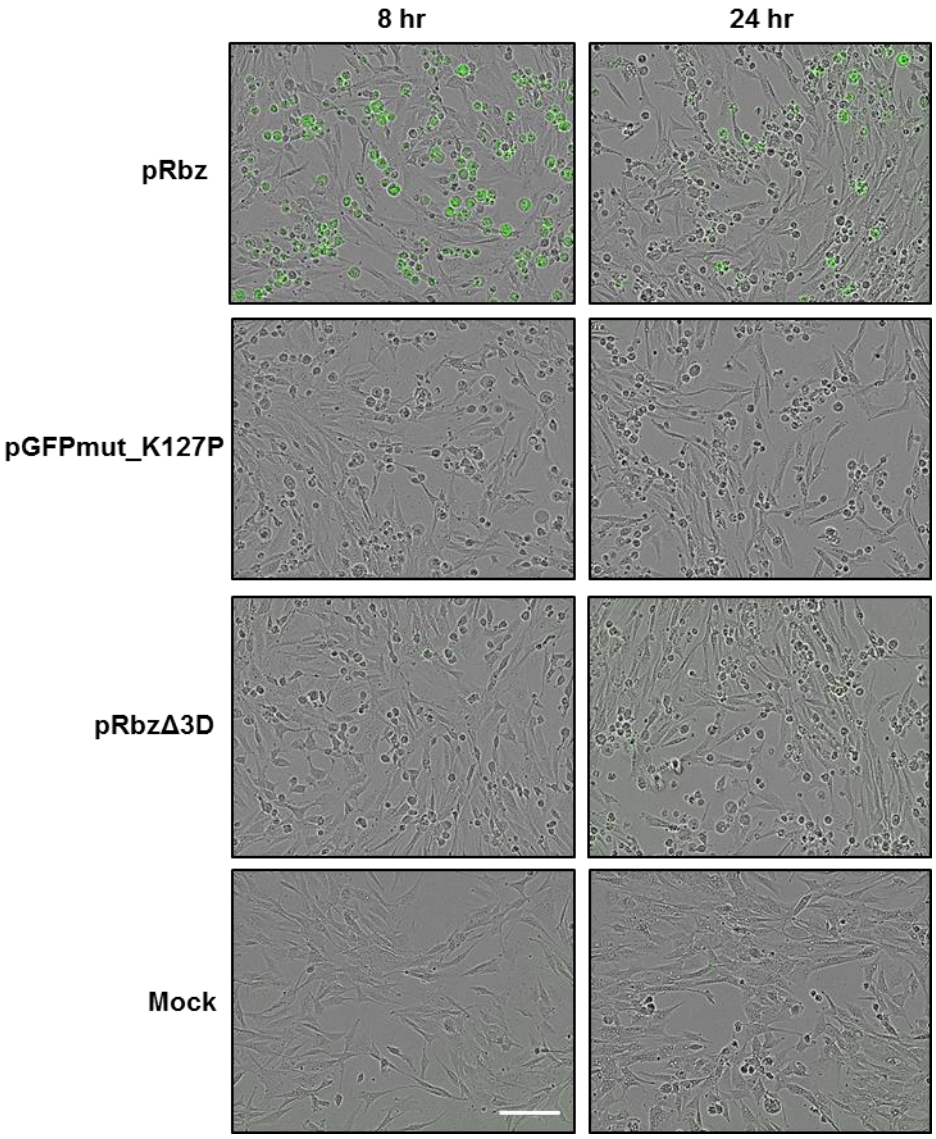


Figure 5.2. GFP expression in FMDV replicons containing the Lys127Pro mutation within *Aequorea* GFP. Expression of Aq.GFP from the pGFPmut_K127P and pGFPmut_K127H (data not shown) replicons was compared to WT and pRbzΔ3D replicons. Following transfection of BHK-21 ells with transcript RNA from each construct, fluorescent images (nine/well) were captured at 2 hr intervals over a 24 hr period using the IncuCyte ZOOM imaging system. Data represents a captured image at both 8 and 24 hr post-transfection, and was obtained from 2 independent transfections, with 4 replicates per transfection. Scale bar represents 100 μm.

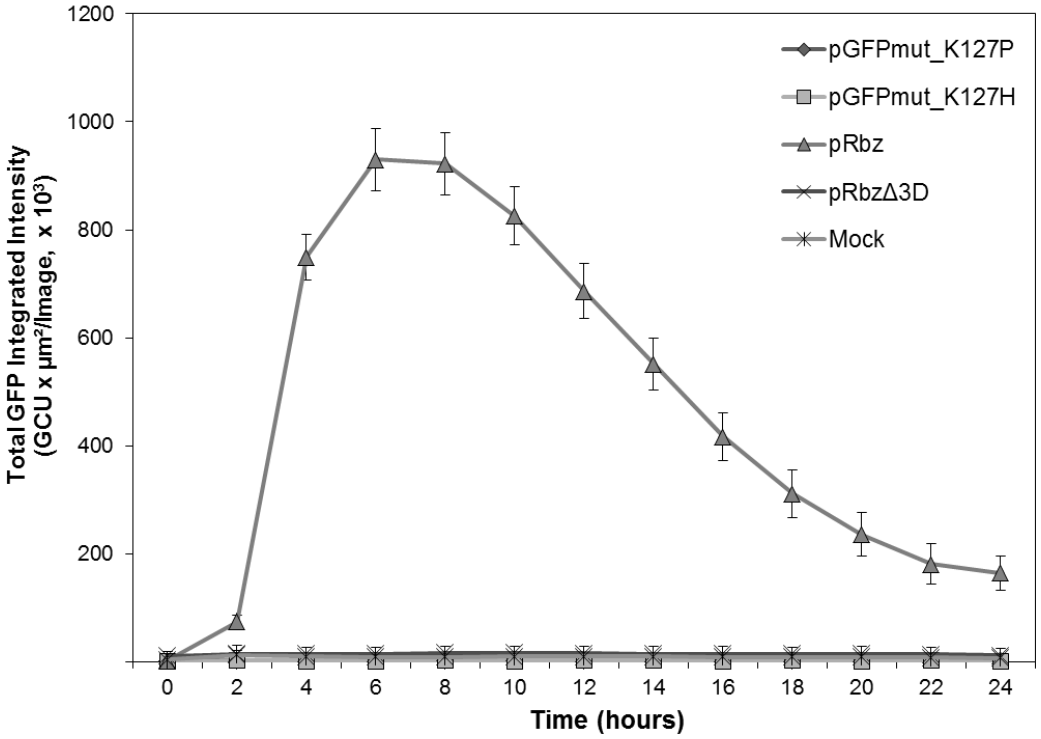


Figure 5.3. GFP fluorescence derived from replicons containing a mutation within *Aequorea* GFP to prevent cleavage by L^{pro}. Data from mock-transfected BHK-21 cells is shown, together with cells transfected with transcript RNA derived from the pGFPmut_K127P, pGFPmut_K127H, pRbz replicons and the polymerase deletion pRbzΔ3D. At the time points indicated, images were captured over a 24 hr period and the GFP fluorescence quantified for each replicon construct: data shown as the total integrated Aq.GFP intensity. Data points/error bars shown are derived from 2 independent transfections, with 4 replicates for each transfection.

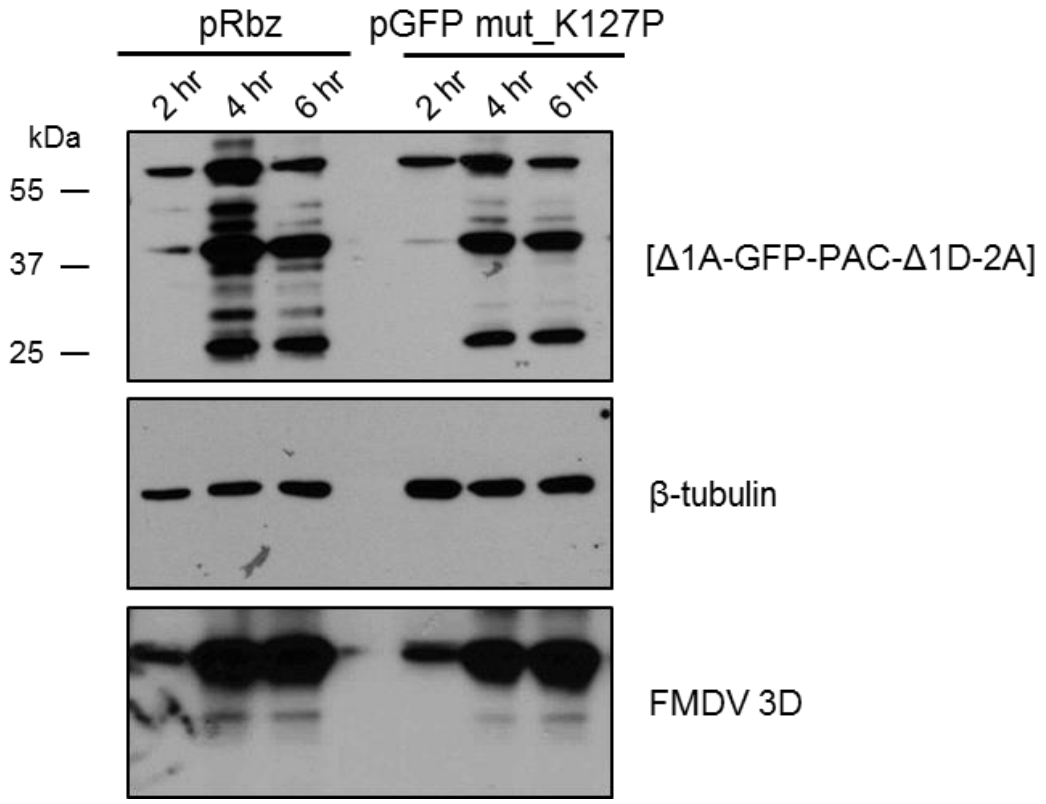


Figure 5.4. Cleavage of *Aequorea* GFP in cells transfected with pRbz and GFP mutant replicons. Extracts were prepared from pRbz or pGFP mut_K127P replicon-transfected BHK-21 cells at the time points indicated. Extracts were separated by 4-20 % SDS-PAGE, transferred to nitrocellulose membranes, and analysed by western blotting with anti-GFP, anti-β-tubulin and anti-3D antibodies. FMDV 3D was used as a measure of viral protein production.

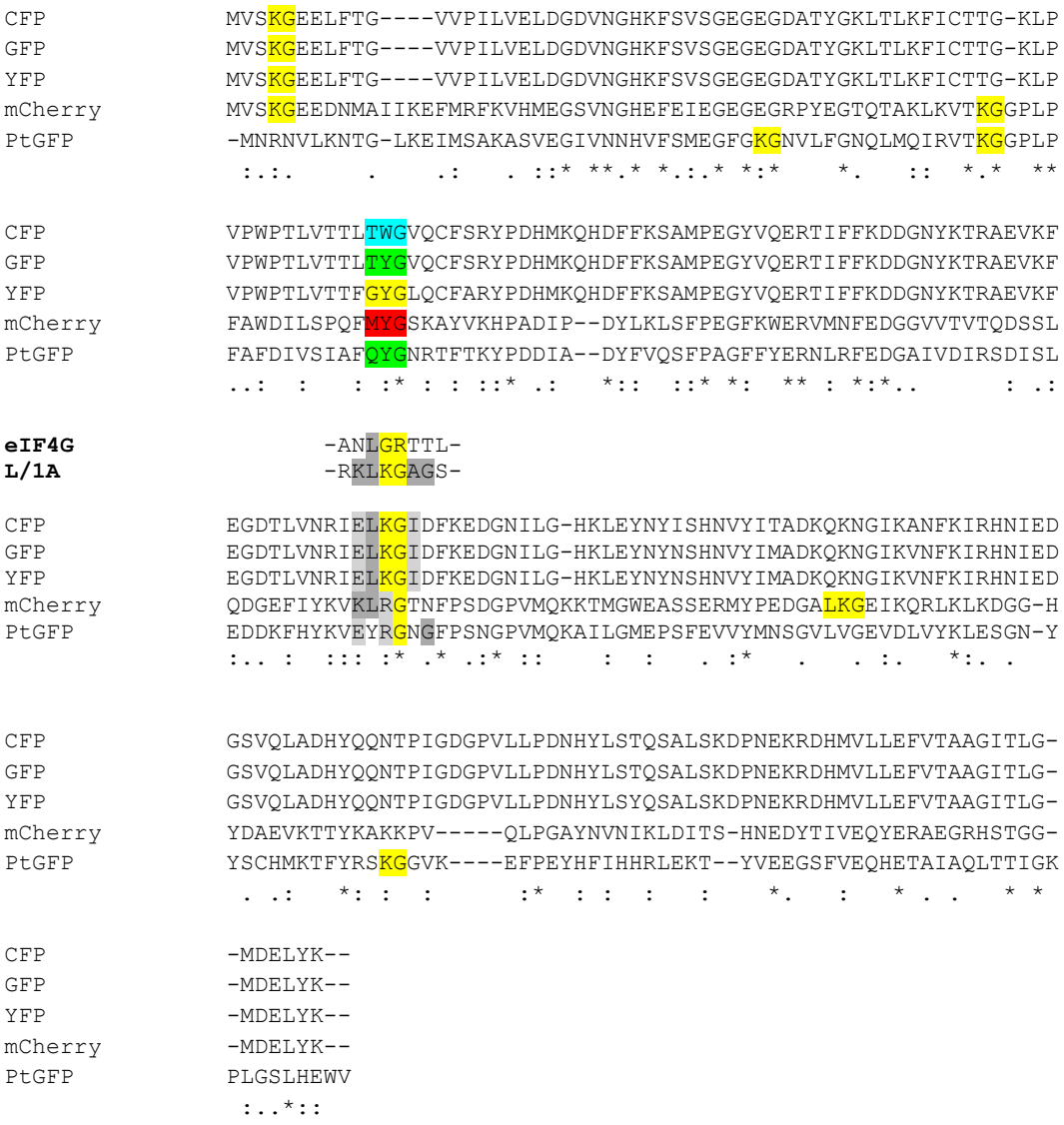


Figure 5.5. Alignment of mCherry and *Ptilosarcus* GFP with *Aequorea* GFP and its derivatives. Residues comprising the chromophore are highlighted with the colour corresponding to each FP. The L^{pro}/1A and eIF4G cleavage sites are shown. The potential L^{pro} cleavage site is annotated with similarity to this site shown in grey or yellow. Lysine-glycine pairs are highlighted throughout all sequences in yellow. Amino acid sequences and alignment generated using DNAMAN.

5.6 *Ptilosarcus* GFP construct design

Pt.GFP was synthesised as a gene block and inserted into the pRbz replicon *via* RE digestion with NsiI and XmaI, creating the pPt.GFP replicon (Figure 5.6, A.3). The pPt.GFP-PAC replicon was generated *via* PCR of the Pt.GFP coding region (primers listed in Table A.1) which was inserted into the pJC3-PAC vector (unpublished, not shown) through RE digestion with BamHI and PstI. The resulting plasmid was digested with NsiI and StuI and the fragment of interest ligated into the pRbz replicon, similarly restricted (A.3). The pGFP replicon was generated previously through PCR of the Aq.GFP sequence, followed by RE digest with EcoRV and XmaI and subsequent ligation into the similarly digested pGFP-PAC replicon (A.2). Replicons lacking PAC were made as a precaution in the event of a C-terminal fusion affecting Pt.GFP fluorescence as previous studies have not used Pt.GFP fusion proteins (Peelle *et al.*, 2001; Schulte *et al.*, 2006).

5.7 *Ptilosarcus* GFP replicon-derived expression

To investigate GFP expression generated by Pt.GFP replicons in comparison to replicons encoding Aq.GFP, plasmids were linearised with AscI or HpaI, respectively, and RNA synthesised as described in previous chapters. BHK-21 cells expressing RFP localised to the nucleus were used during transfections in order to obtain cell numbers for estimation of the transfection efficiency and as an indication of cell death, with the expectation that RFP cell counts from replicon transfected cells would decline over the time period studied. Monolayers were transfected with 1 µg of transcript RNA and fluorescence measured every 2 hr over a 24 hr period using the IncuCyte ZOOM. Examples of microscopy images taken for the duration of the experiment are shown in Figure 5.7. The fluorescent nuclei could be clearly distinguished and changes in cell morphology were more prominent due to all cells containing a fluorescent marker. RFP counts indicated a ~2-3-fold decrease in cell numbers in comparison to mock-transfected cells during the course of the experiment (data not shown), which can be attributed to both cellular death and host cell translational shut-off *via* eIF4G cleavage by L^{pro} (Devaney *et al.*, 1988) leading to the inhibition of cap-dependent RFP synthesis. Cell transfection efficiency was estimated (at 6 hr post-transfection) from GFP and RFP cell counts and was shown to be ~45% for pPt.GFP replicon-transfected cells and ~20% for replicons containing Aq.GFP (data not shown).

GFP counts derived from pPt.GFP replicon-transfected cells were ~10-fold higher on average (Figure 5.8A) and the signal obtained was considerably greater than replicons containing Aq.GFP, with GFP fluorescence ~45-fold higher at 2 hr p.t. and ~30-fold on average thereafter (Figure 5.8B). This increase can in part be attributed to the presence of ribozymes within the pPt.GFP replicons, which were shown to improve replication kinetics in Chapter 4. Comparison of data obtained from Figure 4.3B with data from Figure 5.8B indicates the pPt.GFP replicon is ~20-fold brighter on average. The fluorescent signal arising from Pt.GFP during these experiments is higher than previously reported (Peelle *et al.*, 2001), however this could be largely due to differences in measurement techniques.

The pPt.GFP-PAC and pGFP-PAC replicons displayed a decrease in signal intensity when compared to replicons lacking the PAC coding region, and this suggests the C-terminal fusion affects fluorescence of both GFPs (Figure 5.8). As observed previously within this chapter, the polymerase deletion pRbzΔ3D generated no fluorescence (Figure 5.8).

Due to the high level of GFP fluorescence generated from the pPt.GFP replicon, a number of experiments were undertaken to ensure this increase in brightness was an intrinsic property of this FP and not due to overexpression or increased stability of the protein.

5.8 *Ptilosarcus* GFP structure prediction

Peelle *et al.*, (2001) performed Circular Dichroism (CD) to determine the secondary structure of both *Renilla mulleri* GFP (Rn.GFP) and Pt.GFP. The CD spectra obtained were virtually identical to Aq.GFP and indicated a β-strand conformation with small unstructured chains. The GFPs also exhibited similar stabilities when thermal melting curves were measured (Peelle *et al.*, 2001). The crystal structure of GFP from *Renilla reniformis,* closely related to Renilla mulleri*,* has been solved (Loening *et al.*, 2007) and adopts a β-can conformation highly similar to Aq.GFP. As Rn.GFP and Pt.GFP have comparable fluorescent properties, share a similar amino acid identity to Aq.GFP, and near identical CD spectra, it can be assumed that Pt.GFP also shares a β-can structure. Therefore, the structure of Pt.GFP was predicted using Phyre^2 (Kelley & Sternberg, 2009) to add to this body of evidence, and to eliminate this as a potential reason for the difference in fluorescence generated during replicon studies. Residues 9-231 of the 237 aa sequence were successfully predicted. The PDB generated was

aligned to the Aq.GFP structure (PDB ID: 1EMA) using PyMOL software (Schrodinger, 2010). Superposition of the two GFP structures showed a high degree of structural similarity, with Pt.GFP forming the characteristic β-barrel that is typical of Aq.GFP (Figure 5.9). The C-terminal regions differed in their length, with Pt.GFP having a large unstructured terminus. Several β-strands within Pt.GFP seem to be longer than Aq.GFP; however, apart from the differences mentioned, the proteins appear almost identical.

5.9 *Ptilosarcus* GFP protein expression in replicon-transfected cells

Western blots were carried out on pPt.GFP and pGFP replicon-transfected cells to ensure that protein production was similar between both constructs. BHK-21 monolayers were transfected with 3 µg of transcript RNA and extracts prepared at 2, 4 and 6 hr p.t. (Figure 5.10). Anti-2A and 3D antibodies were used to measure GFP expression and non-structural protein production, respectively. Expression profiles obtained for each construct gave rise to a band correspond to the [Δ1A-GFP-Δ1D-2A] fusion protein, showing no major differences in expression levels between the two replicons and suggested Pt.GFP was not cleaved by L^{pro}. This was also supported by the fluorescent data generated thus far which showed no obvious signs of Pt.GFP cleavage. Plasmid pJC3 (Figure 5.11) was used as a control to indicate a [GFP-2A] fusion protein. A band present within both sets of replicon cell extracts migrated slightly above this GFP-2A product. It is known from our own work on FMDV 2A that the polyclonal antibody raised against the 2A peptide cross-reacts with unknown cellular proteins (unpublished observations; not shown), and it is likely this band has arisen from such non-specific interactions. A Pt.GFP antibody is in production (DCP) and will provide better analysis of Pt.GFP synthesis during the course of replicon experiments.

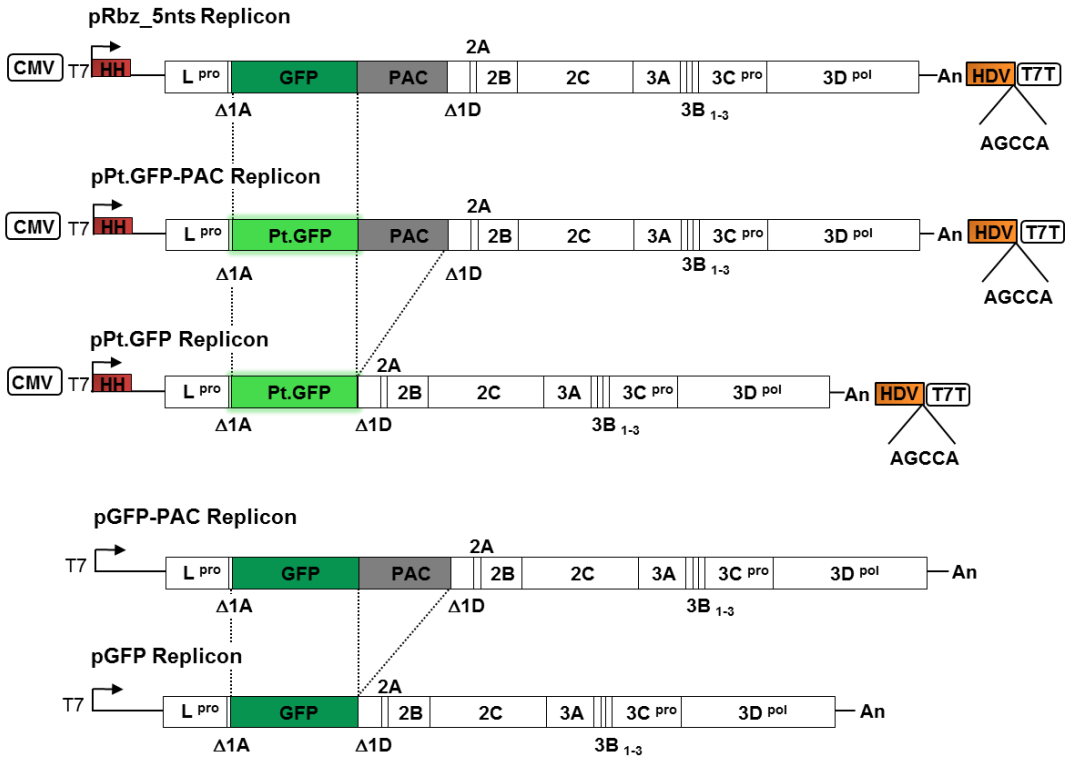


Figure 5.6. *Ptilosarcus* GFP Replicons. Aq.GFP was replaced with the brighter Pt.GFP in the pRbz replicon. A replicon lacking PAC was also created in the event of a C-terminal fusion affecting Pt.GFP fluorescence. A version of the pGFP-PAC replicon without PAC had been created previously, the pGFP replicon, and was used as a comparison to the Pt.GFP replicon.

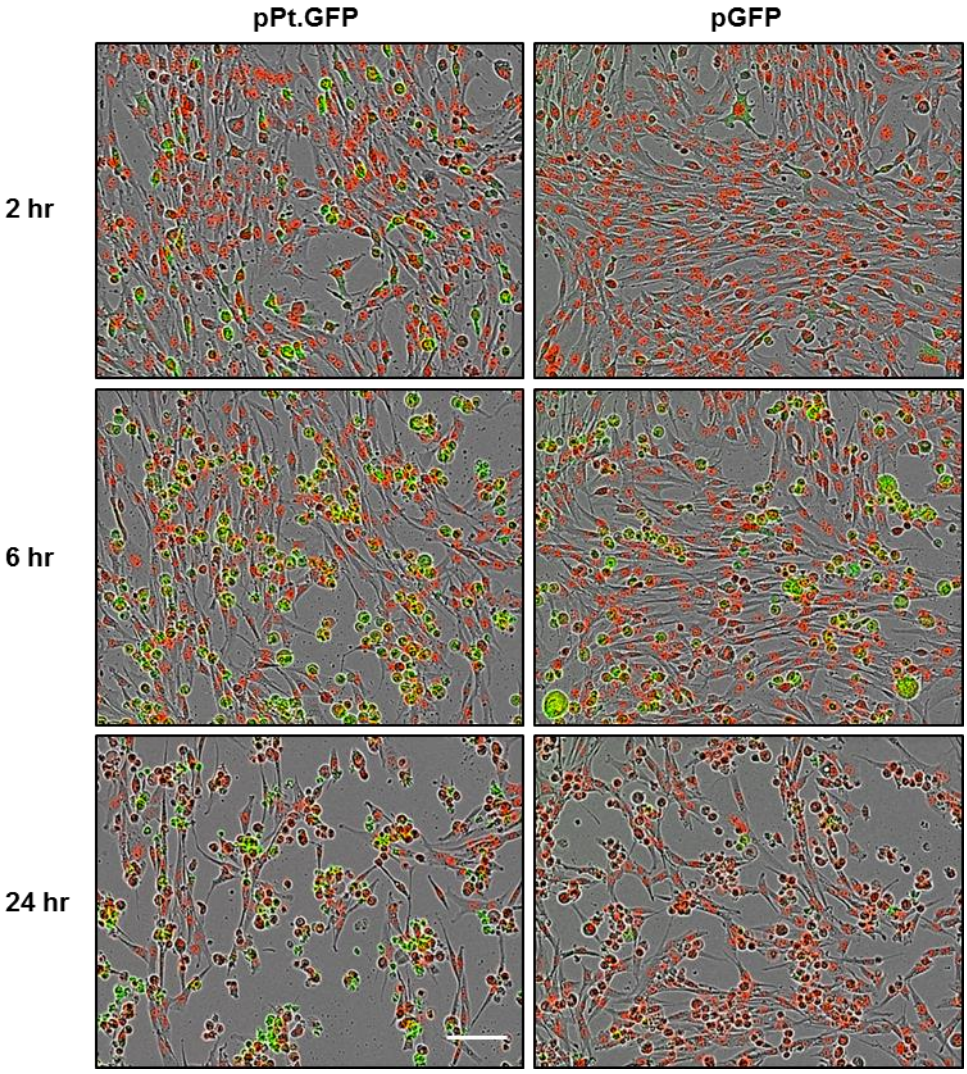


Figure 5.7. *Ptilosarcus* GFP expression in FMDV-transfected nuclear red BHK-21 cells. BHK-21_RFPNuc cell monolayers were transfected with pPt.GFP or pGFP replicon RNA, and fluorescent images captured at 2 hr intervals over a 24 hr period using the IncuCyte ZOOM imaging system. Data represents one image from nine captured at 2, 6 and 24 hr post-transfection, and was obtained from 3 independent transfections, with 4 replicates per transfection. Scale bar represents 100 µm.

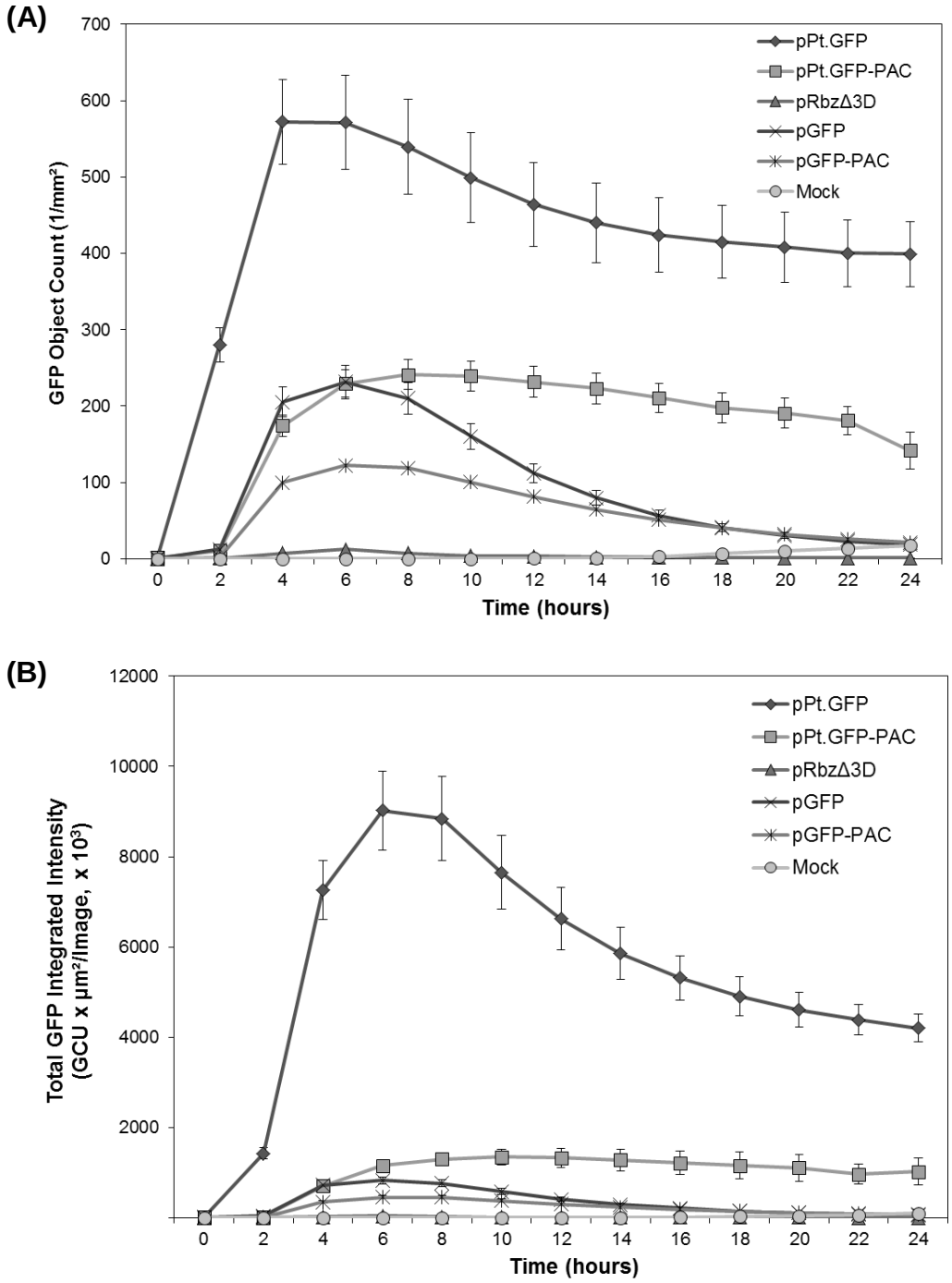


Figure 5.8. Time course of *Ptilosarcus* GFP expression in FMDV-transfected BHK-21-RFPNuc cells. Data from mock-transfected BHK-21 cells is shown, together with cells transfected with transcript RNA derived from the pPt.GFP, pPt.GFP-PAC, pRbzΔ3D, pGFP and pGFP-PAC replicons. At the time points indicated images were captured over a 24 hr period and the GFP fluorescence quantified for each replicon construct: data shown as the green object count/mm^2 (A) or as the total integrated GFP fluorescence intensity (B). Data points/error bars shown are derived from 3 independent transfections, with 4 replicates for each transfection.

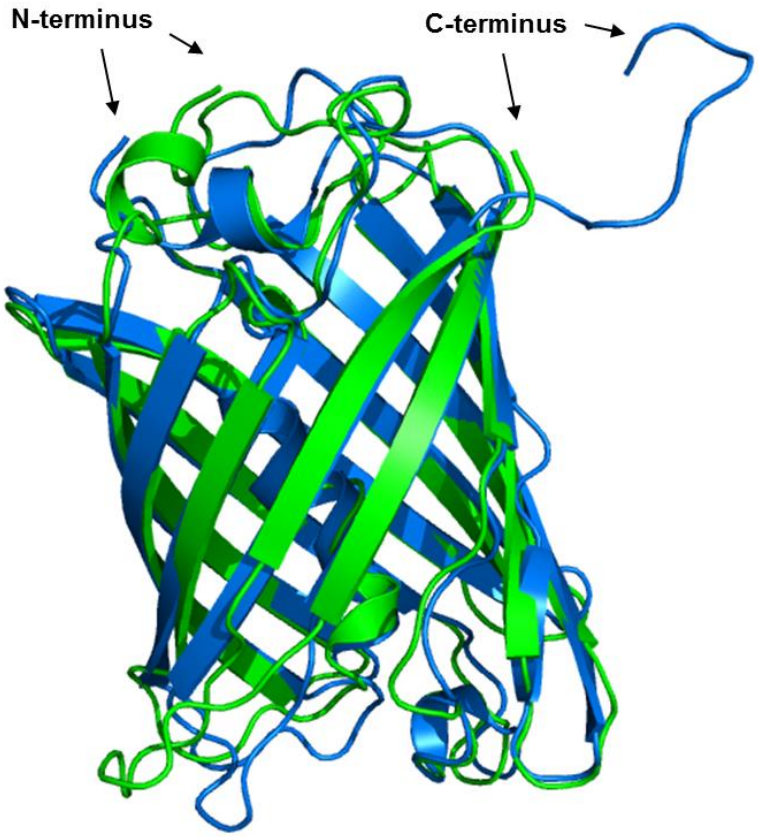


Figure 5.9. Structure of *Aequorea* GFP and predicted structure of *Ptilosarcus* GFP. Aligned cartoon diagram of *Aequorea* (green, PDB ID: 1EMA) and *Ptilosarcus* (blue) GFP. The *Ptilosarcus* GFP structure was predicted using Phyre[2] (Kelley & Sternberg, 2009). N- and C-termini are labelled. Image generated using PyMOL (Schrodinger, 2010).

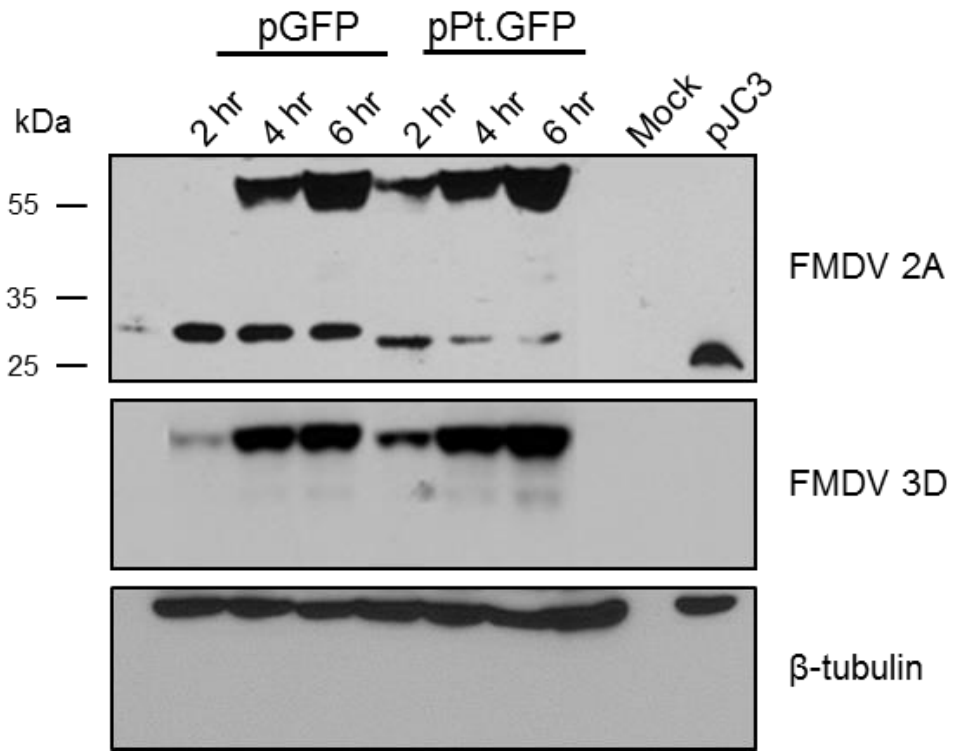


Figure 5.10. *Ptilosarcus* GFP expression in replicon-transfected BHK-21 cells. Extracts from BHK-21 cells transfected with pPt.GFP or pGFP replicon RNA (2-6 hr as indicated) were separated by 12 % SDS-PAGE, transferred to nitrocellulose membranes, and analysed by Western blotting with anti-2A (top panel), anti-3D or β-tubulin antibodies. FMDV 3D expression was used as a measure of non-structural protein production and pJC3 (diluted 1:50) to demonstrate a GFP-2A fusion protein.

5.10 Comparison of GFP half-lives

5.10.1 Construct design

Plasmids containing Pt.GFP and/or flanking sequences from FMDV Δ1A, 1D and 2A were created to test the half-life of Pt.GFP. Primers listed in Table A.1 were used to amplify the desired region from the pPt.GFP replicon and inserted into pJC3 *via* BamHI and ApaI RE sites. This generated a panel of plasmids to study the half-life of Pt.GFP (Figure 5.11). Plasmids containing PAC between the GFP and 2A sequences were also created, pPt.GFP-PAC-JC3 and pJC3-PAC, which encode Pt.GFP and Aq.GFP respectively (not shown).

5.10.2 GFP expression of plasmids created to test GFP half-life

Addition of the FMDV capsid sequences to the Pt.GFP sequence dramatically affected the fluorescent properties of Pt.GFP; pPt.GFPΔ1A/D-JC3 exhibited ~1.5-fold lower fluorescence on average than pPt.GFP-JC3 and ~7-fold lower than pPt.GFPΔ1D-JC3 (Figure 5.12). It was clear that addition of FMDV Δ1D increased Pt.GFP fluorescence, as a ~10-fold increase was observed when comparing the pPt.GFP-JC3 and pPt.GFPΔ1D-JC3 plasmids (Figure 5.12). This is also evident by the ~15-fold difference between pJC3 and pPt.GFP-JC3, which is reduced to a ~3-fold difference when the FMDV Δ1D sequence was added to the C-terminus of Pt.GFP (Figure 5.12). Unexpectedly, pJC3 exhibited a higher signal than all pPt.GFP-JC3-based plasmids, with a ~3-fold increase over the highest expressing Pt.GFP plasmid, pPt.GFPΔ1D-JC3. Plasmids containing PAC produced GFP signals of similar intensity, however while this extension decreased Aq.GFP fluorescence by ~8-fold on average (compare pJC3-PAC with pJC3; Figure 5.12), Pt.GFP expression increased by ~4-fold (compare pPt.GFP-PAC-JC3 with pPt.GFP-JC3; Figure 5.12). This, in combination with data observed for the pPt.GFPΔ1D-JC3, suggests Pt.GFP fluorescence is increased by C-terminal extensions in contrast to hypotheses described in section 5.7 following examination of replicon-derived fluorescence. The fluorescent signal decreased for constructs containing Aq.GFP fused to PAC, supporting data observed with the pGFP-PAC replicon which also produced a decrease in signal intensity in comparison to the pGFP replicon (Figure 5.8). As expected, red fluorescence data showed no major differences between constructs tested (data not shown).

Due to the results obtained pPt. GFPΔ1D-JC3 was used in further experiments to study the half-life of Pt.GFP, as this plasmid generated the best expression level.

5.10.3 GFP expression following cycloheximide treatment

To quantify GFP degradation, BHK-21 cells were transfected with 1 µg of plasmid DNA and fluorescence monitored for 24 hr. Cells were then treated with CHX 24 hr p.t. and fluorescence monitored further: allowing GFP intensity to be measured in the absence of newly synthesised GFP. Figure 5.13 demonstrates the half-life of each GFP is ~36 hr, and, as observed in Figure 5.12 pJC3, expression was higher than pPt.GFPΔ1D-JC3. Green and red counts exhibited trends similar to total intensity data (data not shown). Therefore, the data observed for sections 5.8-5.10 suggests the increased fluorescent properties of Pt.GFP are due to intrinsic factors and the preference for a C-terminal extension, not due to over-expression or increased stability.

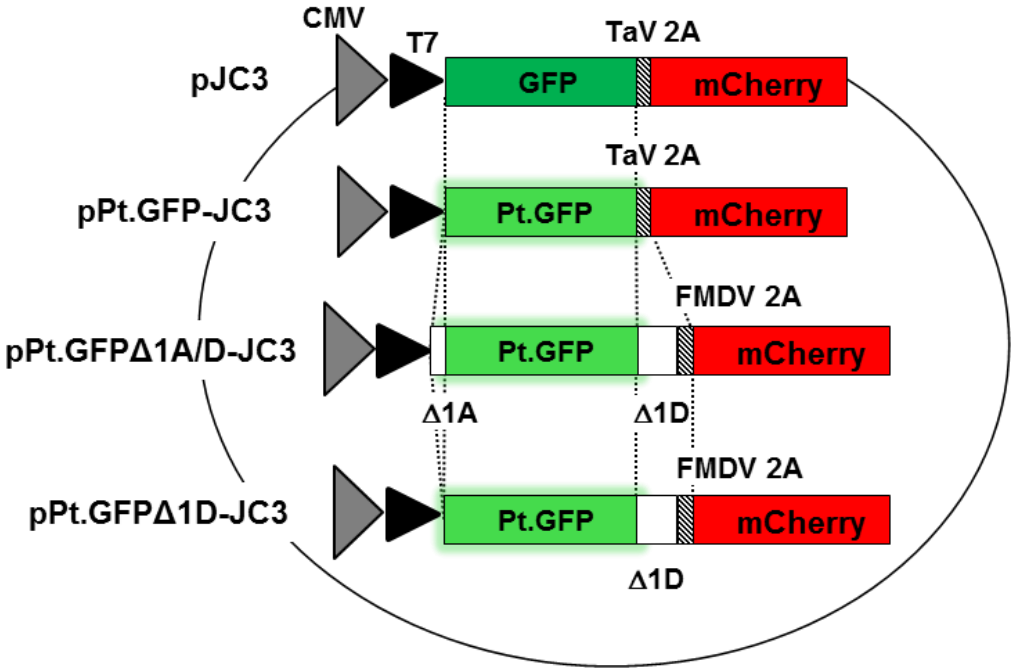


Figure 5.11. Plasmids created to study the half-life of *Ptilosarcus* GFP. The pJC3 plasmid (unpublished) encoding GFP, *Thosea asigna* 2A (TaV 2A; Donnelly *et al.*, 2001), and mCherry, was modified to include Pt.GFP, replacing Aq.GFP. This construct was modified further to include coding regions from the FMDV capsid proteins 1A and 1D and FMDV 2A.

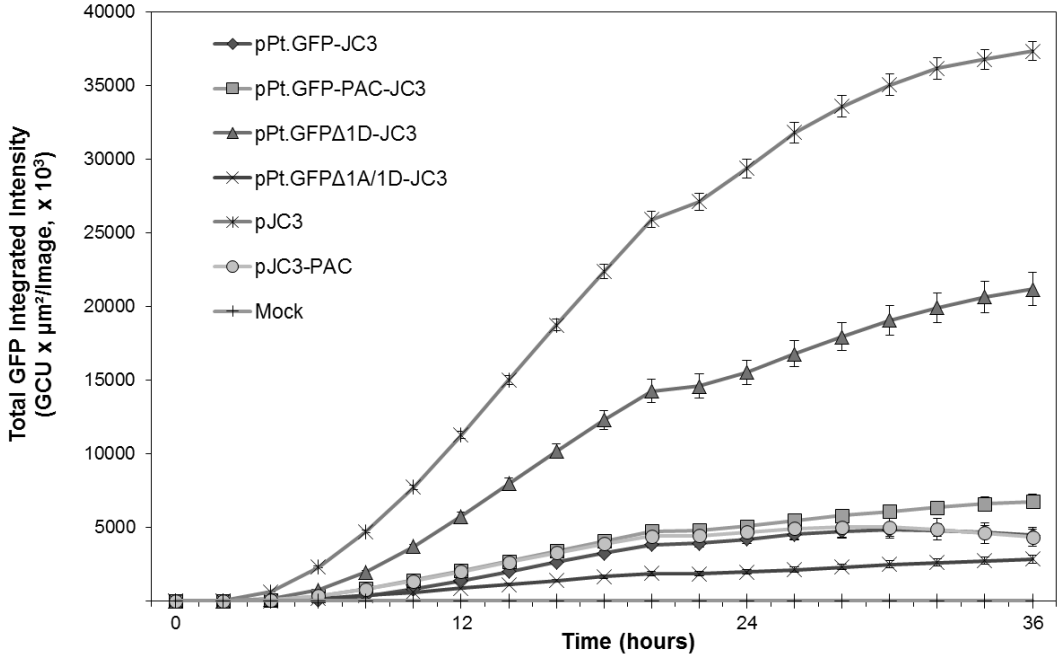


Figure 5.12. GFP expression from plasmids created to compare the fluorescent properties of *Ptilosarcus* or *Aequorea* GFP's. BHK-21 cells were mock-transfected, together with cells transfected with plasmid DNA encoding Pt.GFP, or Aq.GFP; pPt.GFP-JC3, pPt.GFP-PAC-JC3, pPt.GFPΔ1D-JC3, pPt.GFPΔ1A/1D-JC3, pJC3 and pJC3-PAC). At the time points shown images were captured over a 36 hr period and the GFP fluorescence quantified for each construct: data represents the total integrated GFP intensity. Data points/error bars are derived from 2 independent transfections, with 4 replicates for each transfection.

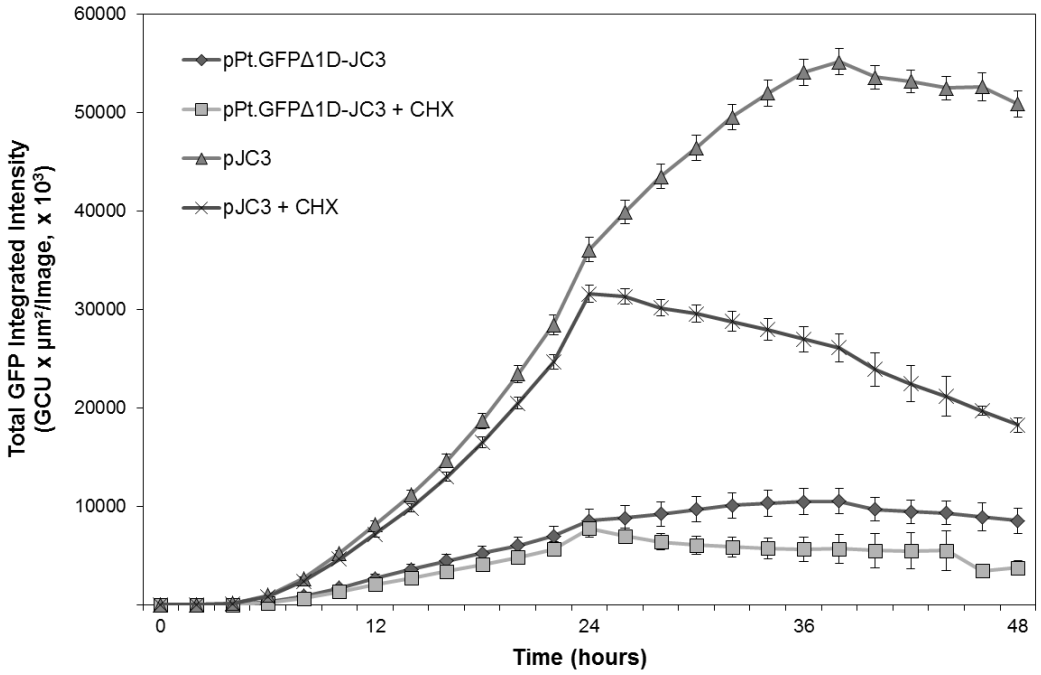


Figure 5.13. Comparison of the half-life of *Ptilosarcus* or *Aequorea* GFP's by measurement of GFP fluorescence following cycloheximide treatment. BHK-21 cells were transfected with plasmid DNA from pPt.GFPΔ1D-JC3, pJC3 or mock-transfected (not shown). Cells were treated with cycloheximide 24 hr post-transfection to determine the protein half-life. At the time points shown images were captured over a 48 hr period and the GFP fluorescence quantified for each construct: data represents the total integrated GFP intensity. Data points/error bars are derived from 2 independent transfections, with 3 replicates for each transfection.

5.11 Construction of the pmCherry replicon

A panel of mCherry replicons were made *via* PCR and cloning. 'Leaderless' versions lacking L or containing the 'spacer' region, as well as a WT and polymerase deletion (not shown) were constructed (Figure 5.14, constructs and all further experiments performed by a senior honours project student under supervision of F.Tulloch). PCR was carried out using primers listed in Table A.1 using the pJC3 plasmid (Figure 5.11) as a template. The pRbz_5nts_mod replicon was generated previously (Dr G. Luke) by deletion of sequences encoding the CMV promoter and neomycin resistance from the pRbz_5nts replicon, and all PCR fragments amplified were inserted into this vector following RE digestion with PsiI or NsiI and XmaI. The pmCherryΔ3D replicon was made by RE digestion of the pmCherry replicon with MluI (not shown).

5.12 pmCherry replicon expression

Plasmids were linearised using AscI and transcript RNA synthesised as described in previous chapters. BHK-21 monolayers were transfected with 1 µg of transcript RNA and fluorescence monitored every hour over a 14 hr period using the IncuCyte imaging system. Fluorescence was detected ~1-2 hr p.t. (Figure 5.15), with the level of mCherry expression ~65-fold lower at this time point when compared to Pt.GFP replicons (Figure 5.8B) and ~15-fold lower than replicons containing Aq.GFP (Chapter 4, Figure 4.3B); however this can be attributed to differences in fluorescent properties between these FPs. The leaderless mCherry replicon displayed fluorescence which was ~20-40-fold lower at 2-3 hr p.t. in comparison to the WT pmCherry replicon. This is in contrast to the signal obtained from the pLL-GFP-PAC replicon which produced fluorescence ~2-fold higher than pGFP-PAC (Figure 3.4B) and suggests mCherry FP is not cleaved by L^{pro}. Western blots on cell extracts with anti-RFP antibodies will confirm if mCherry is indeed intact. Data generated with the pLa-mCherry replicon, which contains the 'spacer' region of L^{pro} (described in Introduction 1.1.2.6), was in agreement with the pLL-mCherry data and suggested this FP was not cleaved. Results obtained from both the 'leaderless' replicons demonstrates that replicons lacking L are attenuated, which is in some agreement with previous observations (Piccone *et al.*, 1995), and exhibits the true potential of this system to measure differences between replicating and attenuated forms. This is particularly important for future work which may involve

investigation of L^{pro} and would require a suitable screen for measuring differences in fluorescence. No fluorescence was detected from cells transfected with pmCherryΔ3D transcript RNA, which is in agreement with data obtained above with the pRbzΔ3D replicon (Figure 5.3, 5.8B). RFP counts exhibited similar trends as RFP fluorescence data (not shown).

5.13 Conclusions

- Mutagenesis of the potential L^{pro} cleavage site within Aq.GFP resulted in a loss of fluorescence. However, it does seem that the amount of GFP cleaved was reduced when cell extracts were probed with anti-GFP antibodies.

- Due to the unsuccessful attempt to mutate the putative cleavage site, other FP replacements were investigated. Alignment of Aq.GFP with both Pt.GFP and mCherry FPs showed low sequence homology and were explored as potential substitutes. Pt.GFP replicons showed a considerable increase in expression, with a ~20-fold increase in overall expression when compared to Aq.GFP. Experiments undertaken to elucidate whether this increase was due to overexpression or increased stability, demonstrated that Pt.GFP was expressed to similar levels as Aq.GFP, with the half-lives of these GFPs being almost identical.

- Replicons encoding mCherry did not seem to undergo cleavage by L^{pro}, and attenuated forms could be more easily distinguished, unlike identical forms investigated previously (Chapter 3) encoding Aq.GFP.

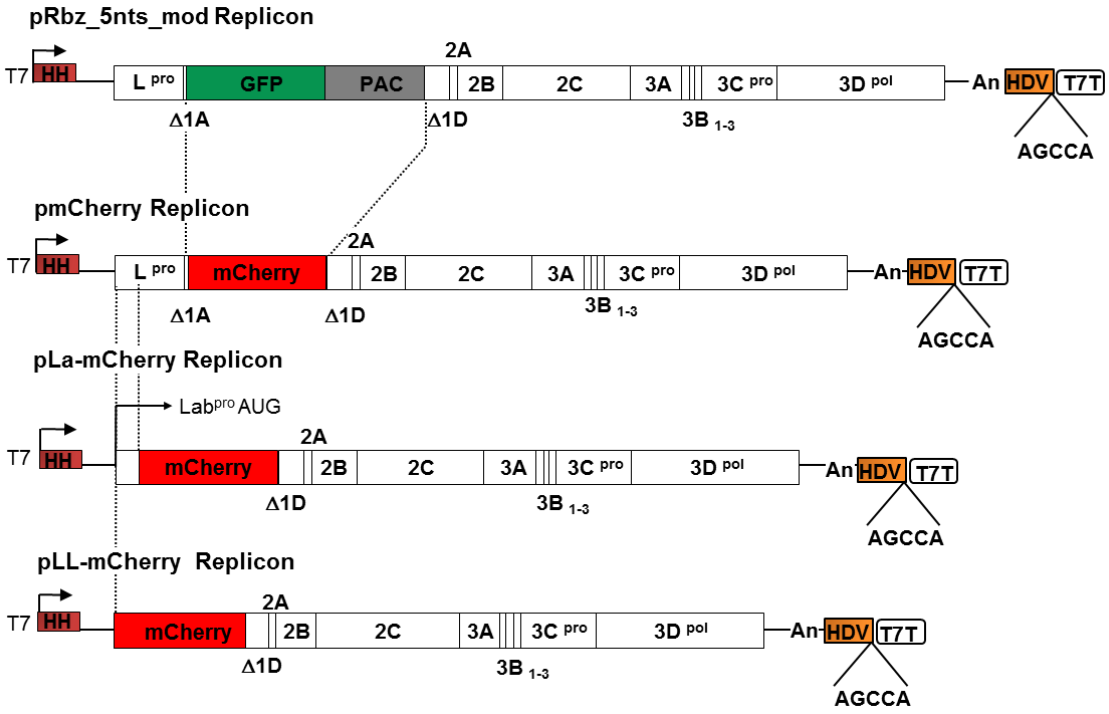


Figure 5.14. mCherry Replicons. The pRbz_5nts_mod replicon was adapted by replacing *Aequorea* GFP with mCherry FP. 'Leaderless' versions lacking either L or the Lb coding region; pLL-mCherry and pLa-mCherry, respectively, and a polymerase deletion pmCherryΔ3D (not shown) were also created.

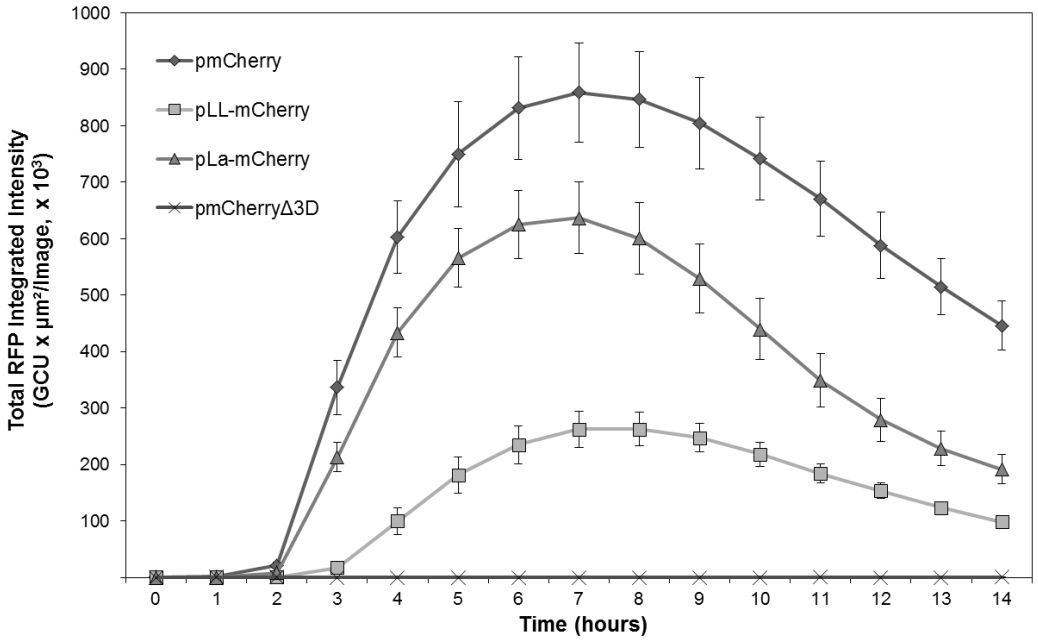


Figure 5.15. Time course of mCherry expression in FMDV replicon-transfected BHK-21 cells. Cells were mock transfected (not shown) or transfected with transcript RNA derived from the pmCherry, pLLmCherry, pLa-mCherry and pmCherryΔ3D replicons. At the time points indicated images were captured over a 14 hr period and the RFP fluorescence quantified for each replicon construct: data shown as the total RFP integrated intensity. The data/error bars shown are derived from 2 independent transfections, with 3 replicates for each transfection.

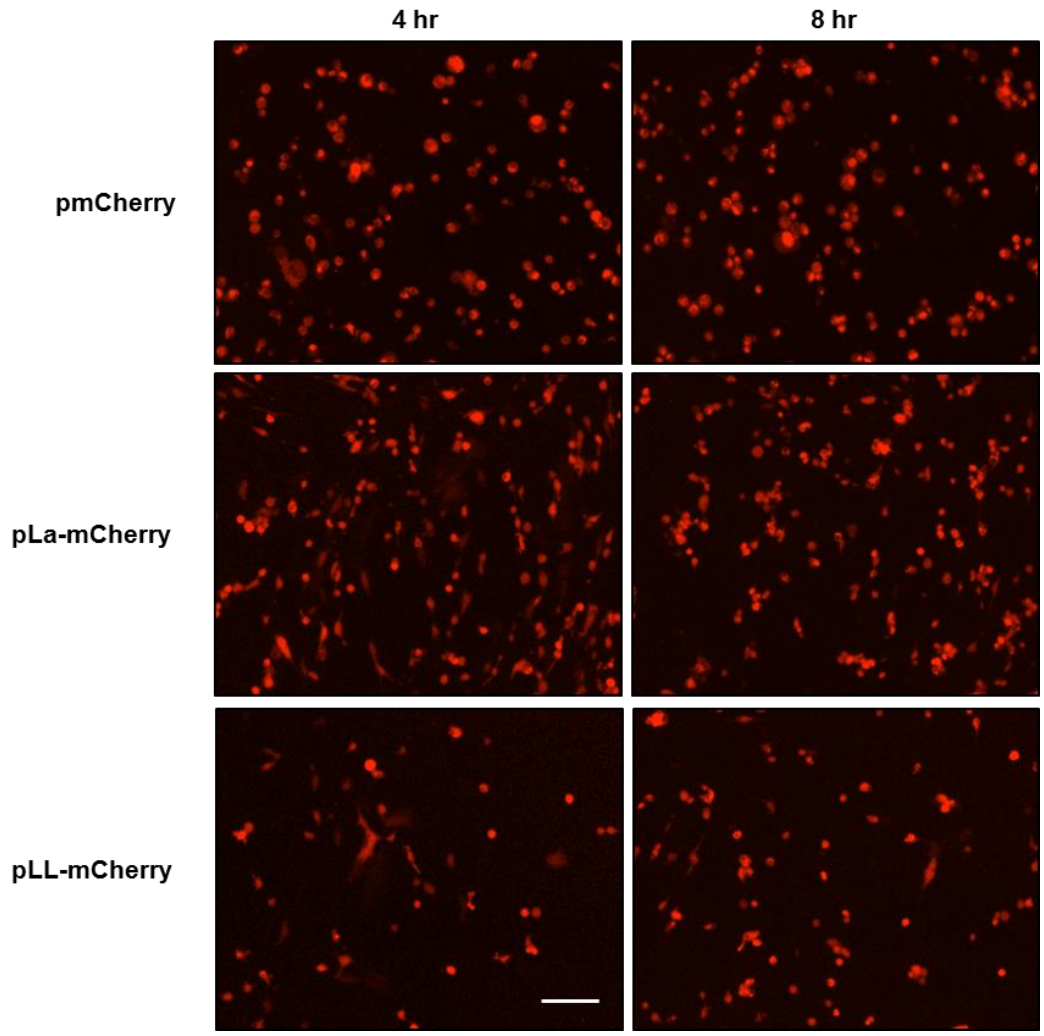


Figure 5.16. mCherry expression in FMDV replicon-transfected BHK-21 cells. Transcript RNA from the pmCherry, pLa-mCherry and pLa-mCherry replicons was introduced into cell monolayers, and fluorescent images (nine/well) captured at 1 hr intervals over a 24 hr period using the IncuCyte ZOOM imaging system. Data is representative of a captured image at 4 and 8 hr post-transfection, and was obtained from 2 independent transfections, with 3 replicates per transfection. Scale bar represents 100 μm.

5.14 Discussion

5.14.1 Mutagenesis of *Aequorea* GFP

Mutagenesis of Lys127 within β-sheet 6 of Aq.GFP abolished fluorescence. No GFP fluorescence was detected from replicons containing mutated Lys127Pro or Lys127His. Western blots performed on replicon-transfected cell extracts with antibodies raised against FMDV non-structural proteins 2A and $3D^{pol}$ confirmed active replication and reduction in the amount of GFP cleaved, however cleavage activity was not completely eliminated. FMDV $3C^{pro}$ cleaves predominantly at Glu-Gly (-EG-) residue pairs (Ryan *et al.*, 1989), and this could account for the remaining cleavage products observed following mutagenesis of the putative L^{pro} cleavage site within Aq.GFP, due to the presence of -EG- pairs within this protein (see Figure 5.5). Deletion analysis of Aq.GFP defined the minimal domain required for fluorescence to be residues 7-229, with only 15 residues able to be deleted without affecting fluorescence (Li *et al.*, 1997). The 7-229 region spans from the α-helix within the N-terminus to the end of the last β-strand of GFP. Deletions or substitutions to the N-terminus, or internally, eliminate fluorescence, whereas C-terminal changes are more tolerable (Li *et al.*, 1997). GFP comprises 11 anti-parallel β-sheets surrounding a central α-helix that is attached to the chromophore consisting of Ser65, Tyr66, and Gly67 (Ormö *et al.*, 1996; Yang *et al.*, 1996). This structure forms a compact 'barrel' or 'can' that ensures protection of the fluorophore from quenching and creates a unique environment for the production of fluorescence (Figure 5.9). Any modifications made to GFP which affect proper folding and structure formation will reduce the fluorescent signal, as will mutations which allow contact with the surrounding environment.

It was hoped that mutagenesis of Lys127 to another positively charged amino acid would lead to retention of activity; however the Lys127His mutant also eliminated fluorescence. The larger, bulky side-chain of Histidine could be affecting important contacts with other residues of GFP, leading to loss of fluorescence. An arginine mutant was not obtained during mutagenesis experiments- perhaps this may have retained some fluorescent activity due to the higher degree of similarity with lysine; however due to the critical residues of GFP being located within region 7-229 of GFP (Li *et al.*, 1997), this may not be achievable regardless of the amino acid substituted.

5.14.2 Alternative reporter proteins

Unsuccessful attempts to mutate Aq.GFP led to the investigation of appropriate substitutions. FPs which arose from different species than the jellyfish GFP would be expected to be more suitable due to their low sequence homology. Therefore *Ptilosarcus* GFP and mCherry from the sea pen and mushroom coral (Peelle *et al.*, 2001; Shaner *et al.*, 2004), respectively, were chosen for subsequent replicon studies.

5.14.2.1 *Ptilosarcus* GFP replicon expression

Replicons encoding Pt.GFP exhibited enhanced fluorescence in comparison to those encoding Aq.GFP. Both GFP count and fluorescence intensity data was substantially higher than data obtained with replicons expressing Aq.GFP. This increase in fluorescence can be attributed to the presence of ribozymes within the Pt.GFP replicons which accounts for ~10-fold of the ~30-fold increase observed. Since further experiments investigating the protein expression profile and stability of Pt.GFP showed these attributes to be comparable to Aq.GFP (discussed below), this suggests the enhanced signal is due to intrinsic characteristics of Pt.GFP.

The best figures of merit to express the 'brightness' of a FP are the molar extinction coefficient (EC) and the quantum yield (QY). The EC is a measurement that determines the efficiency of light absorption by the FP and the QY determines the efficiency of the fluorescence process (number of photons emitted in relation to the number absorbed; Chudakov *et al.*, 2010). It is possible that the enhanced fluorescence could be due to Pt.GFP possessing a higher QY and EC in comparison to Aq.GFP; however these measurements have not been calculated for Pt.GFP and are as yet unknown. Nevertheless, these measurements aren't the only factors that should be taken into account as transcription and translation efficiency, mRNA and protein stability, and chromophore formation can all affect the resulting signal produced by a FP (Chudakov *et al.*, 2010). It should also be noted that Schulte *et al.*, (2006) reported the formation of tetramers for Pt.GFP and it is possible this could lead to increased fluorescence. It is known that oligomeric FPs often exhibit enhanced spectral and biochemical properties in comparison to monomeric FPs (Chudakov *et al.*, 2010).

The pH stability of a FP can also affect the fluorescent signal obtained. Schulte *et al.*, (2006) reported that Pt.GFP was much more robust at low pH when compared to pH-sensitive variants of Aq.GFP (pHluorins). Cells transfected with replicon RNA will undergo cell death due to the process of viral replication and cell protein synthesis shut-off. The pH of the cells will therefore decrease and become acidic (Lagadic-Gossmann *et al.*, 2004), which can quench the chromophore leading to a loss in fluorescence. Aq.GFP is quenched by 50% at pH 5.5 (Tsien, 1998). Apoptotic cells reach pH levels near to (and lower) than this (Lagadic-Gossmann *et al.*, 2004). Schulte and colleagues reported a pH range of 3.8-8.2 for Pt.GFP which was broader than the range of pH 4.8-7.6 for Aq.GFP, thus the robust pH stability of Pt.GFP could attribute to the higher signal produced by this FP.

5.14.2.1.1 *Ptilosarcus* GFP structure prediction

Experiments performed by Peelle *et al.*, (2001) to determine the structure of Pt.GFP using CD indicated a β-can formation almost identical to Aq.GFP. Phyre^2 structure prediction (Kelley & Sternberg, 2009) carried out during this study is in agreement with this finding and eliminates the possibility that structural differences account for the variation in fluorescent signal generated by Pt.GFP and Aq.GFP.

5.14.2.1.2 *Ptilosarcus* GFP half-life

GFP expression was monitored in cells treated with CHX to measure the rate of GFP degradation. Fluorescence data generated indicates that both GFPs have similar half-lives of ~36 hr. Aq.GFP has been reported to have a half-life of 26-30 hr based on previous studies (Corish & Tyler-Smith, 1999; Li *et al.*, 1998). The differences between studies could be due to the measurement techniques (image analyses of radiolabelled protein, western blotting and flow cytometry versus live cell imaging) , cell lines used (human cell lines versus rodent cells) the number of time points taken and the duration of the experiments throughout (3 and 12 hr versus 24 hr). The half-life measured during fluorescence experiments correlates well with the N-end rule, where the terminal amino acid determines the half-life of a protein; proteins with a Met N-terminal residue have half-lives of ~30 hr (Gonda *et al.*, 1989), which is the resulting terminal amino acid of these FPs.

The pPt.GFPΔ1D-JC3 plasmid exhibited a higher fluorescent signal than pPt.GFP-JC3. Peelle *et al.*, (2001) investigated the effect of the addition of a 14 amino acid flag-tag to the C-terminus of Pt.GFP upon fluorescence. They observed a 1.4-fold increase in Pt.GFP expression, which explains observations reported in this study where insertion of the FMDV Δ1D capsid sequence (~ 40 aa) enhanced fluorescence ~ 10-fold.

The addition of ~40 aa of Δ1D increased fluorescence 10-fold whereas a combination of both Δ1A and Δ1D caused a 1.5-fold reduction. It is interesting to note that the N-terminal residue of both Aq.GFP and Pt.GFP will be the terminal glycine from Δ1A (produced by L^{pro} cleavage). The N-terminal region of picornavirus 1A proteins comprises a myristolation signal: it may well be the case that this [Δ1A-Pt.GFP-Δ1D] fusion protein also becomes myristolated and this could affect the half-life or fluorescent properties of Pt.GFP.

PAC C-terminal fusions also increased Pt.GFP fluorescence but, interestingly, not to the same extent as the much shorter Δ1D C-terminal extension. It appears, therefore, that different C-terminal extensions have different affects upon the fluorescent properties of Pt.GFP.

As a general observation, cells expressing plasmid constructs encoding PAC displayed a rounded-up morphology in comparison to cells transfected with constructs lacking the PAC gene (data not shown). Acetyl-coenzyme A (acetyl-CoA) is very important for cellular metabolism (Vara *et al.*, 1985). Since acetyl-CoA is used during the biosynthesis of PAC, it's depletion may produce cytopathic effects.

Plasmid pJC3 displayed fluorescence ~3-fold higher than pPt.GFP Δ1D-JC3. This was unexpected due to the large differences observed between replicons expressing Aq.GFP or Pt.GFP. The main differences between experiments carried out with replicon RNA *versus* plasmid DNA is cytoplasmic replication *versus* nuclear transcription. It is known that Aq.GFP expression was limited when used within plants due to the presence of a cryptic splice site. Expression was restored following alterations to the codon usage within Aq.GFP (Haseloff *et al.*, 1997). Hence, it is possible such a site exists within Pt.GFP and this could explain the observations during these experiments.

5.14.2.2 mCherry replicon systems

Initial fluorescence from mCherry replicons was reduced in comparison to Aq.GFP and Pt.GFP replicons. The QY of mCherry is 0.22 in contrast to 0.60 of Aq.GFP with the relative brightness of mCherry 47% that of Aq.GFP (Chudakov *et al.*, 2010; Shaner *et al.*, 2004); therefore it is unsurprising mCherry fluorescence was ~15-fold lower than Aq.GFP at 2 hr p.t. and ~2-fold overall.
Differences between attenuated forms can clearly be detected by the use of mCherry FP. 'Leaderless' replicons displayed reduced fluorescence with the pLL-mCherry replicon also having a 2-3 hr lag in replication. Replicons lacking L and encoding Aq.GFP failed to show such a decrease, which was due to cleavage of this FP by L^{pro} (Chapter 3; Tulloch *et al.*, 2014a). Replacement with a FP that is not cleaved by the viral protease demonstrates the true potential of this replicon system. Indeed, Pt.GFP replicons of a similar nature have been constructed (undertaken by a senior honours project student) and display identical trends, with 'leaderless' replicons showing attenuation (data not shown). This is further evidence that this system can be used to measure attenuated phenotypes and allows us to use 'leaderless' replicons within our panel of 'control' attenuated genomes as a benchmark for attenuation. The use of FPs with different spectral properties will also be very useful for competition and trans-complementation studies using replicons.

5.14.3 GFP expression derived from replicon forms encoding polymerase deletions, but encoding ribozymes

Interestingly, and rather unexpectedly, replicons encoding polymerase deletions containing all FPs tested throughout Chapter 5 (Aq.GFP, pRbzΔ3D; mCherry, pmCherryΔ3D; Pt.GFP, pPt.GFPΔ3D) generated no fluorescence when transfected into BHK-21 cells. This is in stark contrast to the pGFP-PAC-Δ3D replicon which produces a signal ~3-fold less than the WT pGFP-PAC (Tulloch *et al.*, 2014a). Chapter 4 described the importance of a precise 5'-end to replicon-derived fluorescence, with an increase in fluorescent signal when replicons encoded a self-cleaving ribozyme. This has also been observed with PV replicons encoding luciferase (Herold & Andino, 2000). These observations suggest that transcript RNA which cannot replicate (due to a large deletion within the polymerase) and does not contain non-viral guanosine residues at the 5' terminus, is sequestered into abortive replication complexes and is not translated.

When non-viral guanosines are present, transcript RNA is predominantly translated - as observed for the pGFP-PAC-Δ3D replicon which exhibits a replication curve identical to WT pGFP-PAC, but at a reduced level of fluorescence (Tulloch *et al.*, 2014a). These data pose interesting questions with regards to the initiation of FMDV replication where the mechanisms are still poorly understood.

One of the main advantages of the system described is the ability to study areas of the viral life cycle. Regulation of the primary cleavage between 2A and 2B of the FMDV polyprotein has not been investigated in detail. The next chapter describes the generation of replicons to examine this process.

Chapter 6: Replicative Fitness: The Nature of the 'Primary' Separation between Capsid and Replication Proteins.

6.1 Introduction

The 2A region of picornaviruses is highly variable. Within many genera, such as the aphthoviruses, 2A is an oligopeptide which mediates a novel translational effect- "ribosome skipping". However, within the entero- and sapeloviruses, a 2A proteinase ($2A^{pro}$) is found (Martínez-Salas & Ryan, 2010). The $2A^{pro}$ from these genera mediates a primary cleavage at a tyrosine-glycine pair of the junction between the P1 capsid and the N-terminus of $2A^{pro}$. In aphthoviruses, a secondary cleavage by $3C^{pro}$ at the C-terminus results in full release of $2A^{pro}$ from the [P1-2A] polyprotein processing product (Ryan & Flint, 1997). In common with L^{pro} from FMDV, enterovirus $2A^{pro}$ also cleaves eIF4G inactivating host cell cap-dependent protein synthesis (Glaser & Skern, 2000).

The 18 aa 2A peptide from FMDV is responsible for the primary cleavage event which separates the downstream region encoding proteins involved with replication, from the upstream region which comprises the capsid proteins. This is mediated by "ribosome skipping", where the synthesis of a specific peptide bond is "skipped" when the ribosome encounters the motif DxExNPG$^{\downarrow}$P within 2A (Donnelly *et al.*, 2001b). The cleavage occurs between the C-terminal glycine of 2A and the proline of the downstream protein 2B. The 2A peptide remains attached to the C-terminus of the upstream protein, with the downstream protein containing an N-terminal proline residue (Ryan & Drew, 1994; Ryan *et al.*, 1991). In this model, translation either terminates at the C-terminus of 2A, or translation reinitiates and the downstream region is synthesised. A molar excess of the proteins coded upstream of 2A is also observed (Donnelly *et al.*, 2001b).

It is thought that FMDV 2A mediates translational control of the FMDV polyprotein under increasing cell stress. Here, the model states that increasing cell stress (through infection) leads to increased levels of phosphorylation of eEF2 (eukaryotic elongation factor 2) – decreasing its activity. Ribosome processivity rates would decline throughout infection. A key step in the model of FMDV 2A activity is re-association of prolyl-tRNA into the A site of the ribosome following

egress of eRF1 (Doronina *et al.*, 2008) However, to continue polypeptide chain elongation, prolyl-tRNA must be transclocated from the A-to P-site – mediated by eEF2. It is proposed that termination is in competition with this process of translocation: if the rate of translocation falls too low then termination at the C-terminus of 2A becomes predominant. This model predicts that as infection progresses, an ever-increasing proportion of aminoacyl-tRNAs are used to make capsid proteins – increasing the yield of particles. Alteration of the 'balance' of protein synthesis (capsid *versus* replication proteins) could result in a decrease in FMDV replication - *sensu* particle yield. Thus, insertion of a $2A^{pro}$ could change the molar excess of upstream protein synthesis to an equal stoichiometry of all virus proteins – a more uniform translation profile. Bovine enterovirus $2A^{pro}$ was chosen to compliment the replacement of FMDV 2A with a species-specific (bovine) protein.

This section describes the design and construction of novel FMDV replicons where BEV $2A^{pro}$ has replaced FMDV 2A within the P2 coding region. These replicons will be used to investigate the regulation of FMDV 2A during replication, particularly during cell stress, with protein ratios and cellular factors such as eEF2 and phosphorylated eEF2 monitored throughout to study their role within this mechanism.

The insertion of BEV $2A^{pro}$ has been investigated in replicons containing the 'spacer' region of L^{pro} (lacking the Lb coding region). If full replicative fitness can be restored to that of WT, then it can be postulated that the main function of L^{pro} activity is cleavage of eIF4G; a function common to enterovirus $2A^{pro}$. Otherwise, this could suggest other roles for L^{pro} during FMDV replication which will be studied further.

6.2 Construct design

All replicons generated were made by firstly inserting a StuI-EcoRI RE fragment from the pGFP-PAC replicon into the commercially available pSP72 vector (Promega). The resulting Stu-Eco subclone was restricted with XmaI and ApaI and the BEV 2A region (synthesised as a gene block; Dundee Cell Products) inserted following similar RE digestion. This second subclone was then restricted with XmaI and EcoRI and the resulting fragment ligated into the pRbz_5nts_mod replicon (described in Chapter 5.11), similarly restricted. Finally, L^{pro} was removed from the resulting replicon by RE digestion with KpnI and XmaI and replacement of this region with that of the pLa-GFP-PAC replicon, similarly restricted. This generated the pBEV 2A replicon which contained the 'spacer' region of L^{pro}, self-cleaving ribozymes and BEV $2A^{pro}$ in place of FMDV 2A (Figure 6.1).

Initial experiments demonstrated that the pBEV 2A replicon did not replicate, with no fluorescence detected upon transfection of transcript RNA into BHK-21 cells. However, these results were not completely unexpected, as the pBEV 2A replicon contained a novel $3C^{pro}$ cleavage site at the BEV $2A^{pro}$/ FMDV 2B junction. FMDV $3C^{pro}$ does not normally cleave at this site as FMDV 2A cleaves at its own C-terminus (Ryan *et al.*, 1991), therefore it was assumed that BEV $2A^{pro}$ remained fused to FMDV 2B, hindering replication. *In vitro* transcription/translation (TnT; Promega) experiments implied this was the case due to the presence of a ~40 kDa band in the lane corresponding to the pBEV 2A replicon (Figure 6.2). This band was not present in lanes containing pGFP-PAC and pLa-GFP-PAC, with the size of this band correlating to a [BEV $2A^{pro}$-FMDV 2B] fusion protein (Figure 6.2). It was also evident that the translation profile of the pBEV 2A replicon varied in comparison to the other replicons tested. The ratio of upstream to downstream proteins obtained from the pGFP-PAC and pLa-GFP-PAC replicons contained more upstream ("capsid" i.e. Δ1A-GFP-PAC-Δ1D) proteins/protein precursors; consistent with the translational model for replicons encoding a 2A peptide (Donnelly *et al.*, 2001b). In contrast, a strong band corresponding to [Δ1A-GFP-PAC-Δ1D] was not observed for the pBEV 2A replicon (Figure 6.2), which correlates with the initial GFP fluorescence data (not shown).

The presence of important FMDV proteins such as 2C, the P3 precursor and $3D^{pol}$ within translation reactions programmed with the pBEV 2A replicon suggests that translation is able to occur as usual, but incomplete processing is rendering important proteins such as 2B from performing their critical role during replication: hence the lack of GFP fluorescence observed following introduction of this replicon into BHK cells. This lack of replication could have been due to; i) incomplete processing by BEV $2A^{pro}$ at its N-terminus, ii) incomplete processing by $3C^{pro}$ at the novel BEV $2A^{pro}$/ FMDV 2B junction, or iii) a combination of both these possibilities. Therefore, further constructs were produced in order to address these issues. Firstly, 10 aa of BEV 1D was introduced upstream of BEV $2A^{pro}$ to improve cleavage activity. To examine whether incomplete processing at the C-terminus of BEV $2A^{pro}$ was the root cause of reduced replication *Thosea asigna* virus 2A (TaV 2A), which also encodes a ribosomal 'skipping' peptide (Donnelly *et al.*, 2001b), was inserted immediately after BEV $2A^{pro}$ to induce 'cleavage' at the BEV $2A^{pro}$/ FMDV 2B junction, thus removing the necessity for $3C^{pro}$-mediated processing at the BEV $2A^{pro}$/ FMDV 2B site. This modification was carried out with the aim of replacing TaV with differing $3C^{pro}$ dipeptides if it was shown that improved processing at this region increased replication. Finally, a replicon encoding both the 10 aa of BEV 1D and TaV 2A was generated. All replicons were produced through amplification of the desired region using primers listed in Table A.1 and inserted into the pRbz_5nts_mod replicon as described above. This created the pBEV Δ1D-2A, pBEV-TaV 2A and pBEV Δ1D-TaV 2A replicons, respectively (assembled by Dr E. Minskaia; Figure 6.1).

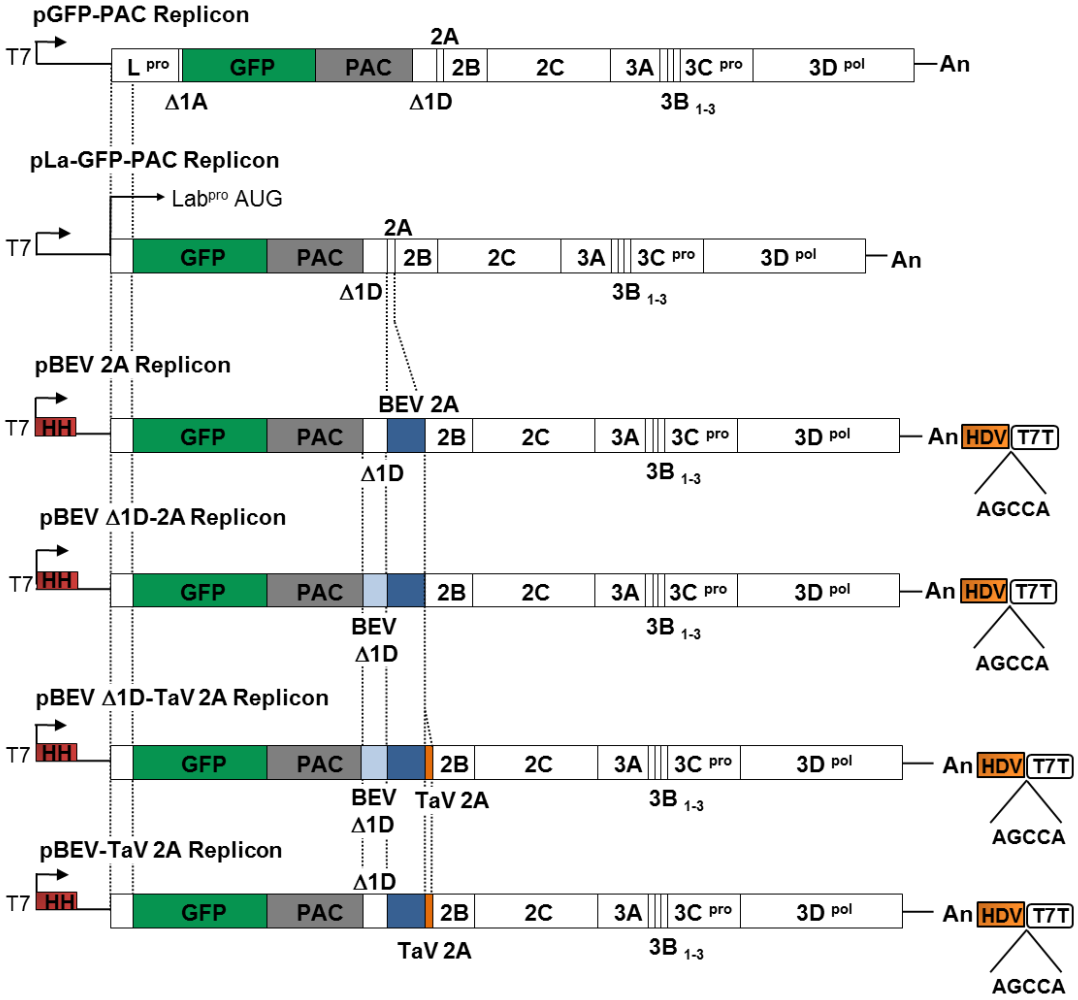


Figure 6.1. BEV 2A Replicons. The pGFP-PAC replicon was modified to create the pLa-GFP-PAC, which contains the coding area between each of the AUG initiation codons in L^{pro}: the 'spacer' region. This was used to create replicons containing BEV $2A^{pro}$ in the improved replicon backbone (pRbz_5nts_mod replicon), which comprises self-cleaving ribozymes. Replicons containing either a small region of the BEV 1D capsid protein, and/or TaV 2A were made to help achieve complete cleavage between BEV $2A^{pro}$ and the [Δ1A-GFP-PAC-Δ1D] fusion protein, and/or FMDV 2B, in order to increase processivity and replication of the pBEV 2A replicon.

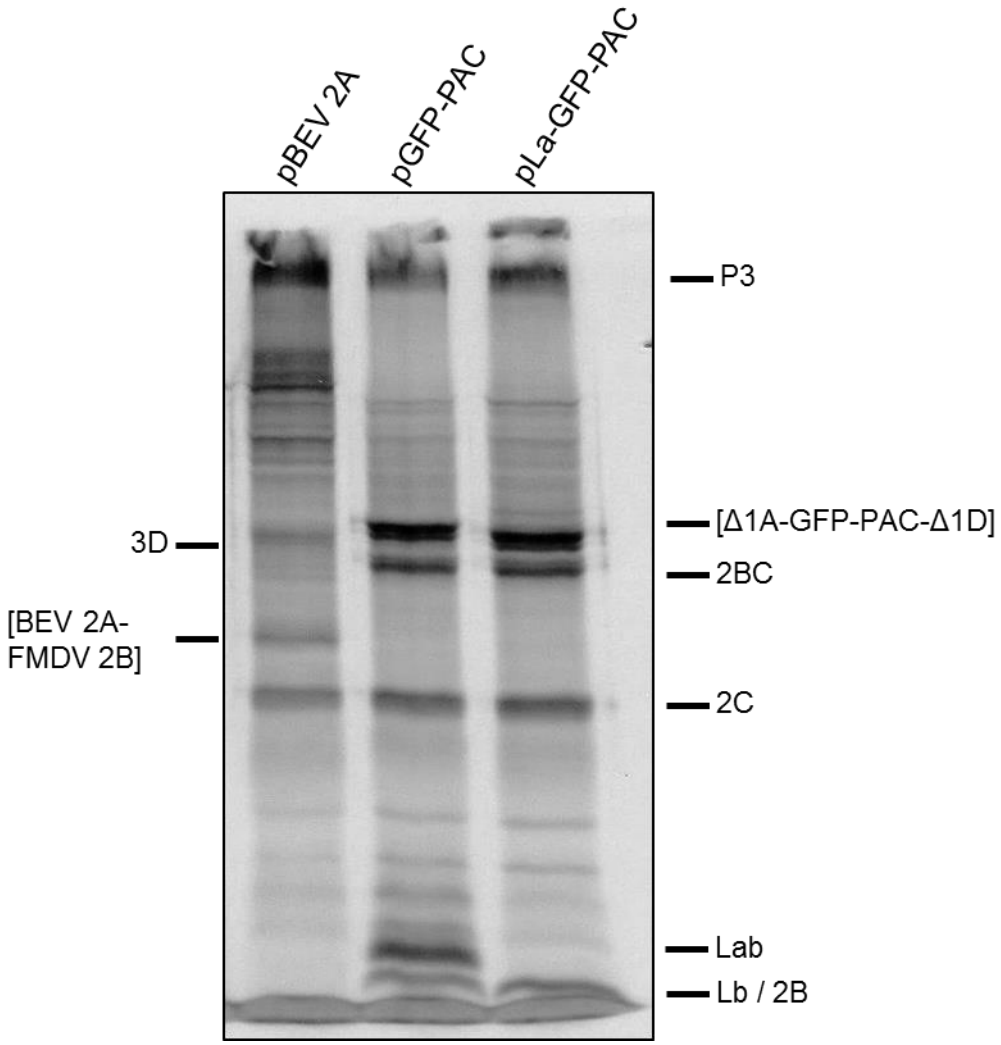


Figure 6.2. *In vitro* translation of pBEV 2A, pGFP-PAC and pLa-GFP-PAC replicons. Rabbit reticulocyte lysates (Promega) were programmed with the plasmids indicated and radiolabelled products examined by 12 % SDS-PAGE. ^{35}S-Met labelled viral proteins were visualised *via* autoradiography and predicted positions are shown.

6.3 BEV 2A replicon-derived GFP expression

Replicons were linearised with AscI and RNA synthesised as described during previous chapters. Transcript RNA (1 µg) was transfected into BHK-21 cells and fluorescence monitored for the duration of the experiment. The 'WT' pRbz_5nts_mod, pBEV-TaV 2A and pBEV Δ1D-TaV 2A replicons displayed fluorescence, with all other replicons tested showing kinetics similar to the replication-incompetent pRbzΔ3D replicon (Figure 6.3, 6.4). The lack of fluorescence in both the pBEV 2A and pBEV Δ1D-2A replicons was most likely due to inefficient cleavage at the BEV $2A^{pro}$/ FMDV 2B junction by $3C^{pro}$. This would result in a [BEV $2A^{pro}$-FMDV 2B] fusion protein, impeding the activity of this protein, and therefore preventing replication to proceed as normal.

There was a lag in the detection of GFP fluorescence of ~2 hr for both pBEV-TaV 2A and pBEV Δ1D-TaV 2A replicons (Figure 6.4), despite encoding self-cleaving ribozymes which were shown to increase replication kinetics (Chapter 4). The total fluorescence intensity was ~10-fold lower for the pBEV Δ1D-TaV 2A replicon and ~40-fold lower for pBEV-TaV 2A in comparison to the pRbz_5nts_mod replicon (Figure 6.4B). On average throughout the experiment, the fluorescent signal generated by pBEV Δ1D-TaV 2A and pBEV-TaV 2A was ~2- and ~6-fold lower, respectively (Figure 6.4B). GFP cell counts were also reduced, with an overall decrease of ~3-fold for pBEV Δ1D-TaV 2A and ~10-fold for the pBEV-TaV 2A replicon (Figure 6.3, 6.4A). This suggests that insertion of 10 aa of BEV 1D and/or TaV 2A improved processing which had an effect on replication. These results also imply that full replicative fitness could not be restored when replicons lacking L^{pro}, but encoding $2A^{pro}$ with a similar function, were transfected into BHK-21 cells. More experiments are needed to fully investigate these findings.

In the case of the pBEV-TaV 2A, it cannot be ruled out that a decrease in GFP signal, and therefore GFP positive cell counts, was due to a loss in fluorescence as a consequence of BEV $2A^{pro}$ remaining fused to FMDV Δ1D, creating a [Δ1A-GFP-PAC-Δ1D-BEV $2A^{pro}$] fusion protein. Examination of protein extracts from transfected cells with anti-GFP and FMDV-specific antibodies would clarify the presence of such a fusion protein, and would also allow for viral protein production to be assessed.

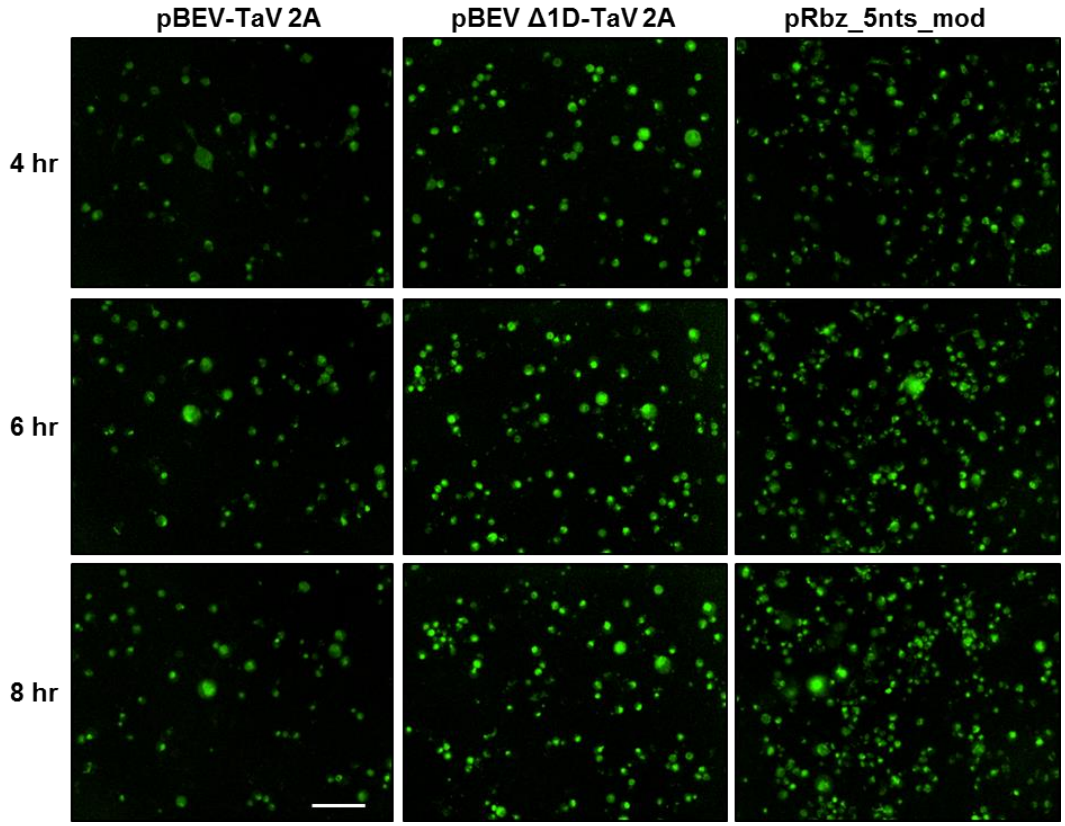


Figure 6.3. GFP expression derived from pBEV 2A replicon-transfected BHK-21 cells. Transcript RNA from the pRbz_5nts_mod, pBEV-TaV 2A and pBEV Δ1D-TaV 2A replicons was introduced into cell monolayers, and fluorescent images captured at 2 hr intervals over a 20 hr period using the IncuCyte ZOOM imaging system. Data from mock, pBEV 2A, pBEV Δ1D-2A and pRbzΔ3D replicon-transfected cells is not shown. A representative of the nine images captured is shown at 4, 6 and 8 hr post-transfection, and was obtained from 1 independent transfection, with 4 replicates per transfection. Scale bar represents 100 μm.

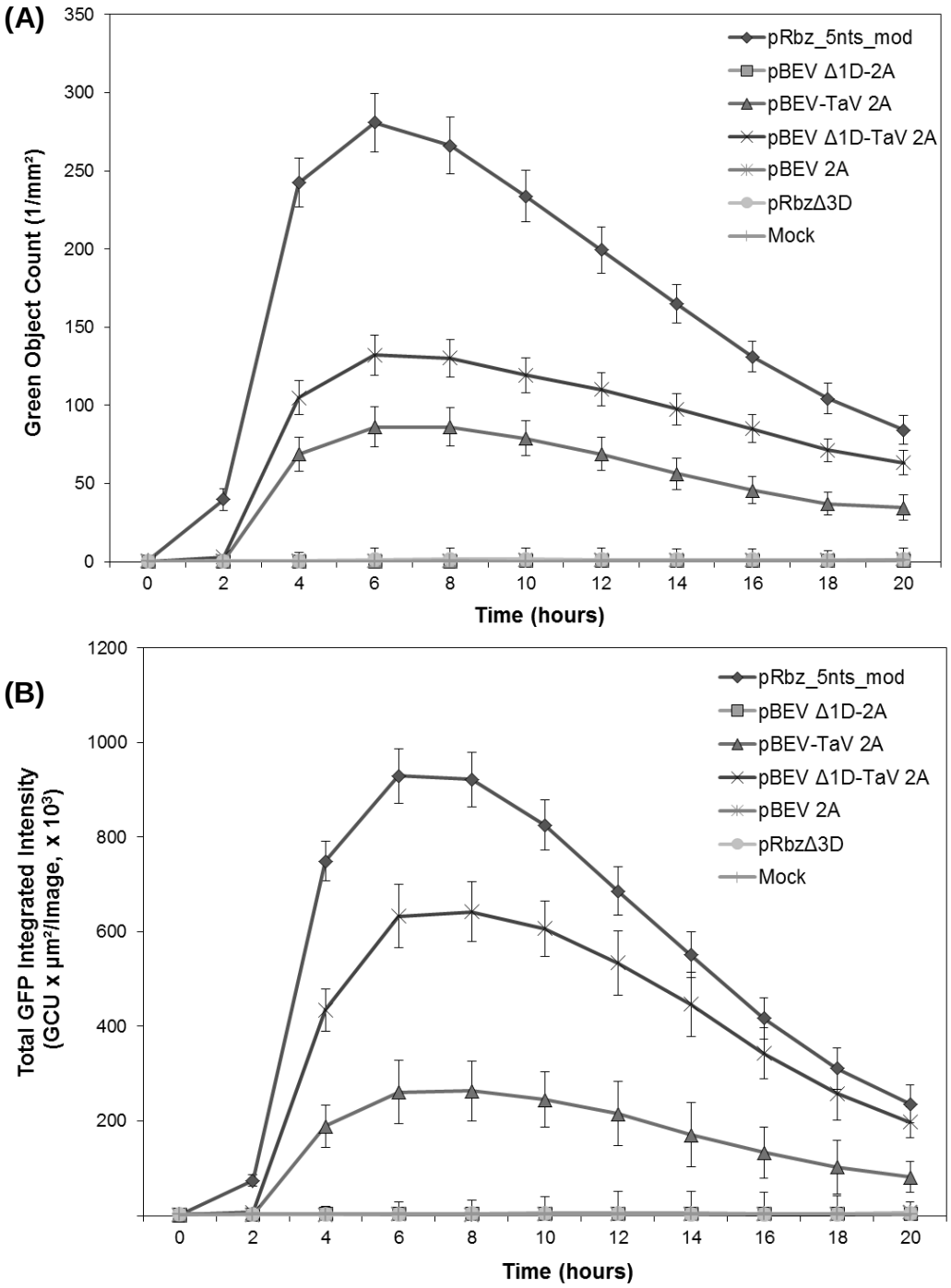


Figure 6.4. pBEV 2A replicon-derived GFP fluorescence. Data from mock-transfected BHK-21 cells is shown, together with cells transfected with transcript RNA derived from the pRbz_5nts_mod, pBEV 2A, pBEV Δ1D-2A, pBEV Δ1D-TaV 2A, pBEV-TaV and pRbzΔ3D replicons. At the time points indicated images were captured over a 20 hr period and the GFP fluorescence quantified for each replicon construct: data shown as the green object count/mm^2 (A) or as the total integrated GFP fluorescence intensity (B). Data points/error bars shown are derived from 1 independent transfection, with 4 replicates.

6.4 Conclusions

- Attenuation of FMDV replicons could be achieved by replacing the FMDV 2A peptide, encoding a ribosomal 'skipping' mechanism, with a 2A proteinase.

- Initial experiments demonstrated a lack of fluorescent signal from the pBEV 2A replicon following transfection of transcript RNA into BHK-21 cells. TnT reactions, programmed with pBEV 2A, pGFP-PAC and pLa-GFP-PAC showed that the ratio of up-and downstream proteins produced by replicons encoding BEV $2A^{pro}$ were quite different from replicons encoding FMDV 2A.

- Insertion of BEV Δ1D and/or TaV 2A increased fluorescence of the pBEV 2A replicon, with a combination of both modifications having the greatest effect. This suggests that cleavage at both junctions flanking BEV $2A^{pro}$ was impaired within the original pBEV 2A replicon.

- GFP expression from both the pBEV Δ1D-TaV 2A and pBEV-TaV 2A replicons was reduced when compared to the pRbz_5nts_mod replicon, indicating that replicons encoding BEV 2A were attenuated, and suggested that full replicative fitness could not be restored to a 'WT' replicon encoding both L^{pro} and FMDV 2A.

- Coupled transcription/translation experiments *in vitro* with each of the newly created replicons will gain insight into the ratio of proteins produced by these constructs, with further experiments investigating regulation of this process during cellular stress.

- Western blots with anti-eIF4G antibodies on replicon-transfected cell extracts will determine if BEV $2A^{pro}$ is active and able to carry out cleavage of this host translation factor. Incomplete processing at the FMDV Δ1D-BEV 2A junction within the pBEV-TaV 2A replicon will be examined using anti-GFP antibodies to determine whether this is the main factor contributing to reduced fluorescence.

6.5 Discussion

6.5.1 FMDV replicons encoding Bovine Enterovirus 2A are replication competent and attenuated

Interestingly, insertion of BEV $2A^{pro}$ into the pRbz_5nts_mod replicon produced replication competent – but attenuated - genomes. Initial experiments showed no detectable replication for the pBEV 2A replicon, most likely due to incomplete polyprotein processing at either the BEV $2A^{pro}$/ 2B junction (by $3C^{pro}$) and the FMDV Δ1D/ BEV $2A^{pro}$ junction (by BEV $2A^{pro}$) – or both. A unique $3C^{pro}$ cleavage site was introduced between the 2A/2B junction, likely producing a [BEV$2A^{pro}$-FMDV 2B] fusion protein as a result of incomplete processing which, in turn, could have affected correct processing of the Δ1D/ BEV $2A^{pro}$ junction. Although the Δ1D/ BEV $2A^{pro}$ region was designed to contain the appropriate residues to facilitate BEV $2A^{pro}$ cleavage (-YG- pair; Stanway, 1990), incomplete processing at this region is probable. Initial IncuCyte experiments revealed no detectable fluorescence; therefore further replicons were constructed to overcome the possible polyprotein processing issues. Ten amino acids of BEV 1D were inserted immediately upstream of BEV $2A^{pro}$ (replacing sequences encoding FMDV Δ1D) to aid in BEV $2A^{pro}$ N-terminal processing. To address the processing issue at the C-terminus of BEV $2A^{pro}$, TaV 2A was inserted immediately after BEV $2A^{pro}$ to induce ribosome skipping-mediated cleavage. $3C^{pro}$- mediated cleavage at this region may have been achieved by the insertion of other $3C^{pro}$ cleavage residues (-EG-, -QG-, -EL- and -QS- dipeptides; Ryan & Flint, 1997), however, introduction of TaV 2A was considered the easiest and quickest modification to perform initially. Finally, a replicon containing both BEV Δ1D and TaV 2A was constructed. If insertion of TaV 2A proves successful, with replication increasing due to induced cleavage at the BEV $2A^{pro}$ C-terminus, the addition of potential $3C^{pro}$ dipeptides will then be investigated in conjunction with removal of TaV 2A.

Fluorescence data generated from these new BEV replicons displayed decreased signal intensities and the number of GFP positive cells was also lower than WT. The pBEV Δ1D-2A replicon exhibited no fluorescence upon cell transfection, suggesting impaired cleavage of BEV $2A^{pro}$ from the polyprotein (2B) at its C-terminus by FMDV $3C^{pro}$. The reduced fluorescent signal of both the

pBEV-TaV 2A and pBEV Δ1D-TaV2A replicons implies attenuation of these replicons. The WT replicon, pRbz_5nts_mod, encodes a fully functional L^{pro} and therefore Aq.GFP will undergo cleavage induced by L^{pro} (Chapter 3; Tulloch *et al.*, 2014a). Leaderless replicons generate higher signal intensities in comparison, due to Aq.GFP remaining un-cleaved. Picornaviral $2A^{pro}$ cleavage within eIF4G occurs between an -RG- pair (Lamphear *et al.*, 1993), with N-terminal cleavage occurring at a -YG- pair within the polyprotein. There are no -RG- dipeptides present within the Aq.GFP coding sequence; however two -YG- pairs exist (Chapter 5, Figure 5.5). It cannot be ruled out that BEV $2A^{pro}$ can cleave at these sites; however like $3C^{pro}$, the context of these dipeptide pairs within the polyprotein determines cleavage and not just their presence (Ryan & Flint, 1997). Nonetheless, western blots using anti-GFP antibodies must be performed to eliminate this possibility. Incomplete cleavage at the FMDV Δ1D/ BEV$2A^{pro}$ junction within the pBEV-TaV2A replicon could also result in the formation of an Aq.GFP fusion protein which could affect fluorescence. Thus, replicon-transfected cells generated from this replicon must also be probed with anti-GFP antibodies for confirmation. Investigation of eIF4G levels over the course of replication will determine if BEV $2A^{pro}$ is active. HRV $2A^{pro}$ cleavage activity of eIF4G was shown to be three times slower than FMDV L^{pro} (Glaser & Skern, 2000), therefore this must be taken into account when comparing the relative levels of fitness between replicons.

6.5.2 Fitness of BEV 2A replicons could not be restored to that of WT replicons encoding L^{pro}

Replicons encoding BEV $2A^{pro}$ could not be restored to a similar level of replicative fitness as a replicon encoding L^{pro} (pRbz_5nts_mod). Both L and 2A proteinases cleave eIF4G, with the cleavage sites being only 6 amino acids apart (Glaser & Skern, 2000). This is the main substrate for each of these enzymes, even though these proteinases are quite different in sequence composition and proteolytic activity (2A – a serine protease *vs* L – a cysteine protease; Ryan & Flint, 1997). FMDV L^{pro}, however, has also been implicated to antagonise the innate immune response *via* degradation of NF-κB (de los Santos *et al.*, 2007, 2009). The observation that replicative fitness could not be restored in replicons lacking the Lb coding sequence, but encoding BEV $2A^{pro}$ which performs the same main function as L^{pro}, may strengthen the findings of de Los Santos *et al*., (2007, 2009).

6.5.3 Regulation of virus protein biogenesis by the FMDV 2A ribosome skipping mechanism can be studied using BEV 2A replicons

Regulation of the 2A-mediated cleavage mechanism ("ribosome skipping") is unknown. 2A 'cleavage' involves host cell translation termination factors, eRF1 and eRF3. These RFs were shown to relieve the pause in translation induced by 2A by hydrolysing the peptidyl-tRNA ester linkage, causing release of the nascent protein (Doronina *et al.*, 2008). Synthesis of sequences downstream of 2A can then commence through re-initiation, or the ribosome subunits dissociate and translation ceases (Brown & Ryan, 2010). Involvement of these RFs raises the possibility that this mechanism could be influenced by cellular pathways that regulate their activity. Translation elongation involves delivery of an aminoacylated tRNA into the ribosomal A site by eukaryotic elongation factor 1A (eEF 1A). Anticodon pairing of the aminoacyl-tRNA in the A-site with the mRNA codon leads to formation of a peptide bond with the peptidyl-tRNA within the P site. eEF2 is responsible for catalysing the movement of both peptidyl- (A-site) and deacylated tRNAs (P-site) within the ribosome (reveiwed in Kaul *et al.*, 2011). It is known that eEF2 is phosphorylated by eEF2 kinase (eEF2K) under cell stress conditions. Phosphorylated eEF2 is inactive and inhibits the activity of un-phosphorylated eEF2 leading to a decrease in translation (Ryazanov *et al.*, 1988). Translation elongation is a process which consumes a large amount of cellular energy; therefore during periods where there is an increased demand for energy, or a decrease in energy supply, it would be beneficial to decrease the rate of protein synthesis. This decrease in translation would lead to a reduction in the synthesis of sequences downstream of 2A, as eEF2 would be required for translation to resume following 2A-mediated cleavage (Brown & Ryan, 2010). Increased synthesis of the capsid coding region (located upstream of 2A) would be favourable during such cellular conditions to encapsidate RNA genomes present within the cytoplasm, leading to a higher yield of infectious particles. Interestingly, eEF2 does indeed become progressively phosphorylated during TMEV (Theiler's murine encephalitis virus) infection (L. Brown, unpublished observations). The proposed model of 2A-mediated cleavage is shown below in Figure 6.5.

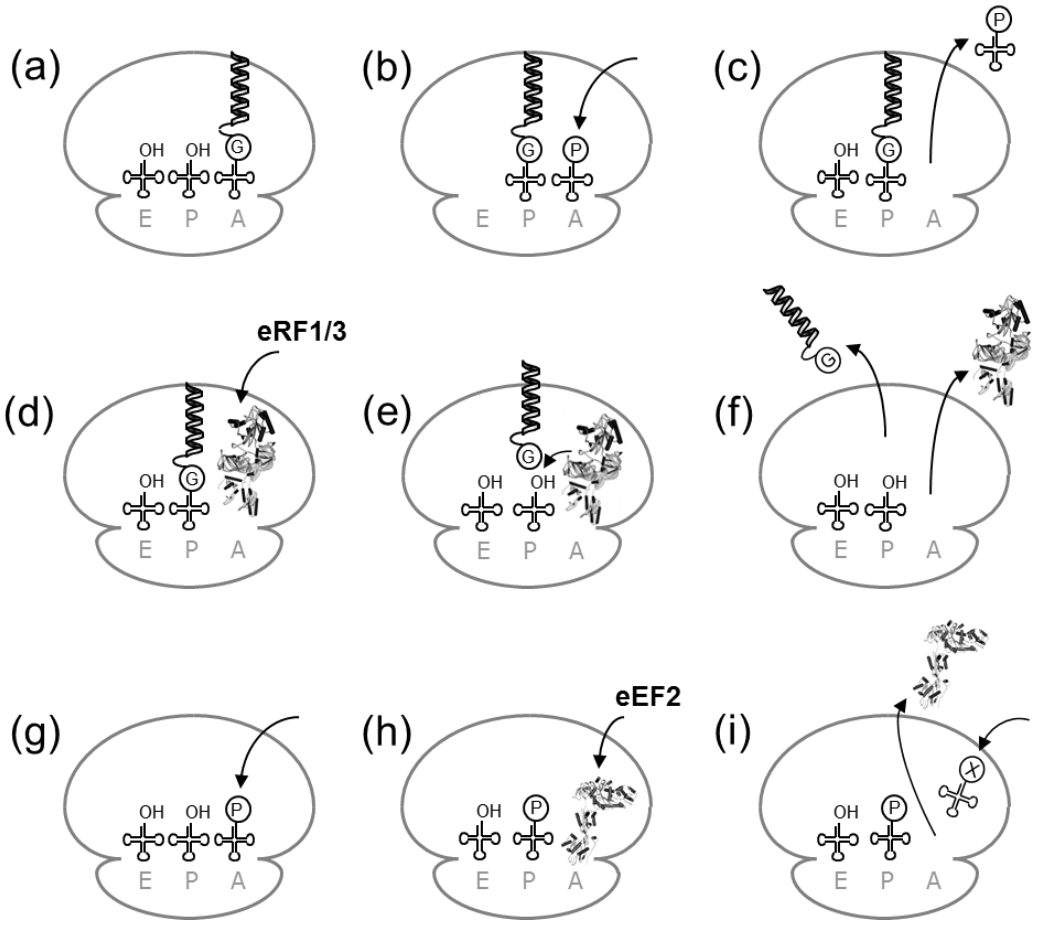


Figure 6.5. Outline of 2A-mediated 'cleavage'. Peptidyl-tRNA is located in the ribosome A-site (a). Peptidyl-tRNA is translocated to the P-site, allowing prolyl-tRNA to enter the A-site (b). Interactions between 2A and the ribosome, and the tight turn, prevents the peptidyl-tRNA ester linkage from undergoing nucleophilic attack, as prolyl-tRNA dissociates from the ribosome (c). eRF1 enters the A-site (d) and hydrolyses the ester bond (e). eRF1 then leaves the A-site (promoted by eRF3), and the nascent peptide is released from the ribosome (f). Prolyl-tRNA (re)enters the A-site (g) and is translocated to the P-site by eEF2 (h). The subsequent amino acid enters the A-site and sequences downstream of 2A are translated (i), or, the ribosome subunits dissociate and translation it terminated.

6.5.4 Translation profiles of FMDV replicons altered by insertion of BEV 2A^{pro}

The translation profile of the pBEV 2A replicon was initially investigated in comparison to pGFP-PAC and pLa-GFP-PAC replicons. Preliminary results indicated that up- and down-stream protein production was altered within the pBEV 2A replicon. Both the pLa-GFP-PAC and pGFP-PAC replicons generated translation profiles indicative of polyproteins encoding a ribosome skipping sequence. The decrease in upstream protein synthesis within the pBEV 2A replicon (i.e. GFP-PAC) could be attributed to the lack of FMDV 2A which mediates the production of a molar excess of proteins encoded upstream. These initial findings can be used to investigate the hypothesis outlined above for 2A-mediated regulation of FMDV translation and prompts further experimentation.

Further studies using the replicons described within this chapter will be designed to examine (i) the ratios of proteins up- and down-stream of 2A synthesised throughout replication, (ii) host-cell shut-off by viral proteinases (L^{pro} & BEV 2A^{pro}), (iii) particle yield derived from infectious copies, and (iv) interactions of L^{pro} (or 2A^{pro}) with the host-cell immune response. Protein ratios will be difficult to quantify due to the multiple precursors synthesised by the FMDV genome. In an attempt to measure differences in protein synthesis between replicons containing BEV 2A^{pro} and FMDV 2A, pulse-labelling experiments will be performed involving immunoprecipitation of both cellular and *in vitro* translation extracts with anti-GFP and anti-3D antibodies. Protein levels will be quantified and differences between each replicon calculated. This will allow *de novo* synthesis of both upstream and downstream to be measured, rather than accumulated protein levels. The levels of eEF2 and phosphorylated-eEF2 will also be monitored within both WT and BEV replicon transfected cells. This will be correlated with protein levels synthesised using the differing 2A's to examine whether this translation factor has a role in regulating FMDV 2A ribosome-skipping.

Before such experiments can commence, polyprotein processing within these replicons must be improved through modification of the C-terminus of BEV $2A^{pro}$ with varying $3C^{pro}$ dipeptides (following removal of TaV 2A) and passaging RNA from replicon transfected cells, or, by virus rescue and passage (Pirbright Institute): genomes would be sequenced throughout the process of passage and the properties of this new form of picornavirus compared with FMDV.

The final results chapter of this thesis describes the influence of dinucleotide frequency on Echovirus 7 growth kinetics and the potential application of this strategy to the FMDV replicon system for vaccine development.

Chapter 7: RNA Virus Attenuation by Codon-pair De-Optimisation is an Artefact of Increases in CpG/UpA Dinucleotide Frequencies.

7.1 Introduction

There are a number of factors which contribute to translation regulation within eukaryotes, bacteria and viruses. These comprise codon usage, relative aminoacyl-tRNA abundancies, elongation rates (ribosome 'processivity') and translation regulation (Gingold & Pilpel, 2011). Biases exist toward certain synonymous codons (the "codon bias") and codon-pairs (the "codon-pair bias"). Codon-pair bias refers to the preferential pairing of certain codons over others (e.g. in human genes the codon-pair GCC-GAA encodes the adjacent amino acids alanine-glutamic acid less often than GCA-GAG, even though GCC and GAA are the most used codons; (Luke *et al.*, 2013). These biases are thought to influence gene expression by altering translation efficiency.

Both codon and codon-pair de-optimisation have been adopted as strategies for attenuating viral replication within RNA viruses (Burns *et al.*, 2006; Coleman *et al.*, 2008; Martrus *et al.*, 2013; Mueller *et al.*, 2006, 2010; Ni *et al.*, 2014). This approach involves replacement of codons within the viral coding region for synonymous non-preferred codons or codon-pairs. It is thought that by altering codon usage, translation efficiency is altered resulting in reduced viral replication. The design of such viruses has the potential for the production of safer, non-reverting, live-attenuated vaccines, since attenuation arises from the incorporation of literally 100s of mutations – each of which only very slightly reduces replicative fitness, but taken together produce significant attenuation. Traditional methods of producing live, attenuated, vaccines involve attenuation of viral genomes by serial passage in tissue-culture, such as the OPV Sabin vaccine which had huge impacts on human health. This method of attenuation is time consuming and relies upon chance alone to generate attenuating mutations which then must undergo another lengthy process of selection to produce the required phenotype. However, using this method often a small number of key (attenuating) mutations may back mutate leading to reversion to virulence. The introduction of a large number of synonymous changes within the genome

reduces, therefore, the risk of back-mutation increasing the genetic stability of the vaccine strain.

Another compositional feature which can affect the replication phenotype is the suppression of CpG and UpA dinucleotide frequencies among RNA and small DNA viruses (Rima & Mcferran, 1997; Simmonds *et al.*, 2013). Artificially increasing the dinucleotide frequencies has been shown to impair replication kinetics of Echovirus 7 (E7) and PV (Atkinson *et al.*, 2014; Burns *et al.*, 2009). The authors of these studies speculate that the selection of disfavoured codon-pairs which attenuate PV and other viruses alters the CpG/UpA frequency within the genome, and it is this which accounts for, the most part (if not all!), the attenuating effects observed. Indeed, Burns *et al.*, (2009) demonstrated that their previous strategy using codon de-optimised PV mutants to attenuate viral replication was due to greatly enhanced frequencies of CpG and UpA dinucleotides, rather than alterations in codon or codon-pair bias.

The study described in this chapter sought to disentangle the relationship between codon-pair bias (CPB) and dinucleotide frequencies in a re-examination of their effects on the replication of an RNA virus, E7. To resolve this functionally, comparisons were made of the replication kinetics and relative fitness of native E7 with a series of novel mutants of E7 in which dinucleotide frequencies and CPB were independently manipulated.

7.2 Construction of mutant viruses and insert compositions

Two regions of the E7 genome were replaced within P1 and P3 (Figure 7.1), used previously to examine the effects of dinucleotide frequencies on virus replication (Atkinson *et al.*, 2014). These areas were selected for mutagenesis due to (i) their lack of predicted RNA secondary structure and (ii) the absence of RNA elements required for replication or translation (Atkinson *et al.*, 2014). Mutants which contained a pre-determined CPB score whilst retaining the composition of CpG and UpA dinucleotide frequencies were generated using computational analysis with a specifically developed computer program (Sequence Mutate in the SSE package, performed by Professor P.Simmonds; Simmonds, 2012). This allowed numerous synonymous changes to be introduced within the E7 genome: the amino acid sequence of all virus proteins remained identical. Gene blocks comprising either (i) altered dinucleotide frequencies, or, (ii) altered CPB scores were synthesised. These sequences were flanked by the

unique restriction sites SalI (genome position 1878) and HpaI (3119) for region 1 and EcoRI (5403) and BglI (6462) for region 2 (Figure 7.1). The E7 cDNA clone pT7:E7 was digested with the enzymes described and the appropriate gene block inserted.

The mutant Min-E possessed the minimum possible CPB score (-0.111) while retaining dinucleotide frequencies identical to WT (CP score: -0.014; CpG: 0.525; UpA: 0.718; Figure 7.2; Table 7.1). Mutants comprising identical CPB scores as Min-E but with slightly elevated dinucleotide frequencies (Min-U; CpG: 0.82; UpA: 0.95), or where dinucleotide frequencies were maximised (Min-H; CpG: 1.3; UpA: 0.98) were also generated (Figure 7.2; Table 7.1). These mutants allowed the effects of dinucleotide frequencies on virus replication to be analysed independently of CPB. The mutant Max-U comprised a maximised CPB score (0.320), but contained similar CpG/UpA frequencies to WT. This mutant could be compared with the previously described CpG/UpA low cu|cu mutant which contained a CPB score slightly greater than WT (CpG; 0; UpA; 0.22; CPB; 0.11; Figure 7.2; Table 7.1). The permuted mutant (P|P) was designed such that codon order was permuted, while retaining protein coding and WT dinucleotide frequencies. The CpG/UpA low mutant (cu|cu) was designed such that (i) all CpG dinucleotides were eliminated and (ii) the maximum number of UpA dinucleotides as possible were removed - as described previously (Atkinson *et al.*, 2014).

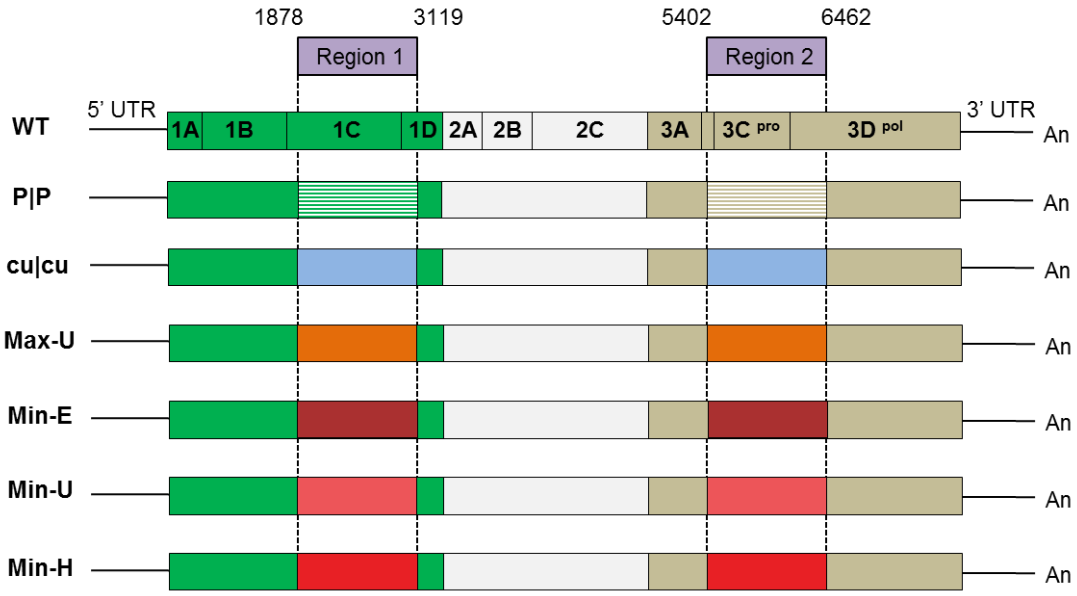


Figure 7.1. Schematic of Echovirus 7 mutant constructs with insert positions aligned to WT. The E7 coding region is shown in boxed areas: the capsid (P1) region is shown in green and the replication proteins domain P3 shown in beige. Differently coloured or striped boxes indicate modified cassettes which have altered CPB or dinucleotide frequencies (described in the text). Nucleotide positions were calculated relative to the pT7:E7 cDNA sequence.

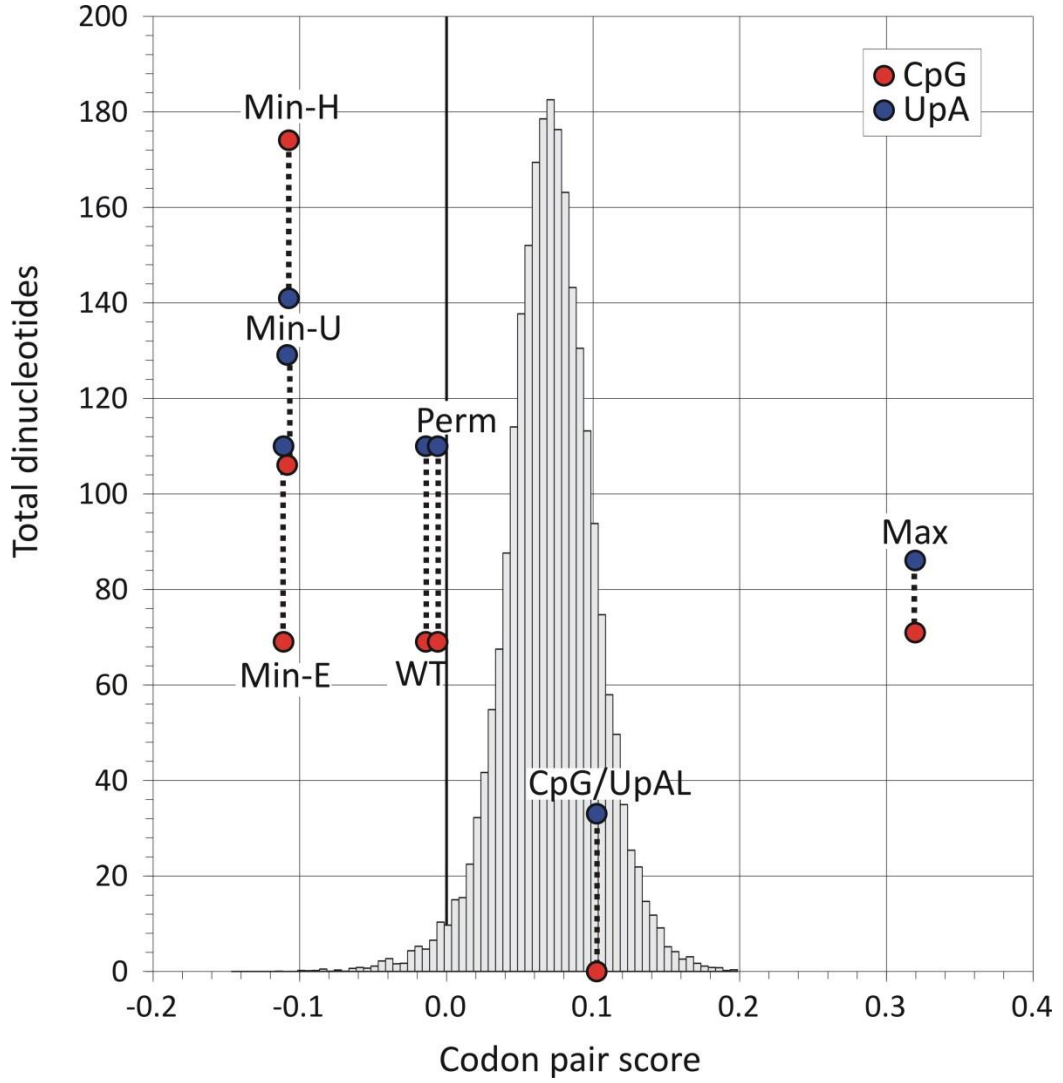


Figure 7.2. Codon-pair scores and numbers of CpG and UpA dinucleotides in native (WT) and mutated regions of E7. Mean codon-pair scores for Regions 1 and 2 shown on x-axis; total of CpG and UpA dinucleotides in each sequences are shown on the y-axis. The histogram shows codon-pair scores for the 35170 human mRNA sequences >200 bases in length (mean 0.072; standard deviation ± 0.031). CpG/UpAL = cu|cu (image generated by Prof. P. Simmonds).

Table 7.1. Composition and codon usage of E7 WT and mutant insert sequences

Region	Sequence (Symbol)	G+C content	CpG Total[1]	O/E ratio[2,3]	UpA Total	O/E ratio[2,3]	Codon Usage CAI[4]	ENc	CP Bias[3]
1	Native (WT)	47.6%	51 (-)	0.730	62 (-)	0.742	0.685	56.5	-0.043
	Permuted (P)	47.6%	51 (0)	0.730	2 (0)	0.742	0.694	55.8	-0.025
	CpG/UpAL (cu)	47.5%	0 (-51)	0	19 (-43)	0.227	0.686	43.5	0.087
	Max-U	50.1%	47(-4)	0.610	43 (-19)	0.573	0.708	49.6	0.328
	Min_E	47.5%	51 (0)	0.736	62 (0)	0.735	0.748	54.3	-0.131
	Min_U	47.5%	69 (+18)	0.992	76 (+14)	0.939	0.709	58.3	-0.134
	Min_H	49.8%	106(+55)	1.400	79 (+17)	0.981	0.696	49.2	-0.130
2	Native (WT)	47.1%	18 (-)	0.320	48 (-)	0.695	0.743	53.2	0.015
	Permuted (P)	47.6%	18 (0)	0.320	48 (0)	0.695	0.739	49.0	0.013
	CpG/UpAL (cu)	48.5%	0 (-18)	0	48 (0)	0.214	0.739	47.2	0.118
	Max-U	46.3%	24 (+6)	0.440	43 (-3)	0.601	0.750	46.1	0.311
	Min-E	45.7%	18 (0)	0.343	48 (0)	0.657	0.785	53.3	-0.091
	Min-U	47.4%	37 (+19)	0.649	50 (+2)	0.738	0.767	57.6	-0.083
	Min-H	47.8%	68 (+50)	1.172	65 (+15)	0.970	0.715	49.7	-0.085

[1]Total number of CpG and UpA dinucleotides in sequence. Changes in numbers between mutated and original WT sequence are indicated in parentheses

[2]Ratio of observed dinucleotide frequency (O) to that expected based on mononucleotide composition (E) *ie.* f(CpG) / f(C) * f(G).

[3]Values deliberately changed are shown in red (maximised) and blue (minimised).

[4]Calculatedfromhttp://genomes.urv.es/CAIcal/(Puigbò *et al.*, 2008).

(table generated by Prof. P. Simmonds).

Table 7.2. Relationship between codon-pair de-optimisation, CpG and UpA Frequencies and Virus Fitness Reduction.

Region	Gene	Prop'n	WT CPB	WT CpG	WT UpA	CPD CPB	CPD CpG	CPD UpA	Replication Reduction	Ref
Poliovirus										
PV-X	Capsid	14.8%	-0.03	0.52	0.75	-0.46	1.34	1.25	x 25	(Coleman *et al.*, 2008)
PV-XY	Capsid	25.9%	-0.03	0.54	0.75	-0.46	1.31	1.27	x 400	
*Influenza A virus**										
HA^{Min}	Segs.4	11.4%	0.02	0.43	0.64	-0.42	1.65	1.11	x 3.5	(Mueller *et al.*, 2010)
HA/NP^{Min}	Segs.4,5	21.3%	0.02	0.44	0.55	-0.42	1.56	1.14	x 14	
$PR8^{3F}$	Segs.1,4,5	29.1%	0.01	0.43	0.53	-0.41	1.55	1.07	x 35	
HIV-1										
A	*gag*	4.6%	0.03	0.47	1.04	-0.43	1.43	1.25	x 7	(Martrus *et al.*, 2013)
B	*gag*	4.7%	0.08	0	0.91	-0.37	1.22	1.15	x 3	
C	*gag*	4.8%	0.03	0.31	1.00	-0.38	1.50	1.09	x 8	
D	*gag*	2.1%	-0.02	0	0.49	-0.42	1.47	0.99	x 1.5	
PRRSV										
SAVE5	gp5	2.6%*	-0.06	0.63	0.73	-0.38	1.37	1.14	x 4*	(Ni *et al.*, 2014)

*Codon-pair minimised sequences of IAV were not provided in (Mueller *et al.*, 2010) and for the purposes of comparison these have been reconstructed in SSE. Note that the codon-pair scores described in Table 1 of that paper (-0.386, -0.420 and -0.421 for PB1, HA and NP respectively) are not minimum scores; these are in fact -0.533, -0.585 and -0.602. Therefore, for the purposes of comparison, codon-pair score minimisation in the current study was targeted to those values. Although the sequences generated by SSE are not identical to those obtained previously, they would demonstrate a similar distortion of dinucleotide frequencies to those used in the previous study (Coleman *et al.*, 2008).

*Mutated region only (positions 147-542 in gp5)

*Data from replication assay in PAM cells.

(table generated by Prof. P. Simmonds).

Table 7.3. Enzymes used in selective digests for competition assays. Pairwise fitness comparisons between WT and mutant viruses are listed with the restriction enzyme used to determine each virus. The RE target and region amplified is also shown.

Virus 1	Virus 2	Region	Enzyme	Target
W|W	P|P	1	*Spe*I	Permuted
W|W	Max-U	1	*Sac*I	Max
W|W	Min-E	1	*Nco*I	WT
W|W	Min-U	1	*Nco*I	WT
W|W	Min-H	1	*Eco*RV	WT
W|W	cu|cu	1	*Eco*RV	WT
P|P	cu|cu	1	*Spe*I	Permuted
Max-U	P|P	1	*Spe*I	Permuted
Max-U	cu|cu	1	*Sac*I	Max
Min-E	Min-U	1	*Cla*I	Min-U
Min-E	Min-H	1	*Eco*RV	Min-E
Min-U	Min-H	1	*Cla*I	Min-U

7.3 The effect of CPB or dinucleotide frequency on viral replication

Coding regions of poliovirus, influenza A virus, PRRSV and HIV-1 have all been subjected to codon-pair de-optimisation (CPD) and effects on virus replication quantified (Coleman *et al.*, 2008; Martrus *et al.*, 2013; Mueller *et al.*, 2010; Ni *et al.*, 2014). Despite their diversity of replication and translation mechanisms, each showed a similar relationship between the extent of CPD and reduction in virus replication ability (Table 7.2). Typically, 10-fold or greater attenuation in cell culture required >12-15% replacement of WT genome with CPD sequences. It is notable, however, that for each virus, CPD invariably increased frequencies of CpG and UpA dinucleotides (Table 7.2), typically from 0.4-0.6 to 1.4-1.6 (CpG) and from 0.5-0.8 to 1.1-1.4 (UpA) in the mutated regions.

RD cell monolayers were infected with transcript RNA transcribed from NotI linearised cDNA from both WT and mutant viruses. Infectious virus was recovered from transfected cells, and supernatants titrated into 96-well plates by measurement of $TCID_{50}$ values. Multi-step growth curves were performed to investigate replication kinetics of the E7 mutants in comparison to WT. RD cells were infected at an MOI of 0.01, and viral supernatants harvested at 6, 24 and 42 hr post infection (p.i.; Figure 7.3). Experiments were performed in triplicate. Replication kinetics of the P|P and cu|cu viruses were evaluated previously (Atkinson *et al.*, 2014) and have been omitted from the current graph. Over the course of the experiment both the Min-U and Min-E viruses showed similar replication to WT, with a slight reduction in titre at 24 hr p.i. Max-U also displayed kinetics similar to that of WT, with a higher titre produced at 42 hr p.i. Contrastingly, the Min-H mutant produced ~30-fold less infectious virus than WT over the duration of the experiment, with a ~70-fold reduction at 24 hr p.i. Cell monolayers were destroyed for all viruses tested with the exception of Min-U and Min-H mutants, which displayed incomplete CPE (data not shown).

7.4 Determination of viral fitness

The relative fitness of wild-type and the various mutant viruses was investigated *via* competition assays. RD cells were co-inoculated with equal MOIs of each virus combination and serially passaged up to ten times. vRNA was extracted from harvested supernatants and PCR amplified across the modified region. The relative compositions of vRNA populations were determined by cleavage with

restriction enzymes of the PCR amplification products: this differentiated between WT and mutant sequences (Table 7.3). Twelve pair-wise combinations were examined (Figure 7.4). Examples of three combinations are shown (Figure 7.5);

(i) Max-U showed similar fitness to WT, but a greater population representation at passage 10 (P10).
(ii) Max-U was completely out-competed by cu|cu by P10, and in the final example shown,
(iii) WT and Min-E displayed equal populations at P5, and P10 (data not shown).

All other combinations tested were at completion by P5 and are summarised along with the examples described (Figure 7.4, 7.6). Thus, the fitness ranking indicated was as follows; cu|cu > Max-U > (WT = P|P = Min-E) > Min-U > Min-H. These results demonstrate that when altered independently from CPB, only dinucleotides were associated with replication fitness.

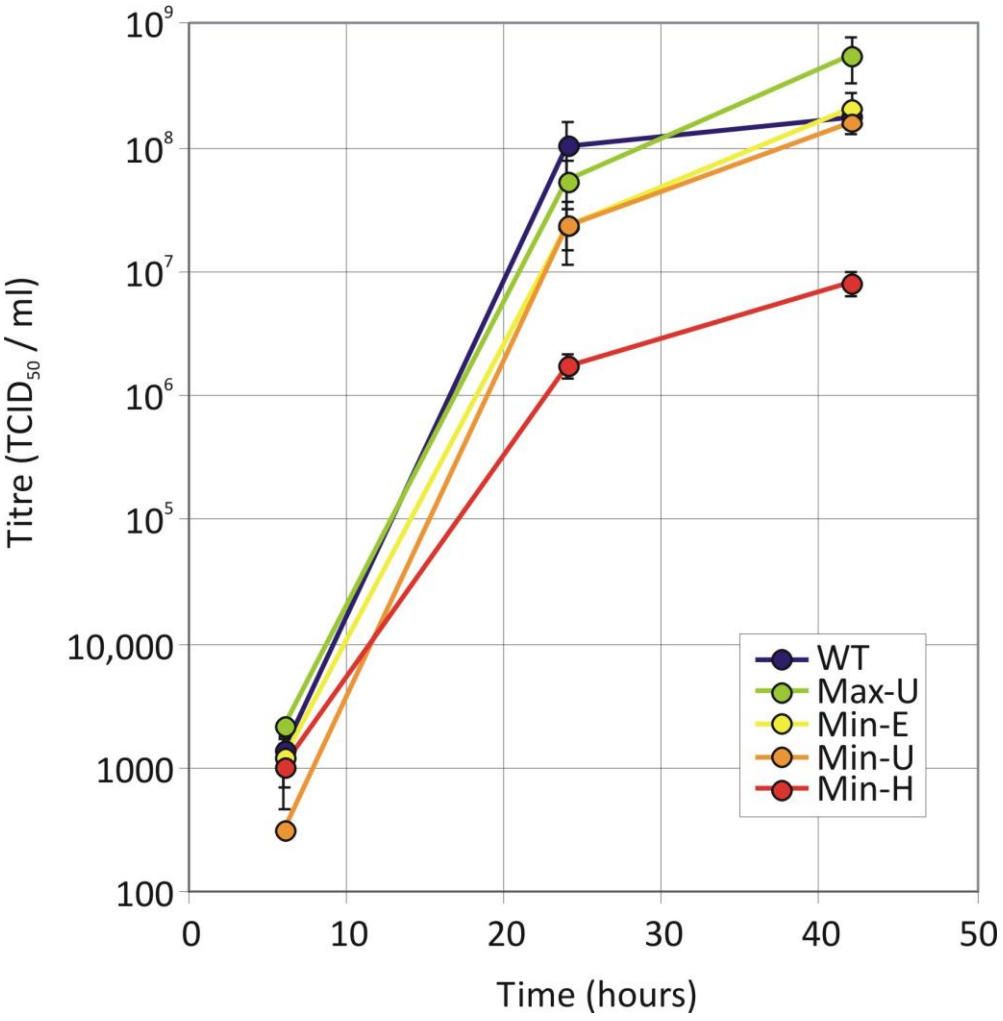


Figure 7.3. Replication kinetics of WT and mutant viruses in multiple-step growth curves. RD cell monolayers were infected at low MOI (0.01), and infectious titre of viral supernatants determined by $TCID_{50}$ assay at the time points indicated. Results are the standard error of the mean of three biological replicates.

(A)

1. WT vs P|P
2. WT vs cu|cu
3. P|P vs cu|cu
4. WT vs Max-U
5. Max-U vs P|P
6. Max-U vs cu|cu
7. WT vs Min-E
8. WT vs Min-U
9. Min-E vs Min-U
10. WT vs Min-H
11. Min-E vs Min-H
12. Min-U vs Min-H

(B)

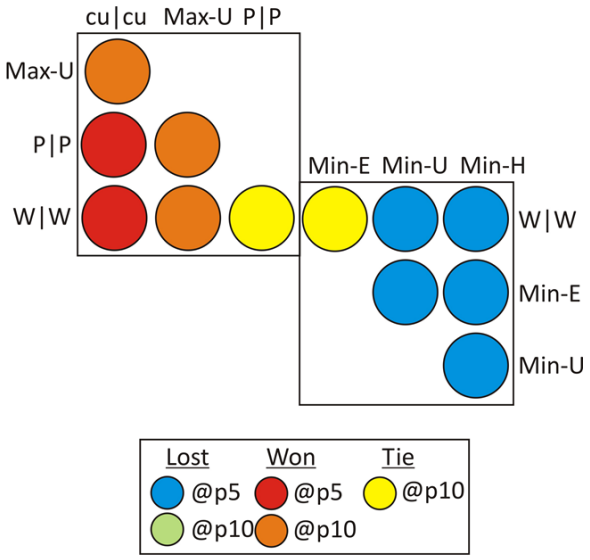


Figure 7.4. Pairwise combinations investigated during competition experiments. (A) Twelve combinations of mutant and WT viruses were examined. (B) Summary of the outcomes of pairwise fitness comparisons for viruses tested. Results indicated by colour shading at passages 5 and 10. For example, Min-E and WT showed equal fitness (yellow shading) and cu|cu outcompeted WT by passage 5 (red) and Max-U by passage 10. W|W = WT (image (B) generated by Prof. P. Simmonds).

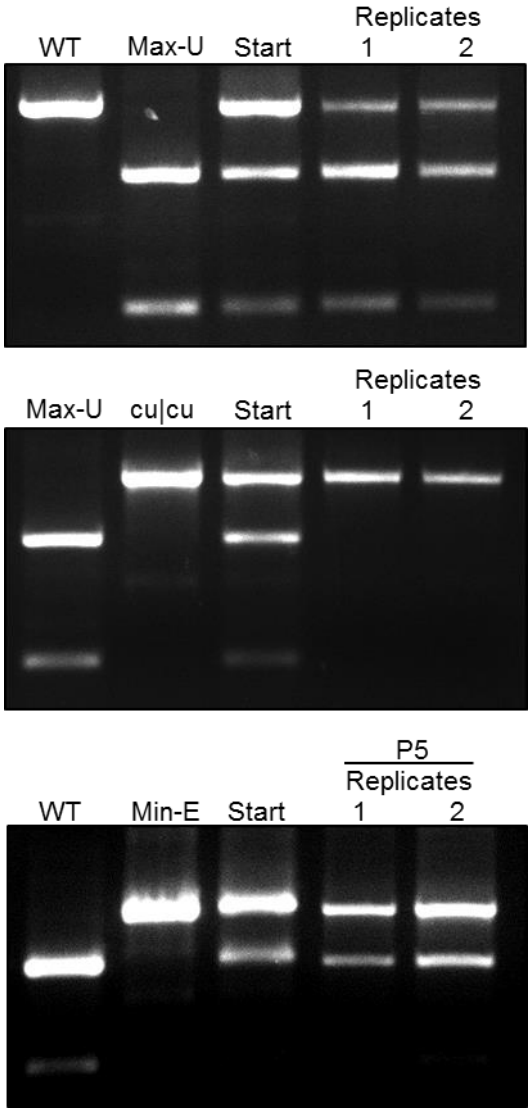


Figure 7.5. Fitness determination of WT and mutant viruses by competition assays. RD cells were co-infected with an equal MOI of WT or mutant viruses and supernatants serially passaged in cells following development of CPE. RNA was isolated from viral supernatant and virus composition verified by selective restriction enzyme digest. Images shown are examples of three pairwise combinations. The virus composition of the starting inoculum (lane 3) and two biological replicates (lanes 4 and 5) at passage 10 (panels 1 and 2) or P5 (panel 3) of WT, Max-U, cu|cu and Min-E starting viruses (lanes 1 and 2) are shown.

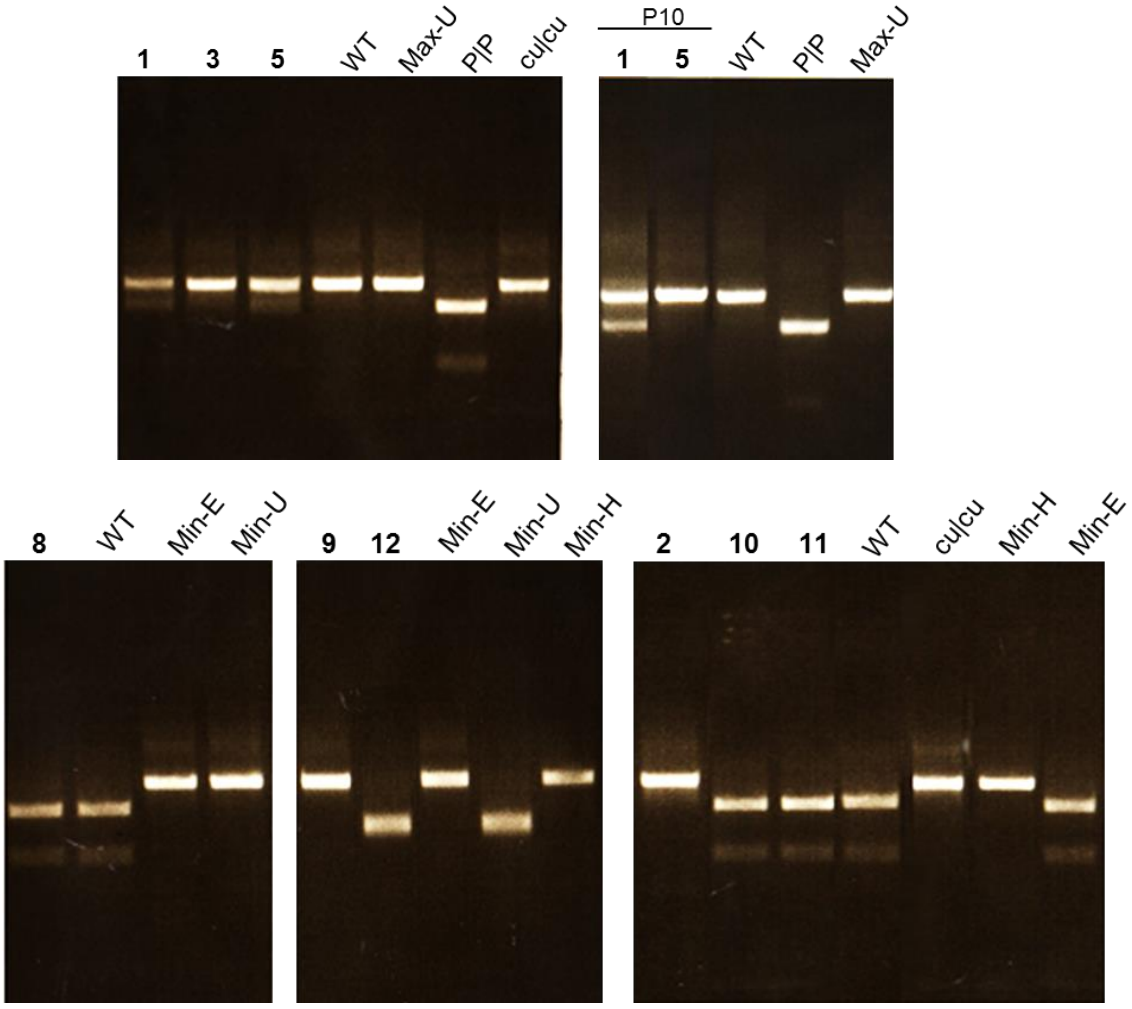


Figure 7.6. Fitness ranking of remaining competition assays examined. Images shown are the results from each pairwise combination, indicated by a number (details listed in Figure 7.4A), and starting viruses. All examples were completed by P5 unless indicated.

7.5 Translation efficiency

The translation profiles of each mutant virus were compared to examine whether altering dinucleotide frequency or CPB score had an effect on the rate of translation, and therefore influenced replication. *In vitro* translation assays were used to avoid any effects mediated through cellular stress-response-related RNA mechanisms that may restrict E7 replication (Atkinson *et al.*, 2014). Rabbit reticulocyte lysates were programmed with transcript RNA (2 µg: transcribed from WT and mutant cDNA), in the presence of [^{35}S]-methionine. Reactions were incubated for 3 hr and subsequently resolved on SDS-PAGE gradient gels. Translation profiles were as expected, with cleaved and partially cleaved proteins observed (Figure 7.7). Translation efficiencies of each mutant in comparison to WT were comparable. Any differences observed did not, therefore, correlate with replicative fitness. This indicated that alteration of CPB or dinucleotide frequencies had no significant effect on viral translation, and therefore cannot be attributed to the differences observed in the replication phenotype documented.

7.6 Conclusions

- CpG and UpA dinucleotide frequencies were primarily responsible for differences observed in viral replication kinetics.
- No fitness differences were observed for mutants with different CPB scores, where dinucleotide frequencies were kept constant.
- Translation efficiencies, however, were not affected by either CPB or dinucleotide frequencies.
- Insight of the mechanism underlying attenuation is crucial for the design and safety of future live attenuated vaccines.

NOTE: The data presented in this chapter were generated by myself, unless specifically stated otherwise. These data made the major contribution towards the publication (Tulloch *et al.*, 2014b).

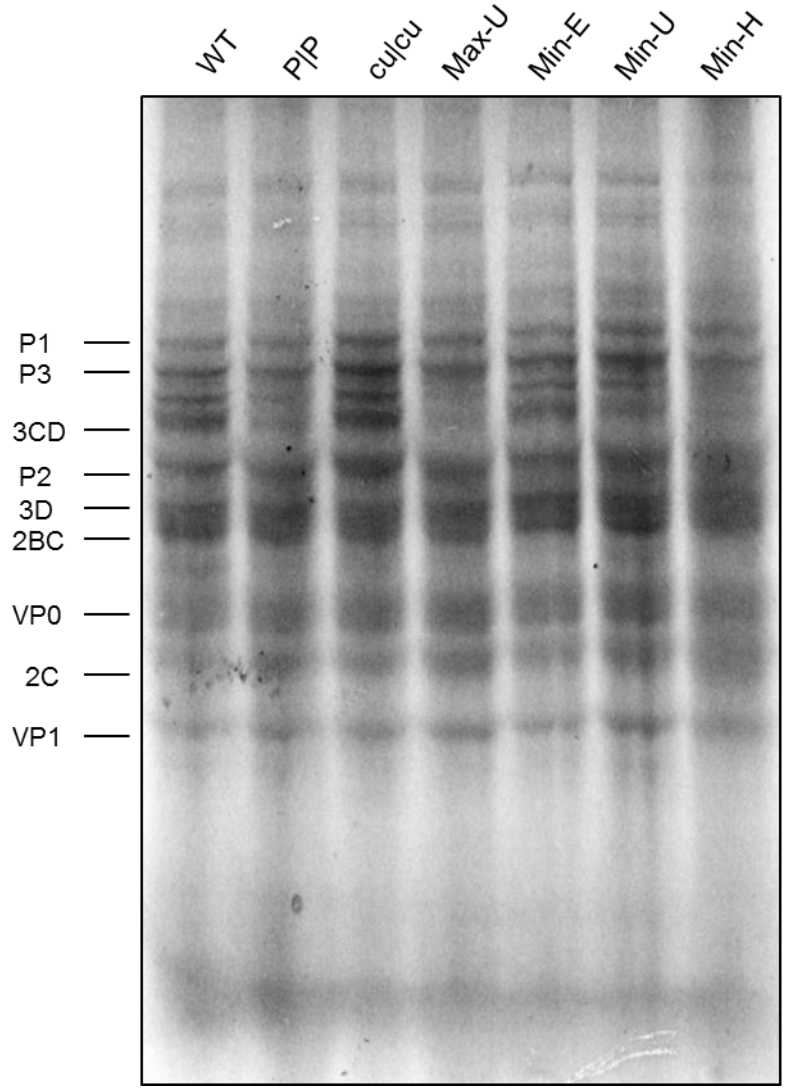


Figure 7.7. *In-vitro* translation profiles of WT and mutant viruses. Rabbit reticulocytes were programmed with 2 µg of transcript RNA generated from WT and mutant cDNA templates. Radiolabelled products were separated by SDS_PAGE. Predicted viral protein positions were calculated based on molecular weights on SDS-PAGE.

7.7 Discussion

7.7.1 Increasing dinucleotide frequency and not CPB generates an attenuated replication phenotype

Studies which have adopted codon-pair de-optimisation as a strategy for the production of live-attenuated vaccines have failed to recognise the effects of this procedure has on dinucleotide frequencies (Coleman *et al.*, 2008; Mueller *et al.*, 2006, 2010). Selection of disfavoured codon-pairs elevated UpA or CpG dinucleotide frequencies to levels 2.5 to 3-fold higher than native sequences (Table 7.2). The frequencies of such dinucleotides are suppressed in most mammalian RNA viruses (Rima & Mcferran, 1997) and have been documented in other studies to contribute towards the observed attenuation (Burns *et al.*, 2009). The mutants constructed within this study provided the influences of CPB and dinucleotides on viral phenotype and fitness to be analysed separately. The fitness ranking observed following competition assay analysis was as follows;

(cu|cu > Max > (WT = P|P = Min-E) > Min-U > Min-H)

This demonstrated that it was the increase in dinucleotide frequencies which contributed to the fitness phenotype observed - and not CPB.

7.7.2 Translation efficiency of E7 and mutant viruses and mechanism of attenuation

It was proposed that reduced viral replication from CPB de-optimised mutants was caused by changes in the translation rate (Coleman *et al.*, 2008; Mueller *et al.*, 2006, 2010). If CPB was responsible for such changes then differences between the translation profiles of WT and CPB-maximised or CPB-minimised mutants would have been observed: no measurable differences were observed either between these different mutants, or, the mutants and WT genomes- as analysed by *in vitro* translation assays. This demonstrates that it is not the virus itself that is intrinsically defective as a template for translation *per se*, but that the modified nature of these viral genomes is now recognised by the cell and inhibiting replication. The mechanism of attenuation remains unknown: it was shown, however, that inhibition of replication of elevated CpG/UpA genomic

forms was not mediated through conventional pattern recognition receptors or PKR (Atkinson *et al.*, 2014). These authors also demonstrated that high and low CpG/UpA mutated genomes were not sensitive to the cellular IFN response. Viruses were, however, sensitive to the addition of the kinase inhibitor C16 (*supposedly* PKR specific; Jammi *et al.*, 2003), with the replication phenotype reversed following treatment. This suggests regulation could occur through an unknown PKR-related component of the stress response pathway. Thus, investigation into the innate cellular responses produced against viruses with altered dinucleotide frequencies is essential for the safe production of these new high CpG/UpA live-attenuated vaccines.

7.7.3 Application to the FMDV replicon system

The regions modified within the E7 genome spanned both the P1 and P3 sequences. The capsid coding region will not be manipulated during studies using the FMDV replicon due to (i) HSE guidelines, and (ii) the strategy proposed for the generation of live-attenuated vaccines during this project. Here we aim to produce an attenuated (replication proteins) 'backbone' such that capsid proteins from a range of serotypes can be inserted into the attenuated backbone cDNA clone – producing either strain or serotype 'chimeric' vaccine strains. This would allow cross-protection across serotypes utilising a single strategy for attenuation which could be applied to all FMDV strains – once the sequence encoding the capsid proteins was determined. Replicons which are attenuated using the mechanism of increased dinucleotide frequency will be compared to a panel of known, published, attenuated FMDV replicons to monitor levels of replication between these genomes.

The E7 work determined the (genomic) basis of attenuation (by whatever cellular mechanism) and now provides a rational basis for interpretation of CPD forms that had already been generated and, critically, for further design of CPD FMDV genomes.

Chapter 8: Concluding Remarks.

This thesis has primarily focused on the design and optimisation of an efficient replicon system for FMDV, which can be used to study replication outwith high disease-secure facilities. The use of the system to screen for attenuated genomes followed by the creation of corresponding infectious copies (by collaborators) is of high priority. Working within high bio-containment facilities such as the Pirbright Institute involves laborious health-and-safety procedures which are very important for safety reasons, however, they lengthen the time taken to execute experiments, especially those involving virus. A replicon system which removes the need for such facilities enables experiments to be performed quicker using techniques which differ from classical methods. This allows for screening of potential areas of interest before commencing time-consuming, but essential, virus work. Screening of attenuated genomes at St Andrews can help to 'rank' replicons in order of interest/importance, reducing the work load involved with viral growth curves and aiding in the quick progression to animal studies.

During experiments to analyse attenuated replicon genomes, it became apparent that the spacer region of L^{pro} could be of some interest for future investigation. Replicons encoding the 84 nt spacer sequence did not demonstrate an initial lag in replication, in contrast to replicons completely lacking L^{pro}, but did exhibit a decrease in fluorescent signal when compared to WT replicons containing a fully functional L^{pro}. This was observed for replicons coding for both mCherry (Chapter 5), and for Pt.GFP (data not shown). Interestingly, insertion of this 84 nt spacer region into a plasmid encoding Aq.GFP and mCherry FPs (such that the 28 aa of the spacer sequence was fused onto the N-terminus of Aq.GFP) produced a perinuclear distribution of fluorescence within transfected BHK-21 cells. In contrast, plasmids encoding Lab^{pro} (able to produce both forms of L) or Lb^{pro} displayed a more uniform distribution of Aq.GFP within the cytoplasm. These interesting observations suggest the spacer region may play a role in the sub-cellular localisation of the Lab, but not the Lb, form of the L proteinase. The two initiation codons within L^{pro} are conserved within all sequenced isolates of FMDV across all seven serotypes (Carrillo *et al.*, 2005; Sangar *et al.*, 1987). The spacer region does not show a high sequence similarity between isolates, unlike the Lb region which is more conserved. It has been suggested that the spacer serves as

an extension which modifies the activity of Lb^{pro} (Sangar *et al.*, 1987), but as yet there is no known alternative function for either forms.

Replicons completely lacking L^{pro} were attenuated (following insertion of a FP which was not cleaved by L) and this finding also merits additional investigation, as this is in disagreement with the work of others (Belsham, 2013; Piccone *et al.*, 1995). As mentioned in the discussion, these laboratories have been unable to rescue viruses lacking L^{pro}: only viruses containing the spacer region of L^{pro} could be rescued.

Thus, the data described within this thesis pose interesting questions regarding the function of the differing forms of L proteinase. Further experiments will involve Deltavision microscopy of cells transfected with replicons/plasmids encoding different forms of L^{pro} and determination of their localisation with the use of anti-L^{pro} antibodies. Hopefully this will help to elucidate the possible function of these forms within cells. Experiments to abolish RNA secondary structure using SHAPE analysis are also planned to confirm whether this plays a role in the function of L^{pro}, as well as experiments investigating a protein-dependent function.

Introduction of ribozymes into polymerase deleted forms of the replicon lead to the rather unexpected observation of no fluorescence detected upon transfection into BHK-21 cells. A polymerase deletion which did not contain these ribozymes (pGFP-PAC-Δ3D) exhibited fluorescence, albeit lower than WT. It should be noted that a lack of fluorescence was observed for all polymerase mutants generated (encoding all FPs examined) during this study. These data suggest that RNA may be sequestered into abortive replication complexes when a precise 5'- end is generated – or at least a complex which supresses translation. If this were the case, however, these data suggest that $3D^{pol}$ is not involved in the suppression of translation, which has been proposed by others to function as the switch from translation to replication (Herold & Andino, 2001; Paul *et al.*, 2000). Further deletions may help 'map' which protein is responsible for this effect. Perhaps when the 5'-end is not authentic, the RNA is predominantly translated, and following removal of non-viral nucleotides (or $3B^{VPg}$ during infection) triggers replication.

The molecular biology of FMDV replication is largely unexplained. Limited studies have been performed and many analogies have to made with the enteroviruses. The findings described within this thesis may help to elucidate the fundamentals of RNA replication, or, at least help to gain more insight into this process. Strand-specific RT-qPCR could be used to compare levels of negative- and positive-

strand RNA produced by both polymerase deletion and WT replicons. This would require sequencing of isolated genomes to determine authenticity of the viral 5'-end.

The use of Pt.GFP within the FMDV replicon not only allows for GFP fluorescence to be measured without cleavage by L^{pro}, but will aid in studies which focus on persistence due to the increased fluorescent properties of this FP. FMDV readily establishes persistent infections, but very little is known about this process at the molecular level.

Experiments at St Andrews will consist of passaging RNA isolated from cells post-transfection. It is proposed that persistent cells will be expressing Pt.GFP at lower levels, and will therefore be easier to detect due to the intrinsic brightness exhibited by this GFP. Cells will be collected by FACS or *via* antibiotic selection, RNA extracted, and introduced into fresh cell monolayers. RNA isolated at each subsequent passage will be sequenced by colleagues at Pirbright and any changes observed will be investigated. Colleagues at Pirbright will also attempt to establish persistently infected cell cultures. Both sets of persistent cultures will be analysed using SILAC analysis (by colleagues at Dundee University) to determine the relative abundance/ subcellular localisation of viral proteins within these cells.

mCherry replicons generated during this work will assist in competition experiments between mutant replicons using the IncuCyte microscope. They will also aid in trans-complementation assays to 'rescue' replicons which have decreased rates of replication. Such experiments have been performed by colleagues at both Leeds University and the Pirbright Instuitue, where mutations/deletions within proteins have been complemented with mCherry replicons and replication has been improved. Virus work will supplement any analyses utilising competition or trans-complementation assays.

The observation that replicons encoding BEV $2A^{pro}$ were attenuated in comparison to WT replicons requires in-depth investigation, not only to rescue virus from corresponding infectious copies, but to elucidate regulation of FMDV 2A 'ribosome-skipping' in further detail. *In vitro* transcription/translation experiments on newly generated BEV replicons will identify whether the ratio of proteins up-and downstream of BEV $2A^{pro}$ is altered in comparison to replicons encoding FMDV 2A. Further examination will include western blotting with anti-eEF2 and phospho-specific anti-eEF2 antibodies. The loss of fluorescence

detected for some of the replicons analysed may have been due to incomplete processing at either of the junctions surrounding BEV $2A^{pro}$. To clarify this, extracts will be probed with anti-GFP and anti-FMDV antibodies, as well as anti-eIF4G to examine if BEV $2A^{pro}$ is active within all constructs generated. Polyprotein processing will be improved by passaging of virus/RNA, with sequencing at the C-terminus of Δ1D to observe any key changes which increase fitness. TaV 2A will be removed and various $3C^{pro}$ dipeptides will be inserted at the C-terminus of BEV 2A in an attempt to induce cleavage at this junction. Particle yields from corresponding infectious copies of BEV replicons will be determined and compared to IncuCyte and protein analysis performed at St Andrews. It is anticipated that these data will correlate and show that reduced protein synthesis from the BEV replicon lowers particle yield.
Studies into the innate immune system (comparison or replicons containing L^{pro} *vs* BEV $2A^{pro}$, mainly NF-κB due to published data; de los Santos *et al.*, 2006, 2007) and host-cell shut-off by the two viral proteinases will commence in the future once the avenues above have been investigated.

Confirmation that increasing dinucleotide frequencies, and <u>not</u> CPB, was the basis of attenuation will greatly facilitate the synthetic biology design process used to create the next generation of attenuated FMDV replicon genomes. Work has already been undertaken to create such replicons, and preliminary data has demonstrated that these genomes are indeed attenuated (experiments performed by Dr E. Minskaia). Work will commence shortly to rescue the corresponding infectious copies, expanding these findings.
Replicons encoding Pt.GFP displayed a stable decline in GFP fluorescence from 12 hr onwards when compared to replicons encoding Aq.GFP (Figure 5.8 *vs* Figure 4.3B) which could be explained by the composition of CpG/UpA dinucleotides between these replicons. It is possible that replicons encoding Aq.GFP have a higher frequency of these dinucleotides, and therefore increase cell death due to enhanced detection by an unknown pathway of the innate immune system; postulated previously as the underlying mechanism of attenuation (Atkinson *et al.*, 2014). The GFP signal generated by the pPt.GFP replicon decreases less rapidly than those encoding Aq.GFP, therefore rather than increased stability of Pt.GFP (which was proposed during Chapter 5), perhaps a reduced frequency of these dinucleotides allows for the viral RNA to escape detection by the immune system, allowing for increased replication and limited cell death. All relevant replicons generated during this thesis will be

analysed using the SSE software developed by Prof. P. Simmonds (Simmonds, 2012) to determine the frequency of CpG/UpA dinucleotides within each genome. Any major differences will be rectified using gene blocks with an 'optimised' sequence.

Hence, the data generated during the construction of this FMDV replicon system highlights the possibilities for the use of this approach and how it will allow us to investigate the fundamentals of FMDV RNA replication. Publications from other groups have resulted from the use of the replicon (Forrest *et al.*, 2014), with collaborators from various institutes using these constructs to study aspects of FMDV biology. Okapi Biosciences (Belgium) are also using this replicon system to screen for small molecule inhibitors of FMDV.

In conclusion, this system will help to produce a step-change in our understanding of how this virus grows and interacts in cells. The work carried out during this project will aid in the development of

- a new generation of more effective vaccines
- new, more biosecure, methods of vaccine production and
- more effective diagnosis.

References

Albariño, C. G., Uebelhoer, L. S., Vincent, J. P., Khristova, M. L., Chakrabarti, A. K., McElroy, A., *et al*. (2013). Development of a reverse genetics system to generate recombinant Marburg virus derived from a bat isolate. *Virology*, *446*, pp. 230–7.

Andino, R., Rieckhof, G. E., Achacoso, P. L., & Baltimore, D. (1993). Poliovirus RNA synthesis utilizes an RNP complex formed around the 5'-end of viral RNA. *The EMBO Journal*, *12*(9), pp. 3587–98.

Atkinson, N. J., Witteveldt, J., Evans, D. J., & Simmonds, P. (2014). The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic Acids Research*, *42*(7), pp. 4527–45.

Bartenschlager, R. (2006). Hepatitis C virus molecular clones: from cDNA to infectious virus particles in cell culture. *Current Opinion in Microbiology*, *9*(4), pp. 416–422.

Bartenschlager, R., & Lohmann, V. (2000). Replication of the hepatitis C virus. *Best Practice & Research Clinical Gastroenterology*, *14*(2), pp. 241–254.

Barton, D. J., Donnell, B. J. O., & Flanegan, J. B. (2001). 5 ' cloverleaf in poliovirus RNA is a cis -acting replication element required for negative-strand synthesis. *The EMBO Journal*, *20*(6), pp. 1439–1448.

Beard, C., & Mason, P. (2000). Genetic determinants of altered virulence of Taiwanese foot-and-mouth disease virus. *Journal of Virology*, *74*(2), pp. 1–6.

Belsham, G. (2005). Translation and replication of FMDV RNA. *Current Topics in Microbiology and Immunology*, *288*, pp. 43–70.

Belsham, G. J. (2013). Influence of the Leader protein coding region of foot-and-mouth disease virus on virus replication. *The Journal of General Virology*, *94*(7), pp. 1486–95.

Belsham, G., McInerney, G., & Ross-Smith, N. (2000). Foot-and-mouth disease virus 3C protease induces cleavage of translation initiation factors eIF4A and eIF4G within infected cells. *Journal of Virology*, *74*(1), pp. 272–280.

Bienz, K., Egger, D., Troxler, M., & Pasamontest, L. (1990). Structural Organization of Poliovirus RNA Replication Is Mediated by Viral Proteins of the P2 Genomic Region. *Journal of Virology*, *64*(3), pp. 1156–1163.

Blattner, F. (2014). SeqMan Pro Sanger Sequencing Software. Retrieved from http://www.dnastar.com.

Blight, K. J., & Norgard, E. A. (1996). HCV Replicon Systems, in: S. Tan (Ed.), *Hepatitis C Viruses: Genomes and Molecular Biology*, pp. 311–351. Norfolk, UK: Horizon Bioscience.

Boshart, M., Weber, F., Jahn, G., Dorschhler, K., Fleckenstein, B., & Schaffner, W. (1985). A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell*, *41*(2), pp. 521–530.

Brown, C., Piccone, M., Mason, P., McKenna, T., & Grubman, M. (1996). Pathogenesis of wild-type and leaderless foot-and-mouth disease virus in cattle. *Journal of Virology*, *70*(8), pp. 5638–5641.

Brown, D. M., Kauder, S. E., Cornell, C. T., Jang, G. M., Racaniello, V. R., & Semler, B. L. (2004). Cell-Dependent Role for the Poliovirus 3 ' Noncoding Region in Positive-Strand RNA Synthesis. *Journal of Virology*, *78*(3), pp. 1344–1351.

Brown, J. D., & Ryan, M. D. (2010). Ribosome "Skipping": "Stop-Carry On" or "StopGo" Translation, in: J. Atkins & R. Gesteland (Eds.), *Recoding: Expansion of Decoding Rules Enriches Gene Expression*, pp. 101–122. New York: Springer.

Burman, A., Clark, S., Abrescia, N. G., Fry, E. E., Stuart, D. I., & Jackson, T. (2006). Specificity of the VP1 GH Loop of Foot-and-Mouth Disease Virus for alpha-v integrins. *Journal of Virology*, *80*(19), pp. 9798–810.

Burns, C. C., Campagnoli, R., Shaw, J., Vincent, A., Jorba, J., & Kew, O. (2009). Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. *Journal of Virology*, *83*(19), pp. 9957–69.

Burns, C. C., Shaw, J., Campagnoli, R., Vincent, A., Quay, J., & Kew, O. (2006). Modulation of Poliovirus Replicative Fitness in HeLa Cells by Deoptimization of Synonymous Codon Usage in the Capsid Region. *Journal of Virology*, *80*(7), pp. 3259–3272.

Cao, X., Bergmann, I., Füllkrug, R., & Beck, E. (1995). Functional analysis of the two alternative translation initiation sites of foot-and-mouth disease virus. *Journal of Virology*, *69*(1), pp. 560–563.

Carrillo, C., Tulman, E., Delhon, G., Lu, Z., Carreno, A., Vagnozzi, A., *et al*. (2005). Comparative genomics of foot-and-mouth disease virus. *Journal of Virology*, *79*(10), pp. 6487–6504.

Carter, A., Morris, C., & McAllister, W. (1981). Revised transcription map of the late region of bacteriophage T7 DNA. *Journal of Virology*, *37*(2), pp. 636–642.

Chalfie, M. (1995). Green fluorescent protein. *Photochemistry and Photobiology*, *62*(4), pp. 5–10.

Chang, Y., Zheng, H., Shang, Y., Jin, Y., Wang, G., Shen, X., *et al*. (2009). Recovery of infectious foot-and-mouth disease virus from full-length genomic cDNA clones using an RNA polymerase I system. *Acta Biochimica et Biophysica Sinica*, *41*(12), pp. 998–1007.

Chinsangaram, J., Mason, P. W., & Grubman, M. J. (1998). Protection of swine by live and inactivated vaccines prepared from a leader proteinase-deficient serotype A12 foot-and-mouth disease virus. *Vaccine*, *16*(16), pp. 1516–1522.

Chudakov, D. M., Matz, M. V, Lukyanov, S., & Lukyanov, K. A. (2010). Fluorescent proteins and their applications in imaging living cells and tissues. *Physiological Reviews*, *90*(3), pp. 1103–63.

Clark, J. M. (1988). Novel non-templated nucleotide addition reactions catalyzed by prokaryotic and eukaryotic DNA polymerases. *Nucleic Acids Research*, *16*(9), pp. 9677–9686.

Clarke, B. E., Brown, A. L., Currey, K. M., Newton, S. E., Rowlands, D. J., & Carroll, A. R. (1987). Potential secondary and tertiary structure in the genomic RNA of foot and mouth disease virus. *Nucleic Acids Research*, *15*(17), pp. 7067–79.

Cohen, J. I., Ticehurst, J. R., Feinstone, S. M., Rosenblum, B., & Purcell, R. H. (1987). Hepatitis A Virus cDNA and its RNA Transcripts are Infectious in Cell Culture. *Journal of Virology*, *61*(10), pp. 3035–3039.

Coleman, J. R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E., & Mueller, S. (2008). Virus attenuation by genome-scale changes in codon pair bias. *Science*, *320*(5884), pp. 1784–7.

Corish, P., & Tyler-Smith, C. (1999). Attenuation of green fluorescent protein half-life in mammalian cells. *Protein Engineering Design and Selection*, *12*(12), pp. 1035–1040.

Crawford, N. M., & Baltimore, D. (1983). Genome-linked protein VPg of poliovirus is present as free VPg and VPg-pUpU in poliovirus-infected cells. *Proceedings of the National Academy of Sciences of the United States of America*, *80*, pp. 7452–7455.

Curry, S., Fry, E., Blakemore, W., Abu-ghazaleh, R., Jackson, T., King, A., *et al*. (1997). Dissecting the Roles of VP0 Cleavage and RNA Packaging in Picornavirus Capsid Stabilization : The Structure of Empty Capsids of Foot-and-Mouth Disease Virus. *Journal of Virology*, *71*(12), pp. 9743–9752.

de los Santos, T., de Avila Botton, S., Weiblen, R., & Grubman, M. J. (2006). The Leader Proteinase of Foot-and-Mouth Disease Virus Inhibits the Induction of Beta Interferon mRNA and Blocks the Host Innate Immune Response. *Journal of Virology*, *80*(4), pp. 1906–1914.

de los Santos, T., Diaz-San Segundo, F., & Grubman, M. J. (2007). Degradation of nuclear factor kappa B during foot-and-mouth disease virus infection. *Journal of Virology*, *81*(23), pp. 12803–15.

de los Santos, T., Segundo, F. D., Zhu, J., Koster, M., Dias, C. C., & Grubman, M. J. (2009). A conserved domain in the leader proteinase of foot-and-mouth disease virus is required for proper subcellular localization and function. *Journal of Virology*, *83*(4), pp. 1800–10.

Devaney, M. A., Vakharia, V. N., Lloyd, R. E., Ehrenfeld, E., & Grubman, M. J. (1988). Leader protein of foot-and-mouth disease virus is required for cleavage of the p220 component of the cap-binding protein complex. *Journal of Virology*, *62*(11), pp. 4407–9.

Doel, T. (2003). FMD vaccines. *Virus Research*, *91*, pp. 81–99.

Dong-Sheng, H., Kang-Ning, L., Xiao-Min, L., Shao-Rong, L., Dan-Ping, S., & Ming, L. (2011). Genomic comparison of foot-and-mouth disease virus R strain and its chick-passaged attenuated strain. *Veterinary Microbiology*, *150*(1-2), pp. 185–90.

Donnelly, M., Hughes, L., Luke, G., Mendoza, H., ten Dam, E., Gani, D., *et al*. (2001a). The “cleavage”activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring “2A-like”sequences. *Journal of General Virology*, *82*, pp. 1027–1041.

Donnelly, M. L. L., Luke, G., Mehrotra, A., Li, X., Hughes, L. E., Gani, D., *et al*. (2001b). Analysis of the aphthovirus 2A/2B polyprotein “ cleavage ” mechanism indicates not a proteolytic reaction, but a novel translational effect : a putative ribosomal “ skip .” *Journal of General Virology*, *82*, pp. 1013–1025.

Doronina, V. a, Wu, C., de Felipe, P., Sachs, M. S., Ryan, M. D., & Brown, J. D. (2008). Site-specific release of nascent chains from ribosomes at a sense codon. *Molecular and Cellular Biology*, *28*(13), pp. 4227–39.

Du, L., Villarreal, S., & Forster, A. C. (2012). Multigene expression in vivo: supremacy of large versus small terminators for T7 RNA polymerase. *Biotechnology and Bioengineering*, *109*(4), pp. 1043–50.

Dubensky, T., Driver, D., Polo, J., Belli, B., Latham, E., Ibanez, C., *et al*. (1996). Sindbis virus DNA-based expression vectors: utility for in vitro and in vivo gene transfer. *Journal of Virology*, *70*(1), pp. 508–519.

Duke, G. M., Osorio, J. E., & Palmenberg, A. C. (1990). Attenuation of Mengo virus through genetic engineering of the 5′ noncoding poly (C) tract. *Nature*, *343*, pp. 474–476.

Duke, G. M., & Palmenberg, A. C. (1989). Cloning and synthesis of infectious cardiovirus RNAs containing short, discrete poly(C) tracts. *Journal of Virology*, *63*(4), pp. 1822–6.

Ellard, F. M., Drew, J., Blakemore, W. E., Stuart, D. I., & King, A. M. (1999). Evidence for the role of His-142 of protein 1C in the acid-induced disassembly of foot-and-mouth disease virus capsids. *The Journal of General Virology*, *80 (Pt 8)*, pp. 1911–8.

Ertel, K. J., Brunner, J. E., & Semler, B. L. (2010). Mechanistic consequences of hnRNP C binding to both RNA termini of poliovirus negative-strand RNA intermediates. *Journal of Virology*, *84*(9), pp. 4229–42.

Escarmís, C., Toja, M., Medina, M., & Domingo, E. (1992). Modifications of the 5’ untranslated region of foot-and-mouth disease virus after prolonged persistence in cell culture. *Virus Research*, *26*(2), pp. 113–25.

Falk, M., Grigera, P. R., Bergmann, I. E., Zibert, A., Multhaup, G., & Beck, E. (1990). Foot-and-Mouth Disease Virus Protease 3C Induces Specific Proteolytic Cleavage of Host Cell Histone H3. *Journal of Virology*, *64*(2), pp. 748–756.

Falk, M. M., Sobrino, F., & Beck, E. (1992). VPg Gene Amplification Correlates with Infective Particle Formation in Foot-and-Mouth Disease Virus. *Journal of Virology*, *66*(4), pp. 2251–2260.

Forrest, S., Lear, Z., Herod, M. R., Ryan, M. D., Rowlands, D. J., & Stonehouse, N. J. (2014). Inhibition of the FMDV sub-genomic replicon by RNA aptamers. *The Journal of General Virology*.

Forss, S., & Schaller, H. (1982). A tandem repeat gene in a picornavirus. *Nucleic Acids Research*, *10*(20), pp. 6441–6450.

Forss, S., Strebel, K., Beck, E., & Schaller, H. (1984). Nucleotide sequence and genome organization of foot-and-mouth disease virus. *Nucleic Acids Research*, *12*(16), pp. 6587–6601.

Fox, G., Parry, N., Barnett, P., McGinn, B., Dowlands, D., & Brown, F. (1989). The cell attachment site on foot-and-mouth disease virus includes the amino acid sequence RGD (arginine-glycine-aspartic acid). *Journal of General Virology*, *70*, pp. 625–637.

Fry, E., & Stuart, D. (2010). Virion Structure, in: E. Ehrenfeld, E. Domingo, & R. Roos (Eds.), *The Picornaviruses*, pp. 59–71. Washington, DC: ASM Press.

Gamarnik, A. V, & Andino, R. (2000). Interactions of Viral Protein 3CD and Poly (rC) Binding Protein with the 5 ' Untranslated Region of the Poliovirus Genome. *Journal of Virology*, *74*(5), pp. 2219–2226.

Gamarnik, A. V., & Andino, R. (1998). Switch from translation to RNA replication in a positive-stranded RNA virus. *Genes & Development*, *12*(15), pp. 2293–2304.

Ghanem, A., Kern, A., & Conzelmann, K. (2012). Significantly improved rescue of rabies virus from cDNA plasmids. *European Journal of Cell Biology*, *91*(1), pp. 10–6.

Gingold, H., & Pilpel, Y. (2011). Determinants of translation efficiency and accuracy. *Molecular Systems Biology*, *7*(481), p. 481.

Giomi, M., Bergmann, I., Scodeller, E., Auge de Mello, P., Gomez, I., & La Torre, J. (1984). Heterogeneity of the polyribocytidylic acid tract in aphthovirus: biochemical and biological studies of viruses carrying polyribocytidylic acid tracts of different lengths. *Journal of Virology*, *51*(3), pp. 799–805.

Glaser, W., & Skern, T. (2000). Extremely efficient cleavage of eIF4G by picornaviral proteinases L and 2A in vitro. *FEBS Letters*, *480*(2-3), pp. 151–155.

Gonda, D. K., Bachmair, A., Wunning, I., Tobias, J. W., Lane, W. S., & Varshavsky, A. (1989). Universality and Structure of the N-end Rule. *Journal of Biological Chemistry*, *264*(28), pp. 16700–16712.

Goodfellow, I., Chaudhry, Y., Richardson, A., Meredith, J., Almond, J. W., Barclay, W., *et al*. (2000). Identification of a cis-Acting Replication Element within the Poliovirus Coding Region. *Journal of Virology*, *74*(10), pp. 4590–4600.

Gorbalenya, A. E., & Koonin, E. V. (1989). Viral proteins containing the purine NTP-binding sequence pattern. *Nucleic Acids Research*, *17*(21), pp. 8413–8440.

Grubman, M., & Baxt, B. (2004). Foot-and-mouth disease. *Clinical Microbiology Reviews*, *17*(2), pp. 465–493.

Grubman, M. J., Zellner, M., Bablanian, G., Mason, P. W., & Piccone, M. E. (1995). Identification of the active-site residues of the 3C proteinase of foot-and-mouth disease virus. *Virology*, *213*(2), pp. 581–9.

Grubman, M., Rodríguez, L., & de los Santos, T. (2010). Foot-and-mouth disease, in: E. Ehrenfeld, E. Domingo, & R. Roos (Eds.), *The Picornaviruses*, pp. 397–410. Washington, DC: ASM Press.

Gu, C., Zheng, C., Shi, L., Zhang, Q., Li, Y., Lu, B., *et al*. (2007). Plus- and minus-stranded foot-and-mouth disease virus RNA quantified simultaneously using a novel real-time RT-PCR. *Virus Genes*, *34*(3), pp. 289–98.

Harris, T., & Brown, F. (1977). Biochemical analysis of a virulent and an avirulent strain of foot-and-mouth disease virus. *Journal of General Virology*, *34*, pp. 87–105.

Haseloff, J., Siemering, K. R., Prasher, D. C., & Hodge, S. (1997). Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(6), pp. 2122–2127.

He, B., Kukarin, A., Temiakov, D., Chin-Bow, S. T., Lyakhov, D. L., Rong, M., *et al*. (1998). Characterization of an Unusual, Sequence-specific Termination Signal for T7 RNA Polymerase. *Journal of Biological Chemistry*, *273*(30), pp. 18802–18811.

Herold, J., & Andino, R. (2000). Poliovirus Requires a Precise 5 ′ End for Efficient Positive-Strand RNA Synthesis. *Journal of Virology*, *74*(14), pp. 6394–6400.

Herold, J., & Andino, R. (2001). Poliovirus RNA replication requires genome circularization through a protein–protein bridge. *Molecular Cell*, *7*, pp. 581–591.

Hoenen, T., Groseth, A., de Kok-Mercado, F., Kuhn, J. H., & Wahl-Jensen, V. (2011). Minigenomes, transcription and replication competent virus-like particles and beyond: Reverse genetics systems for filoviruses and other negative stranded hemorrhagic fever viruses. *Antiviral Research*, *91*(2), pp. 195–208.

Ikeda, R. A., & Richardson, C. C. (1986). Interactions of the RNA polymerase of bacteriophage T7 with its promoter during binding and initiation of transcription. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(11), pp. 3614–3618.

Imburgio, D., Rong, M., Ma, K., & McAllister, W. T. (2000). Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry*, *39*(34), pp. 10419–30.

Inouye, S., & Tsuji, F. I. (1994). Aequorea green fluorescent protein. *FEBS Letters*, *341*(2-3), pp. 277–280.

Israelsson, S., Sävneby, A., Ekström, J.-O., Jonsson, N., Edman, K., & Lindberg, A. M. (2014). Improved replication efficiency of echovirus 5 after transfection of colon cancer cells using an authentic 5' RNA genome end methodology. *Investigational New Drugs*, *32*(6), pp. 1063–70.

Jamal, S. M., & Belsham, G. J. (2013). Foot-and-mouth disease: past, present and future. *Veterinary Research*, *44*(116), pp. 1–14.

Jammi, N. V., Whitby, L. R., & Beal, P. a. (2003). Small molecule inhibitors of the RNA-dependent protein kinase. *Biochemical and Biophysical Research Communications*, *308*(1), pp. 50–57.

Jeng, S., Gardner, J., & Gumport, R. (1990). Transcription termination by bacteriophage T7 RNA polymerase at rho-independent terminators. *Journal of Biological Chemistry*, *265*(7), pp. 3823–3830.

Jeng, S., Gardnerq, J., & Gumport, R. (1992). Transcription Termination in Vitro by Bacteriophage T7 RNA Polymerase. *Journal of Biological Chemistry*, *267*(27), pp. 19306–19312.

Juleff, N., Windsor, M., Reid, E., Seago, J., Zhang, Z., Monaghan, P., *et al*. (2008). Foot-and-mouth disease virus persists in the light zone of germinal centres. *PloS One*, *3*(10): e3434.

Kajigaya, S., Arakawa, H., Kuge, S., Koi, T., Imura, N., & Nomoto, A. (1985). Isolation and characterization of defective-interfering particles of poliovirus Sabin 1 strain. *Virology*, *142*(2), pp. 307–16.

Kandolf, R., & Hofschneider, P. H. (1985). Molecular cloning of the genome of a cardiotropic Coxsackie B3 virus: full-length reverse-transcribed recombinant cDNA generates infectious virus in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(14), pp. 4818–22.

Kaplan, G., Lubinski, J., Dasgupta, A., & Racaniello, V. R. (1985). In vitro synthesis of infectious poliovirus RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *82*(24), pp. 8424–8.

Kaplan, G., & Racaniello, V. R. (1988). Construction and characterization of poliovirus subgenomic replicons. *Journal of Virology*, *62*(5), pp. 1687–96.

Kapoor, A., Victoria, J., Simmonds, P., Wang, C., Shafer, R. W., Nims, R., *et al*. (2008). A highly divergent picornavirus in a marine mammal. *Journal of Virology*, *82*(1), pp. 311–20.

Kaul, G., Pattan, G., & Rafeequi, T. (2011). Eukaryotic elongation factor-2 (eEF2): its regulation and peptide chain elongation. *Cell Biochemistry and Function*, *29*, pp. 227–234.

Kelley, L., & Sternberg, M. (2009). Protein structure prediction on the web: a case study using the Phyre server. , pp. 363–371.

Khromykh, A. A., Kondratieva, N., Sgro, J.-Y., Palmenberg, A., & Westaway, E. G. (2003). Significance in replication of the terminal nucleotides of the flavivirus genome. *Journal of Virology*, *77*(19), pp. 10623–10629.

Kirchweger, R., Ziegler, E., Lamphear, B. J., Waters, D., Liebig, H. D., Sommergruber, W., *et al*. (1994). Foot-and-mouth disease virus leader proteinase: purification of the Lb form and determination of its cleavage site on eIF-4 gamma. *Journal of Virology*, *68*(9), pp. 5677–84.

Kirkegaard, K., & Semler, B. (2010). Genome Replication II: the Process, in: E. Ehrenfeld, E. Domingo, & R. Roos (Eds.), *The Picornaviruses*, pp. 127–139. Washington, DC: ASM Press.

Klump, W. M., Bergmann, I., Müller, B. C., Ameis, D., & Kandolf, R. (1990). Complete nucleotide sequence of infectious Coxsackievirus B3 cDNA: two initial 5' uridine residues are regained during plus-strand RNA synthesis. *Journal of Virology*, *64*(4), pp. 1573–83.

Knipe, T., Rieder, E., Baxt, B., Ward, G., & Mason, P. W. (1997). Characterization of Synthetic Foot-and-Mouth Disease Virus Provirions Separates Acid-Mediated Disassembly from Infectivity. *Journal of Virology*, *71*(4), pp. 2851–2856.

Knowles, N. (2013). Creation of new species within the Picornaviridae. *Picornavirus Study Group*. Retrieved from http://www.picornastudygroup.com.

Knowles, N. (2014). Aphthovirus. *Picornaviridae*. Retrieved from http://www.picornaviridae.com/aphthovirus/aphthovirus.htm

Knowles, N., Davies, P., & Henry, T. (2001). Emergence in Asia of foot-and-mouth disease viruses with altered host range: characterization of alterations in the 3A protein. *Journal of Virology*, *75*(3), pp. 1551–1556.

Knowles, N., Hovi, T., King, A. M. Q., & Stanway, G. (2010). Overview of Taxonomy, in: E. Ehrenfeld, E. Domingo, & R. Roos (Eds.), *The Picornaviruses*, pp. 19–32. Washington, DC: ASM Press.

Kuge, S., Saito, I., & Nomoto, A. (1986). Primary structure of poliovirus defective-interfering particle genomes and possible generation mechanisms of the particles. *Journal of Molecular Biology*, *192*(3), pp. 473–487.

Kühn, R., Luz, N., & Beck, E. (1990). Functional analysis of the internal translation initiation site of foot-and-mouth disease virus. *Journal of Virology*, *64*(10), pp. 4625–4631.

Lagadic-Gossmann, D., Huc, L., & Lecureur, V. (2004). Alterations of intracellular pH homeostasis in apoptosis: origins and roles. *Cell Death and Differentiation*, *11*(9), pp. 953–61.

Lamphear, B., Yan, R., Yang, F., Waters, D., Liebig, H., Klump, H., *et al*. (1993). Mapping the Cleavage Site in Protein Synthesis Initiation Factor eIF-4G of the 2A Protease from Human Coxsackievirus and Rhinovirus. *Journal of Biological Chemistry*, *268*(26), pp. 19200–19203.

Lawrence, P., & Rieder, E. (2009). Identification of RNA helicase A as a new host factor in the replication cycle of foot-and-mouth disease virus. *Journal of Virology*, *83*(21), pp. 11356–66.

Lazouskaya, N. V, Palombo, E. A., Poh, C., & Barton, P. A. (2014). Construction of an infectious cDNA clone of Enterovirus 71: insights into the factors ensuring experimental success. *Journal of Virological Methods*, *197*, pp. 67–76.

Levy, H., Bostina, M., Filman, D., & Hogle, J. (2010). Cell Entry: a Biochemical and Structural Perspective, in: E. Ehrenfeld, E. Domingo, & R. Roos (Eds.), *The Picornaviruses*, pp. 87–103. Washington, DC: ASM Press.

Li, X., Zhang, G., Ngo, N., Zhao, X., Kain, S. R., & Huang, C.-C. (1997). Deletions of the Aequorea victoria Green Fluorescent Protein Define the Minimal Domain Required for Fluorescence. *Journal of Biological Chemistry*, *272*(45), pp. 28545–28549.

Li, X., Zhao, X., Fang, Y., Jiang, X., Duong, T., Fan, C., *et al*. (1998). Generation of Destabilized Green Fluorescent Protein as a Transcription Reporter. *Journal of Biological Chemistry*, *273*(52), pp. 34970–34975.

Liljeström, P., Lusa, S., Huylebroeck, D., & Garoff, H. (1991). In vitro mutagenesis of a full-length cDNA clone of Semliki Forest virus: the small 6,000-molecular-weight membrane protein modulates virus release. *Journal of Virology*, *65*(8), pp. 4107–4113.

Lin, J.-Y., Chen, T.-C., Weng, K.-F., Chang, S.-C., Chen, L.-L., & Shih, S.-R. (2009). Viral and host proteins involved in picornavirus life cycle. *Journal of Biomedical Science*, *16*, p. 103.

Liu, Z., Cashion, L., & Twu, J. (1997). A systematic comparison of relative promoter/enhancer activities in mammalian cell lines. *Analytical Biochemistry*, *152*(1989), pp. 150–152.

Loening, A. M., Fenn, T. D., & Gambhir, S. S. (2007). Crystal structures of the luciferase and green fluorescent protein from Renilla reniformis. *Journal of Molecular Biology*, *374*(4), pp. 1017–28.

Lohmann, V., Korner, F., Koch, J., Herian, U., Theilmann, L., & Bartenschlager, R. (1999). Replication of Subgenomic Hepatitis C Virus RNAs in a Hepatoma Cell Line. *Science*, *285*(5424), pp. 110–113.

Luke, G. A., Pathania, U. S., Tulloch, F., & Martin, D. (2013). Advances in Genetic Engineering & Biotechnology Codon Pair Bias and Viral Vaccine Design. *Advances in Genetic Engineering & Biotechnology*, *2*(1), pp. 1–2.

Lundquist, R. E., Sullivan, M., & Maizel, J. V. (1979). Characterization of a new isolate of poliovirus defective interfering particles. *Cell*, *18*(3), pp. 759–769.

Lynnon Biosoft. (2001). DNAMAN Bioinformatics Software. Retrieved from http://www.lynnon.com.

Manicassamy, B., Manicassamy, S., Belicha-Villanueva, A., Pisanelli, G., Pulendran, B., & García-Sastre, A. (2010). Analysis of in vivo dynamics of influenza virus infection in mice using a GFP reporter virus. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(25), pp. 11531–6.

Marc, D., Girard, M., & Werf, S. Van Der. (1991). A Gly 1 to Ala substitution in poliovirus capsid protein VP0 blocks its myristoylation and prevents viral assembly. *Journal of General Virology*, *72*(5), pp. 1151–1157.

Martin, C. T., Muller, D. K., & Coleman, J. E. (1988). Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry*, *27*(11), pp. 3966–74.

Martínez-Salas, E., & Ryan, M. (2010). Translation and Protein Processing, in: E. Ehrenfeld, E. Domingo, & R. Roos (Eds.), *The Picornaviruses*, pp. 141–161. Washington, DC: ASM Press.

Martrus, G., Nevot, M., Andres, C., Clotet, B., & Martinez, M. A. (2013). Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture. *Retrovirology*, *10*(78).

Mason, P. W., Bezborodova, S. V, & Henry, T. M. (2002). Identification and characterization of a cis-acting replication element (cre) adjacent to the internal ribosome entry site of foot-and-mouth disease virus. *Journal of Virology*, *76*(19), pp. 9686–94.

Mason, P. W., Rieder, E., & Baxt, B. (1994). RGD sequence of foot-and-mouth disease virus is essential for infecting cells via the natural receptor but can be bypassed by an antibody-dependent enhancement pathway. *Proceedings of the National Academy of Sciences of the United States of America*, *91*, pp. 1932–1936.

Matzura, O., & Wennborg, A. (1996). RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. , pp. 247–249.

Mccullough, K. C., Simone, F. D. E., Brocchi, E., Capucci, L., Crowther, J. R., & Kihm, U. (1992). Protective Immune Response against Foot-and-Mouth Disease. *Journal of Virology*, *66*(4), pp. 1835–1840.

McInerney, G. M., King, a M., Ross-Smith, N., & Belsham, G. J. (2000). Replication-competent foot-and-mouth disease virus RNAs lacking capsid coding sequences. *The Journal of General Virology*, *81*(7), pp. 1699–702.

McKnight, K. L., & Lemon, S. M. (1996). Capsid coding sequence is required for efficient replication of human rhinovirus 14 RNA. *Journal of Virology*, *70*(3), pp. 1941–52.

McKnight, K. L., & Lemon, S. M. (1998). The rhinovirus type 14 genome contains an internally located RNA structure that is required for viral replication. *RNA*, *4*(12), pp. 1569–84.

Milligan, J. F., Groebe, D. R., Witherell, G., & Uhlenbeck, O. C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Research*, *15*(21), pp. 8783–8798.

Mizutani, S., & Colonno, R. J. (1985). In vitro synthesis of an infectious RNA from cDNA clones of human rhinovirus type 14. *Journal of Virology*, *56*(2), pp. 628–32.

Moffat, K., Howell, G., Knox, C., Belsham, G. J., Monaghan, P., Ryan, M. D., *et al*. (2005). Effects of foot-and-mouth disease virus nonstructural proteins on the structure and function of the early secretory pathway: 2BC but not 3A blocks endoplasmic reticulum-to-Golgi transport. *Journal of Virology*, *79*(7), pp. 4382–95.

Monaghan, P., Cook, H., Jackson, T., Ryan, M., & Wileman, T. (2004). The ultrastructure of the developing replication site in foot-and-mouth disease virus-infected BHK-38 cells. *Journal of General Virology*, *85*(4), pp. 933–946.

Mowat, G. N., & Chapman, W. G. (1962). Growth of foot-and-mouth disease virus in a fibroblastic cell line derived from hamster kidneys. *Nature*, *194*, pp. 253–5.

Mueller, S., Coleman, J. R., Papamichail, D., Ward, C. B., Nimnual, A., Futcher, B., *et al*. (2010). Live attenuated influenza virus vaccines by computer-aided rational design. *Nature Biotechnology*, *28*(7), pp. 723–6.

Mueller, S., Papamichail, D., Coleman, J. R., Skiena, S., & Wimmer, E. (2006). Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *Journal of Virology*, *80*(19), pp. 9687–96.

Mueller, S., & Wimmer, E. (1998). Expression of foreign proteins by poliovirus polyprotein fusion: analysis of genetic stability reveals rapid deletions and formation of cardioviruslike open reading frames. *Journal of Virology*, *72*(1), pp. 20–31.

Nathanson, N., & Landmuir, A. (1995). The Cutter Incident. *American Journal of Epidemiology*, *142*(2), pp. 29–60.

Nayak, A., Goodfellow, I., & Belsham, G. (2005). Factors required for the uridylylation of the foot-and-mouth disease virus 3B1, 3B2, and 3B3 peptides by the RNA-dependent RNA polymerase (3Dpol) in vitro. *Journal of Virology*, *79*(12), pp. 7698–7706.

Ni, Y., Zhao, Z., Opriessnig, T., Subramaniam, S., Zhou, L., Cao, D., *et al*. (2014). Computer-aided codon-pairs deoptimization of the major envelope GP5 gene attenuates porcine reproductive and respiratory syndrome virus. *Virology*, *450-451*, pp. 132–9.

Nomoto, A., Jacobson, A., Lee, Y. F., Dunn, J., & Wimmer, E. (1979). Defective interfering particles of poliovirus: Mapping of the deletion and evidence that the deletions in the genomes of DI(1), (2) and (3) are located in the same region. *Journal of Molecular Biology*, *128*(2), pp. 179–196.

Nomoto, A., Kitamura, N., Golini, F., & Wimmer, E. (1977). The 5'-terminal structures of poliovirion RNA and poliovirus mRNA differ only in the genome-linked protein VPg. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), pp. 5345–9.

Novella, I. S., Borregob, B., Mateub, M. G., Domingo, E., Giralt, E., & Andreud, D. (1993). Use of substituted and tandem-repeated peptides to probe the relevance of the highly conserved RGD tripeptide in the immune response against foot-and-mouth disease virus. *FEBS Letters*, *330*(3), pp. 253–259.

Núñez, J., Baranowski, E., Molina, N., Ruiz-Jarabo, C., Sánchez, C., Domingo, E., *et al*. (2001). A single amino acid substitution in nonstructural protein 3A can mediate adaptation of foot-and-mouth disease virus to the guinea pig. *Journal of Virology*, *75*(8), pp. 3977–3983.

Ochs, K., Zeller, a., Saleh, L., Bassili, G., Song, Y., Sonntag, a., *et al*. (2003). Impaired Binding of Standard Initiation Factors Mediates Poliovirus Translation Attenuation. *Journal of Virology*, *77*(1), pp. 115–122.

Ormö, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., & Remington, S. J. (1996). Crystal structure of the Aequorea victoria green fluorescent protein. *Science*, *273*(5280), pp. 1392–5.

Pacheco, J. M., Henry, T. M., O'Donnell, V. K., Gregory, J. B., & Mason, P. W. (2003). Role of Nonstructural Proteins 3A and 3B in Host Range and Pathogenicity of Foot-and-Mouth Disease Virus. *Journal of Virology*, *77*(24), pp. 13017–13027.

Palmenberg, A., Neubauer, D., & Skern, T. (2010). Genome Organisation and Encoded Proteins, in: E. Ehrenfeld, E. Domingo, & R. P. Roos (Eds.), *The Picornaviruses*, pp. 3–18. Washington, DC: ASM Press.

Parsley, T., Towner, J., Blyn, L., Ehrenfeld, E., & Semler, B. L. (1997). Poly (rC) binding protein 2 forms a ternary complex with the 5'-terminal sequences of poliovirus RNA and the viral 3CD proteinase. *RNA*, *3*, pp. 1124–1134.

Pattnaik, A. K., Ball, L. A., LeGrone, A. W., & Wertz, G. W. (1992). Infectious defective interfering particles of VSV from transcripts of a cDNA clone. *Cell*, *69*(6), pp. 1011–20.

Paul, A., Boom, J. van, Filippov, D., & Wimmer, E. (1998). Protein-primed RNA synthesis by purified poliovirus RNA polymerase. *Nature*, *393*, pp. 280–284.

Paul, A., Rieder, E., & Kim, D. (2000). Identification of an RNA hairpin in poliovirus RNA that serves as the primary template in the in vitro uridylylation of VPg. *Journal of Virology*, *74*(22), pp. 10359–10370.

Peelle, B., Gururaja, T. L., Payan, D. G., & Anderson, D. C. (2001). Characterization and use of green fluorescent proteins from Renilla mulleri and Ptilosarcus guernyi for the human cell display of functional peptides. *Journal of Protein Chemistry*, *20*(6), pp. 507–19.

Pekosz, A., He, B., & Lamb, R. (1999). Reverse genetics of negative-strand RNA viruses: closing the circle. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(August), pp. 8804–8806.

Percy, N., Barclay, W. S., Sullivan, M., & Almond, J. W. (1992). A Poliovirus Replicon Containing the Chloramphenicol Acetyltransferase Gene Can Be Used To Study the Replication and Encapsidation of Poliovirus RNA. *Journal of Virology*, *66*(8), pp. 5040–46.

Perrotta, a T., & Been, M. D. (1998). A toggle duplex in hepatitis delta virus self-cleaving RNA that stabilizes an inactive and a salt-dependent pro-active ribozyme conformation. *Journal of Molecular Biology*, *279*(2), pp. 361–73.

Perrotta, A. T., & Been, M. D. (1990). The self-cleaving domain from the genomic RNA of hepatitis delta virus: sequence requirements and the effects of denaturant. *Nucleic Acids Research*, *18*(23), pp. 6821–6827.

Perrotta, A. T., & Been, M. D. (1991). A pseudoknot-like structure required for efficient self-cleavage of hepatitis delta virus RNA. *Nature*, *350*, pp. 434–436.

Phan, T. G., Kapusinszky, B., Wang, C., Rose, R. K., Lipton, H. L., & Delwart, E. L. (2011). The fecal viral flora of wild rodents. *PLoS Pathogens*, *7*(9), p. e1002218.

Piccone, M. E., Pacheco, J. M., Pauszek, S. J., Kramer, E., Rieder, E., Borca, M. V, *et al*. (2010). The region between the two polyprotein initiation codons of foot-and-mouth disease virus is critical for virulence in cattle. *Virology*, *396*(1), pp. 152–9.

Piccone, M. E., Rieder, E., Mason, P. W., & Grubman, M. J. (1995). The foot-and-mouth disease virus leader proteinase gene is not required for viral replication. *Journal of Virology*, *69*(9), pp. 5376–5382.

Pilipenko, E. V, Blinov, V. M., Chernov, B. K., Dmitrieva, T. M., & Agol, V. I. (1989). Conservation of the secondary structure elements of the 5'-untranslated region of cardio- and aphthovirus RNAs. *Nucleic Acids Research*, *17*(14), pp. 5701–11.

Pilipenko, E. V, Pestova, T. V, Kolupaeva, V. G., Khitrina, E. V, Poperechnaya, A. N., Agol, V. I., *et al*. (2000). A cell cycle-dependent protein serves as a template-specific translation initiation factor. *Genes & Development*, *14*, pp. 2028–2045.

Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G., & Cormier, M. J. (1992). Primary structure of the Aequorea victoria green-fluorescent protein. *Gene*, *111*(2), pp. 229–33.

Puigbò, P., Bravo, I. G., & Garcia-Vallve, S. (2008). CAIcal: a combined set of tools to assess codon usage adaptation. *Biology Direct*, *3*, p. 38.

Racaniello, V. R., & Baltimore, D. (1981). Cloned poliovirus complementary DNA is infectious in mammalian cells. *Science*, *214*(4523), pp. 916–919.

Rambaut, A. (2008). FigTree. Retrieved from http://tree.bio.ed.ac.uk/software/figtree.

Reed, L., & Muench, H. (1938). A simple method of estimating fifty per cent endpoints. *The American Journal of Hygiene*, *27*(3), pp. 493–497.

Reuter, G., Boros, A., Kiss, T., Delwart, E., & Pankovics, P. (2014). Complete genome characterization of mosavirus (family Picornaviridae) identified in droppings of a European roller (Coracias garrulus) in Hungary. *Archives of Virology*, *159*(10), pp. 2723–9.

Rice, C., Levis, R., Strauss, J., & Huang, H. (1987). Production of infectious RNA transcripts from Sindbis virus cDNA clones: mapping of lethal mutations, rescue of a temperature-sensitive marker, and in vitro mutagenesis to generate defined mutants. *Journal of Virology*, *61*(12), pp. 3809–3819.

Rieder, E., Bunch, T., Brown, F., & Mason, P. (1993). Genetically engineered foot-and-mouth disease viruses with poly (C) tracts of two nucleotides are virulent in mice. *Journal of Virology*, *67*(9), pp. 5139–5145.

Rima, B. K., & Mcferran, N. V. (1997). Dinucleotide and stop codon frequencies in single-stranded RNA viruses. *Journal of General Virology*, *78*, pp. 2859–2870.

Rizzo, M. A., Davidson, M. W., & Piston, D. W. (2009). Fluorescent protein tracking and detection: fluorescent protein structure and color variants. *Cold Spring Harbor Protocols*, *4*(12).

Rodríguez, L. L., & Grubman, M. J. (2009). Foot and mouth disease virus vaccines. *Vaccine*, *27 Suppl 4*, pp. D90–4.

Rodríguez Pulido, M., Sobrino, F., Borrego, B., & Sáiz, M. (2009). Attenuated foot-and-mouth disease virus RNA carrying a deletion in the 3' noncoding region can elicit immunity in swine. *Journal of Virology*, *83*(8), pp. 3475–85.

Rowlands, D., & Minor, P. (2010). Vaccine Strategies, in: E. Ehrenfeld, E. Domingo, & R. Roos (Eds.), *The Picornaviruses*, pp. 431–447. Washington, DC: ASM Press.

Rozovics, J., & Semler, B. L. (2010). Genome Replication I: the Players, in: E. Ehrenfeld, E. Domingo, & R. Roos (Eds.), *The Picornaviruses*, pp. 107–126. Washington, DC.

Ryan, M. D., Belsham, G. J., & King, A. M. Q. (1989). Specificity of Enzyme-Substrate Interactions in Foot-and-Mouth Disease Virus Polyprotein Processing. *Virology*, *173*, pp. 35–45.

Ryan, M. D., & Drew, J. (1994). Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *The EMBO Journal*, *13*(4), pp. 928–933.

Ryan, M. D., & Flint, M. (1997). Virus-encoded proteinases of the picornavirus super-group. *The Journal of General Virology*, *78*(4), pp. 699–723.

Ryan, M. D., King, A. M. Q., & Thomas, G. P. (1991). Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *Journal of General Virology*, *72*(11), pp. 2727–2732.

Ryazanov, A., Shestakova, E., & Natapov, P. (1988). Phosphorylation of elongation factor 2 by EF-2 kinase affects rate of translation. *Nature*, *334*(14), pp. 170–173.

Saiz, M., Gomez, S., Martínez-Salas, E., & Sobrino, F. (2001). Deletion or substitution of the aphthovirus 3′ NCR abrogates infectivity and virus replication. *Journal of General Virology*, *14*(82), pp. 93–101.

Sambrook, J., Fritsch, E., & Maniatis, T. (1989). *Molecular cloning : a laboratory manual.* New York: Cold Spring Harbour.

Sangar, D. V., Newton, S. E., Rowlands, D. J., & Clarke, B. E. (1987). All foot and mouth disease virus serotypes initiate protein synthesis at two separate AUGs. *Nucleic Acids Research*, *15*(8), pp. 3305–3315.

Sarnow, P. (1989). Role of 3'-end sequences in infectivity of poliovirus transcripts made in vitro. *Journal of Virology*, *63*(1), pp. 467–70.

Schenborn, E., & Mierendorf, R. (1985). A novel transcription property of SP6 and T7 RNA polymerases: dependence on template structure. *Nucleic Acids Research*, *13*(17), pp. 6223–6236.

Schlegel, A., Giddings, T. H., Ladinsky, M. S., & Kirkegaard, K. (1996). Cellular Origin and Ultrastructure of Membranes Induced during Poliovirus Infection. *Journal of Virology*, *70*(10), pp. 6576–6588.

Schrodinger, L. (2010). The PyMOL Molecular Graphics System.

Schulte, A., Lorenzen, I., Böttcher, M., & Plieth, C. (2006). A novel fluorescent pH probe for expression in plants. *Plant Methods*, *2*(7).

Seago, J., Juleff, N., Moffat, K., Berryman, S., Christie, J. M., Charleston, B., *et al*. (2013). An infectious recombinant foot-and-mouth disease virus expressing a fluorescent marker protein. *The Journal of General Virology*, *94*(7), pp. 1517–27.

Semler, B. L., Dorner, A. J., & Wimmer, E. (1984). Production of infectious poliovirus from cloned cDNA is dramatically increased by SV40 transcription and replicaton signals. *Nucleic Acids Research*, *12*(12), pp. 5123–5141.

Serrano, P., Pulido, M. R., Sáiz, M., & Martínez-Salas, E. (2006). The 3' end of the foot-and-mouth disease virus genome establishes two distinct long-range RNA-RNA interactions with the 5' end region. *The Journal of General Virology*, *87*(10), pp. 3013–22.

Shaner, N. C., Campbell, R. E., Steinbach, P. a, Giepmans, B. N. G., Palmer, A. E., & Tsien, R. Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology*, *22*(12), pp. 1567–72.

Sharma, N., Donnell, B. J. O., & Flanegan, J. B. (2005). 3 ′ -Terminal Sequence in Poliovirus Negative-Strand Templates Is the Primary cis -Acting Element Required for VPgpUpU-Primed Positive-Strand Initiation. *Journal of Virological Methods*, *79*(6), pp. 3565–3577.

Shimomura, O., Johnson, F. H., & Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *Journal of Cellular and Comparative Physiology*, *59*, pp. 223–39.

Silvestri, L. S., Parilla, J. M., Morasco, B. J., Ogram, S. a, & Flanegan, J. B. (2006). Relationship between poliovirus negative-strand RNA synthesis and the length of the 3' poly(A) tail. *Virology*, *345*(2), pp. 509–19.

Simmonds, P. (2012). SSE: a nucleotide and amino acid sequence analysis platform. *BMC Research Notes*, *5*(50).

Simmonds, P., Xia, W., Baillie, J. K., & McKinnon, K. (2013). Modelling mutational and selection pressures on dinucleotides in eukaryotic phyla--selection against CpG and UpA in cytoplasmically expressed RNA and in RNA viruses. *BMC Genomics*, *14*(1), p. 610.

Spector, D. H., & Baltimore, D. (1974). Requirement of 3 ' -Terminal Poly (adenylic acid) for the Infectivity of Poliovirus RNA. *Proceedings of the National Academy of Sciences of the United States of America*, *71*(8), pp. 2983–2987.

Stanway, G. (1990). Structure, function and evolution of picornaviruses. *Journal of General Virology*, *71*, pp. 2483–2501.

Strebel, K., & Beck, E. (1986). A second protease of foot-and-mouth disease virus. *Journal of Virology*, *58*(3), pp. 893–9.

Sutmoller, P., Barteling, S. S., Olascoaga, R. C., & Sumption, K. J. (2003). Control and eradication of foot-and-mouth disease. *Virus Research*, *91*(1), pp. 101–144.

Sweeney, T. R., Cisnetto, V., Bose, D., Bailey, M., Wilson, J. R., Zhang, X., *et al*. (2010). Foot-and-mouth disease virus 2C is a hexameric AAA+ protein with a coordinated ATP hydrolysis mechanism. *The Journal of Biological Chemistry*, *285*(32), pp. 24347–59.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, *28*, pp. 2731–2739.

Taniguchi, T., Palmieri, M., & Weissmann, C. (1978). QB DNA-containing hybrid plasmids giving rise to QB phage formation in the bacterial host. *Nature*, *274*, pp. 223–228.

Telling, R. C., & Elsworth, R. (1965). Submerged culture of hamster kidney cells in a stainless steel vessel. *Biotechnology and Bioengineering*, *7*(3), pp. 417–434.

Todd, S., Towner, J. S., Brown, D. M., & Semler, B. L. (1997). Replication-Competent Picornaviruses with Complete Genomic RNA 3 ' Noncoding Region Deletions. *Journal of Virology*, *71*(11), pp. 8868–8874.

Tong, L., Lin, L., Zhao, W., Wang, B., Wu, S., Liu, H., *et al*. (2011). Destabilization of coxsackievirus B3 genome integrated with enhanced green fluorescent protein gene. *Intervirology*, *54*(5), pp. 268–75.

Triana-Alonso, F., Dabrowski, M., Wadzack, J., & Nierhaus, K. (1995). Self-coded 3′-extension of run-off transcripts produces aberrant products during in vitro transcription with T7 RNA polymerase. *Journal of Biological Chemistry*, *270*(11), pp. 6298–6307.

Tsien, R. (1998). The green fluorescent protein. *Annual Review of Biochemistry*, *67*, pp. 509–544.

Tulloch, F., Pathania, U., Luke, G. A., Nicholson, J., Stonehouse, N. J., Rowlands, D. J., *et al*. (2014a). FMDV replicons encoding green fluorescent protein are replication competent. *Journal of Virological Methods*, *209C*, pp. 35–40.

Tulloch, F., Atkinson, N. J., Evans, D. J., Ryan, M. D., & Simmonds, P. (2014b). RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. *eLife*, *3*: e04531.

Uhlenbeck, O. (1987). A small catalytic oligoribonucleotide. *Nature*, *328*(6131), pp. 596–600.

Vara, J., Perez-Gonzalez, J., & Jimenez, A. (1985). Biosynthesis of puromycin by Streptomyces alboniger: characterization of puromycin N-acetyltransferase. *Biochemistry*, *24*(1973), pp. 8074–8081.

van Kuppeveld, F. J. M., Belov, G., & Ehrenfeld, E. (2010). Remodelling Cellular Membranes, in: E. Ehrenfeld, E. Domingo, & R. Roos (Eds.), *The Picornaviruses*, pp. 181–193. Washington, DC: ASM Press.

van Kuppeveld, F. J. M., de Jong, A. S., Melchers, W. J. G., & Willems, P. H. G. M. (2005). Enterovirus protein 2B po(u)res out the calcium: a viral strategy to survive? *Trends in Microbiology*, *13*(2), pp. 41–4.

van der Werf, S., Bradley, J., Wimmer, E., Studier, F. W., & Dunn, J. J. (1986). Synthesis of infectious poliovirus RNA by purified T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(8), pp. 2330–4.

Ward, C. D., Stokes, M. A. M., & Flanegan, J. B. (1988). Direct Measurement of the Poliovirus RNA Polymerase Error Frequency In Vitro. *Journal of Virology*, *62*(2), pp. 558–562.

Whitton, J. L., Cornell, C. T., & Feuer, R. (2005). Host and Virus Determinants of Picornavirus Pathogenesis and Tropism. *Nature Reviews Microbiology*, *3*, pp. 765–776.

Witwer, C., Rauscher, S., Hofacker, I. L., & Stadler, P. F. (2001). Conserved RNA secondary structures in Picornaviridae genomes. *Nucleic Acids Research*, *29*(24), pp. 5079–5089.

Yang, F., Moss, L. G., & Phillips, G. N. (1996). The molecular structure of green fluorescent protein. *Nature Biotechnology*, *14*(10), pp. 1246–51.

Zibert, A., Maass, G., Strebel, K., Falk, M. M., & Beck, E. (1990). Infectious foot-and-mouth disease virus derived from a cloned full-length cDNA. *Journal of Virology*, *64*(6), pp. 2467–73.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridisation prediction. , pp. 3406–15.

Appendices

A.1 Primers used throughout this study for cloning purposes

Table A.1 Oligonucleotides used during vector construction. Primer sequences are shown in the 5'- 3' direction with the name and description listed. All primers were designed to maintain a correct ORF.

Primer Name	Primer sequence (5'- 3') and description
JN7	CCCGGGTGGCACCGGGCTTGCGGGTCATGC Reverse primer encoding the C-terminus of Puromycin-N-acetyl-transferase (PAC)
JN9	TTATAACCACTGAACACATGGATATCGTGTCCAAAGGGGAAGAGCTG Forward primer encoding the N-terminus of GFP, with a PsiI site (underlined)
JN10	TACGTAATCGATATCGTGTCCAAAGGGGAAGAGCTG Forward primer encoding the N-terminus of GFP, containing a SnaBI site (underlined)
GFP_fwd	AATACTATGCATATGGATATCGTGTCCAAAGGG Forward primer encoding the N-terminus of GFP, containing an EcoRV site (underlined)
GFP_rev _SmaI	CCCGGGCTTATACAGCTCGTCCATG Reverse primer encoding the C-terminus of GFP, containing a SmaI site (underlined)
PMR for 2	CGGACCCAACATGTGTGCAA Forward primer encoding region of the 5'-UTR from FMDV
PMR for 14	GTTGCAACCCTGATGTTGAT Forward primer encoding region of $3D^{pol}$ from FMDV
HDVrbz_ 5nt_AscI rev	GGCGCGCCTGGCTCTCCCTTAGCCATCCGA Reverse primer encoding HDV ribozyme with 5 extra nucleotides (bold) and an AscI site (underlined)
Pt.GFP_ BamHI fwd	GGCCGCGGATCCAGATGCATATGAACCGCAACGT Forward primer encoding Pt.GFP and a BamHI site (underlined)
Pt.GFP_ PstI_rev	CGCGGCCTGCAGCACCCACTCGTGCAG Reverse primer encoding Pt.GFP and a PstI site (underlined)
Pt.GFP Δ1A_ BamHI fwd	GCGCGCGGATCCCCTTTTATAACCACTGAA Forward primer encoding Δ1A and a BamHI site (underlined)
PMR rev 8	CCGACGTTAGGTCGAACTTC Reverse primer encoding 5'-end of 2B
EM234	CTACGGGGTCTGACGCTCAGTGGAAC Forward primer encoding region of pUC ori
EM235	GTTCCACTGAGCGTCAGACCCCGTAG Reverse primer encoding region of pUC ori
GFPmut fwd	CCGCATCGAGCTGCNCGGCATCGACTTCAAGGAGG Forward primer encoding L^{pro}- GFP cleavage site with mutated Lysine (bold and underlined)

GFPmut rev	CCTCCTTGAAGTCGATGCCGNGCAGCTCGATGCGG Reverse primer encoding L^{pro}- GFP cleavage site with mutated Lysine (bold and underlined)
PMR rev 9	GTCCTCCTGCATCTGGTTGAT Reverse primer encoding 5'-end of 2B
mCherry fwd_NsiI	ATGCATATGGATATCGTGAGCAAGGGCGAGGAG Forward primer encoding 5'-end of mCherry and a NsiI site (underlined)
mCherry fwd_PsiI	TTATAACCACTGAACACATGGTGAGCAAGGGCGAGGAGGAT Forward primer encoding 5'-end of mChery and a PsiI site (underlined)
La-mCherry fwd_PsiI	TTATAACCACTGAACACATGAATACAACTGACTGTTTTATCGCTTTGGTACAGGCTA TCAGAGAGATTAAAGCACTTTTTCTATCACGCACCACAGGGAAA**ATGGTGAGCAAGG GCGAGGAGGATAACATGGCC** Forward primer encoding the 'spacer' region of L^{pro}, the 5'-end of mCherry (bold) and a PsiI site (underlined)
mCherry rev_XmaI	CCCGGGTTTTGTACAATTCATCCAT Reverse primer encoding mCherry and a XmaI site (underlined)
BEV_ Δ1D fwd	CCGCGCCCCGGGGGATTTGTTCTAACAGAGCCAGCTTGACTAGCTATGGACCCTTTG GGCAGCAGCAGGGT Forward primer encoding BEV Δ1D and a region of BEV 2A with a XmaI site (underlined)
BEV_ Δ1D rev	CCGCGCGGGCCCGTCACGTGCTTTGAGCTGTTCCAT Reverse primer encoding region of BEV 2A with an ApaI site (underlined)
BEV_Tav 2A fwd	CCGCGCCCCGGGTCACCGAGTTGCTTTACCGG Forward primer encoding region of BEV 2A with a XmaI site (underlined)
BEV_Tav 2A rev	CGCGCCGGGCCCAGGATTTTCCTCCACGTCCCCGCATGTTAGAAGACTTCCCCTGCC CTCCAGATGTTCCATGGCGTCATCCTC Reverse primer encoding TaV 2A and a region of BEV 2A with an ApaI site (underlined)
BEV_Δ1 D fwd_2	CGCGGCCCCGGGGGATTTGTTCTAACAGA Forward primer encoding region of BEV Δ1D with a XmaI site (underlined)
R1 OS	CCCAATTTGATGTAACACCACACATGG Forward primer used to amplify region1 of E7 for competition assays
R1 OAS	CCCATACTCGGATGTGCTTGGG Reverse primer used to amplify region1 of E7 for competition assays

A.2 pGFP-PAC replicon sequence

Replicon O1K pGFP-PAC

```
1     TTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAG
                                                          SspI
61    GGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGA
121   AGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAAT
181   AAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACC
241   ATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCG
301   CGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCT
361   TGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGC
421   GGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCAT
481   TCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCC
                                  SphI
541   GTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCC
601   GGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCG
661   AGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGC

        SgrAI                                                SpeI -> T7
721   GCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCTGGCTAGGACTAGTAATACG

        promoter  -> FMDV
781   ACTCACTATAGGTTTGAAAGGGGGCATTAGGGTCTCACCCCTAGTAAGCCAACGACAGTC
841   CCTGCGTTGCACTCCACACTTACGTTGTACACACGCGGGACCCGATGGGCTATCGTTCAC

            PvuII
901   CCACCTACAGCTGGACTCACGGCGCCGCGTGGCCATTTAGCTGGATTGTGCGGACGAACA
961   CGCTTGCGCACCTCGCGTGACCGGTTAGTACTCTTACCACTCTCCGCCTACTTGGTCGTT
1021  AGCGCTGTCTTGGGCATTCCTGTGGGGGCCGTTCGACGCTCCACGGGAACTCTCCTGTGT
1081  GACATCTACGGTGATGGGGCCGTTTCGCGTGGGCTGGTCGTTTGGACTGCTTCGGCTGTC
                        NheI
1141  ACCCGGCGCCCGCCTTTCAGCTAGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
1201  CCCCCCCCCCCCCCCGCTCCCCCCCCCCCAAGTTTTACCGTCGTTCCCGACGTAAAAGGGA
              HinDIII                                        HinDIII
1261  GGTAACCACAAGCTTGAAACCGTCCGGCCCGACGTAAAAGGGTGGTAACCACAAGCTTAC
1321  TGCCGTCTTTCCCGACGTTAAAGGGATGAAACCACAAGACTTACCTTCGCTCGGAAGTAA
1381  AACGACAAACACACACAGTTTTGCCCGTTTTCATGAGAAATGGGACGTCTGCGCACGAAA
1441  CGCGCCGTCGCTTGAGGAGGACTTGTACAAACACGATCTATGCAGGTTTCCCCAACTGAC
1501  ACAAACCGTGCAACTTGAAACTCCGCCTGGTCTTTCCAGGTCTAGAGGGGTAACATTTTG
1561  TACTGTGTTTGACTCCACGCTCGATCCACTAGCGAGTGTTAGTAGCGGTACTGCTGTCTC
1621  GTAGCGGAGCATGTTGGCCGTGGGAACACCTCCTTGGTAACAAGGACCCACGGGGCCGAA
1681  AGCCATGTCCTAACGGACCCAACATGTGTGCAACCCCAGCACGGCAGCTTTACTGTGAAA
1741  CCCACTTCAAGGTGACATTGATACTGGTACTCAAACACTGGTGACAGGCTAAGGATGCCC
          KpnI
1801  TTCAGGTACCCCGAGGTAACAAGCGACACTCGGGATCTGAGAAGGGGACTGGGACTTCTT
                        HinDIII
1861  TAAAGTGCCCAGTTTAAAAAGCTTCTACGCCTGAATAGGTGACCGGAGGCCGGCACCTTT

          PsiI              → Lab
1921  CCTTTTATAACCACTGAACACATGAATACAACTGACTGTTTTATCGCTTTGGTACAGGCT
1                         M  N  T  T  D  C  F  I  A  L  V  Q  A

                                                    → Lb
1981  ATCAGAGAGATTAAAGCACTTTTTCTATCACGCACCACAGGGAAAATGGAACTGACACTG
14     I  R  E  I  K  A  L  F  L  S  R  T  T  G  K  M  E  L  T  L
```

```
2041 TACAACGGTGAGAAGAAGACCTTTTACTCCAGGCCCAACAACCACGACAACTGCTGGTTG
34    Y  N  G  E  K  K  T  F  Y  S  R  P  N  N  H  D  N  C  W  L

2101 AACGCCATCCTCCAGTTGTTCAGGTACGTTGAAGAACCATTCTTCGACTGGGTCTACAGT
54    N  A  I  L  Q  L  F  R  Y  V  E  E  P  F  F  D  W  V  Y  S

2161 TCGCCTGAGAACCTCACGCTTGAAGCCATCAAGCAGTTGGAGGATCTCACAGGACTTGAA
74    S  P  E  N  L  T  L  E  A  I  K  Q  L  E  D  L  T  G  L  E

2221 CTGCATGAGGGTGGACCACCTGCTCTCGTGATCTGGAACATCAAGCACTTGCTCCACACC
94    L  H  E  G  G  P  P  A  L  V  I  W  N  I  K  H  L  L  H  T

2281 GGCATCGGCACCGCCTCGCGACCCAGCGAGGTGTGCATGGTGGATGGTACGGACATGTGC
114   G  I  G  T  A  S  R  P  S  E  V  C  M  V  D  G  T  D  M  C
                                  AflII
2341 TTGGCTGATTTCCATGCTGGCATTTTCCTTAAGGGGCAAGAACACGCTGTGTTTGCGTGT
134   L  A  D  F  H  A  G  I  F  L  K  G  Q  E  H  A  V  F  A  C

2401 GTCACCTCCAACGGGTGGTACGCGATTGACGATGAGGACTTCTACCCCTGGACGCCGGAC
154   V  T  S  N  G  W  Y  A  I  D  D  E  D  F  Y  P  W  T  P  D

2461 CCGTCCGACGTTCTGGTGTTTGTCCCGTACGATCAAGAACCACTCAACGGGGAATGGAAA
174   P  S  D  V  L  V  F  V  P  Y  D  Q  E  P  L  N  G  E  W  K

                          → 1A
2521 GCCAAGGTTCAACGCAAGCTCAAAGGGGCTGGACAATCCAGTCCAGCGACCGGCTCGCAG
194   A  K  V  Q  R  K  L  K  G  A  G  Q  S  S  P  A  T  G  S  Q

                          → GFP
                        NsiI  NdeI  EcoRV
2581 AACCAATCTGGCAATACTATGCATATGGATATCGTGTCCAAAGGGGAAGAGCTGTTCACC
214   N  Q  S  G  N  T  M  H  M  D  I  V  S  K  G  E  E  L  F  T

2641 GGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTG
234   G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H  K  F  S  V

                    eGFP active site residues
2701 TCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACC
254   S  G  E  G  E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T

2761 ACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG
274   T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  T  Y  G  V  Q

2821 TGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCC
294   C  F  S  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P

2881 GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGC
314   E  G  Y  V  Q  E  R  T  I  F  F  K  D  D  G  N  Y  K  T  R


                          possible L^pro cleavage site
2941 GCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGAC
334   A  E  V  K  F  E  G  D  T  L  V  N  R  I  E  L  K  G  I  D

3001 TTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAAC
354   F  K  E  D  G  N  I  L  G  H  K  L  E  Y  N  Y  N  S  H  N

3061 GTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCAC
374   V  Y  I  M  A  D  K  Q  K  N  G  I  K  V  N  F  K  I  R  H
```

```
3121 AACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGC
394   N  I  E  D  G  S  V  Q  L  A  D  H  Y  Q  Q  N  T  P  I  G

3181 GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAA
414   D  G  P  V  L  L  P  D  N  H  Y  L  S  T  Q  S  A  L  S  K

3241 GACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATC
434   D  P  N  E  K  R  D  H  M  V  L  L  E  F  V  T  A  A  G  I

                                          → PAC
                                    ApaI
3301 ACTCTCGGCATGGACGAGCTGTATAAGGGGCCCCACCACCACCACCACCACACCGAGTAC
454   T  L  G  M  D  E  L  Y  K  G  P  H  H  H  H  H  H  T  E  Y

3361 AAGCCCACGGTGCGCCTCGCCACCCGCGACGACGTCCCCAGGGCCGTACGCACCCTCGCC
474   K  P  T  V  R  L  A  T  R  D  D  V  P  R  A  V  R  T  L  A
                                          BspEI
                                          AccIII
3421 GCCGCGTTCGCCGACTACCCCGCCACGCGCCACACCGTCGATCCGGACCGCCACATCGAG
494   A  A  F  A  D  Y  P  A  T  R  H  T  V  D  P  D  R  H  I  E

3481 CGGGTCACCGAGCTGCAAGAACTCTTCCTCACGCGCGTCGGCCTCGACATCGGGAAGGTG
514   R  V  T  E  L  Q  E  L  F  L  T  R  V  G  L  D  I  G  K  V

3541 TGGGTCGCGGACGACGGCCCCGCGGTGGCGGTCTGGACCACGCCGGAGAGCGTCGAAGCG
534   W  V  A  D  D  G  P  A  V  A  V  W  T  T  P  E  S  V  E  A

3601 GGGGCGGTGTTCGCCGAGATCGGCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCC
554   G  A  V  F  A  E  I  G  P  R  M  A  E  L  S  G  S  R  L  A
                   StuI
3661 GCGCAGCAACAGATGGAAGGCCTCCTGGCGCCGCACCGGCCCAAGGAGCCCGCGTGGTTC
574   A  Q  Q  Q  M  E  G  L  L  A  P  H  R  P  K  E  P  A  W  F

3721 CTGGCCACCGTCGGCGTCTCGCCCGACCACCAGGGCAAGGGTCTGGGCAGCGCCGTCGTG
594   L  A  T  V  G  V  S  P  D  H  Q  G  K  G  L  G  S  A  V  V
                   EagI
3781 CTCCCCGGAGTGGAGGCGGCCGAGCGCGCCGGGGTGCCCGCCTTCCTGGAGACCTCCGCG
614   L  P  G  V  E  A  A  E  R  A  G  V  P  A  F  L  E  T  S  A

3841 CCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTTCACCGTCACCGCCGACGTCGAGGTG
634   P  R  N  L  P  F  Y  E  R  L  G  F  T  V  T  A  D  V  E  V

                                                → Δ1D
                                                SmaI
3901 CCCGAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCCGGTGCCACCCCGGGTCACCGAG
654   P  E  G  P  R  T  W  C  M  T  R  K  P  G  A  T  R  V  T  E

3961 TTGCTTTACCGGATGAAGAGGGCCGAAACATACTGTCCAAGGCCCTTGCTGGCAATCCAC
674   L  L  Y  R  M  K  R  A  E  T  Y  C  P  R  P  L  L  A  I  H

                                          SgrAI
                                          AgeI                → 2A
4021 CCAACTGAAGCCAGACACAAACAGAAAATTGTGGCACCGGTGAAACAGACTTTGAATTTT
694   P  T  E  A  R  H  K  Q  K  I  V  A  P  V  K  Q  T  L  N  F
```

→ **2B**

ApaI

4081 GACCTTCTCAAGTTGGCGGGA**GACGTC**GAGTCCAACCCT**GGGCCC**TTCTTTTTCTCCGAC
714 D L L K L A G D V E S N P G P F F F S D

4141 GTTAGGTCGAACTTCTCCAAACTGGTGGAAACCATCAACCAGATGCAGGAGGACATGTCA
734 V R S N F S K L V E T I N Q M Q E D M S

4201 ACAAAACACGGGCCTGACTTTAACCGGTTAGTGTCCGCATTTGAGGAGTTGGCCATTGGA
754 T K H G P D F N R L V S A F E E L A I G

HinDIII

4261 GTGAAAGCCATCAGAACCGGTCTCGACGAAGCCAAACCCTGGTAC**AAGCTT**ATCAAGCTC
774 V K A I R T G L D E A K P W Y K L I K L

4321 CTAAGCCGCCTGTCGTGCATGGCCGCTGTGGCAGCACGGTCCAAGGACCCAGTCCTTGTG
794 L S R L S C M A A V A A R S K D P V L V

4381 GCCATCATGCTGGCCGACACCGGTCTCGAGATTCTGGACAGCACCTTCGTCGTGAAGA**AG**
814 A I M L A D T G L E I L D S T F V V K K

4441 **ATCT**CCGACTCGCTCTCCAGTCTCTTTCACGTGCCGGCCCCCGTCTTCAGTTTCGGAGCA
834 I S D S L S S L F H V P A P V F S F G A

4501 CCGGTCCTGTTGGCCGGGTTGGTCAAAGTTGCCTCGAGTTTCTTCCGGTCCACACCCGAA
854 P V L L A G L V K V A S S F F R S T P E

→ **2C**

4561 GACCTTGAGAGAGCAGAGAAACAGCTCAAAGCACGTGACATCAACGACATCTTCGCCATT
874 D L E R A E K Q L K A R D I N D I F A I

4621 CTCAAGAACGGCGAGTGGCTGGTCAAACTGATCCTTGCCATCCGCGACTGGATTAAGGCT
894 L K N G E W L V K L I L A I R D W I K A

4681 TGGATCGCCTCAGAAGAGAAGTTTGTCACCATGACAGACTTGGTGCCTGGCATCCTTGAA
914 W I A S E E K F V T M T D L V P G I L E

4741 AAGCAGCGGGACCTGAACGACCCGAGCAAGTACAAGGAAGCCAAGGAGTGGCTCGACAAC
934 K Q R D L N D P S K Y K E A K E W L D N

4801 GCGCGCCAAGCGTGTTTGAAGAGCGGGAACGTCCACATTGCCAACCTGTGCAAAGTGGTC
954 A R Q A C L K S G N V H I A N L C K V V

4861 GCACCAGCACCCAGCAAGTCGAGGCCCGAACCCGTGGTTGTTTGCCTCCGCGGCAAATCT
974 A P A P S K S R P E P V V V C L R G K S

4921 GGCCAGGGCAAGAGCTTCCTTGCAAACGTGCTTGCACAGGCAATTTCCACCCACTTCACC
994 G Q G K S F L A N V L A Q A I S T H F T

4981 GGCAGAATCGACTCAGTGTGGTACTGCCCACCTGACCCTGACCACTTCGACGGTTACAAC
1014 G R I D S V W Y C P P D P D H F D G Y N

5041 CAGCAAACCGTTGTTGTGATGGATGATTTGGGCCAGAACCCTGACGGCAAGGACTTCAAA
1034 Q Q T V V V M D D L G Q N P D G K D F K

5101 TACTTTGCCCAAATGGTCTCGACCACAGGGTTTATCCCGCCCATGGCATCACTCGAGGAC
1054 Y F A Q M V S T T G F I P P M A S L E D

5161 AAAGGTAAACCTTTCAACAGCAAAGTCATCATCGCGACCACCAACTTGTACTCGGGCTTC
1074 K G K P F N S K V I I A T T N L Y S G F

```
5221 ACCCCGAGGACCATGGTATGTCCCGACGCACTGAACCGGAGGTTTCACTTTGACATCGAT
1094  T  P  R  T  M  V  C  P  D  A  L  N  R  R  F  H  F  D  I  D


5281 GTGAGTGCTAAGGATGGGTACAAAATTAACAGCAAATTGGACATTATCAAAGCACTCGAA
1114  V  S  A  K  D  G  Y  K  I  N  S  K  L  D  I  I  K  A  L  E
5341 GACACCCACGCCAACCCAGTGGCAATGTTTCAATACGACTGTGCCCTTCTCAACGGCATG
1134  D  T  H  A  N  P  V  A  M  F  Q  Y  D  C  A  L  L  N  G  M

5401 GCCGTTGAAATGAAGAGAATGCAACAAGACATGTTCAAGCCTCAACCACCCCTCCAGAAT
1154  A  V  E  M  K  R  M  Q  Q  D  M  F  K  P  Q  P  P  L  Q  N

5461 GTGTACCAGCTTGTTCAGGAGGTGATCGATCGGGTCGAGCTCCACGAGAAAGTGTCGAGT
1174  V  Y  Q  L  V  Q  E  V  I  D  R  V  E  L  H  E  K  V  S  S

                          → 3A
5521 CACCCGATCTTCAAGCAGATCTCAATTCCTTCTCAAAAATCTGTGTTGTACTTTCTCATT
1194  H  P  I  F  K  Q  I  S  I  P  S  Q  K  S  V  L  Y  F  L  I
                             MfeI EcoRI
5581 GAGAAGGGCCAACATGAGGCAGCAATTGAATTCTTTGAGGGCATGGTCCACGACTCCATC
1214  E  K  G  Q  H  E  A  A  I  E  F  F  E  G  M  V  H  D  S  I

5641 AAAGAGGAACTCCGACCCCTCATCCAACAAACTTCATTTGTGAAACGCGCTTTCAAGCGC
1234  K  E  E  L  R  P  L  I  Q  Q  T  S  F  V  K  R  A  F  K  R

5701 CTGAAGGAAAATTTTGAGATTGTTGCTCTGTGTTTAACGCTTTTGGCAAACATTGTGATC
1254  L  K  E  N  F  E  I  V  A  L  C  L  T  L  L  A  N  I  V  I

5761 ATGATCCGTGAGACTCGCAAGAGGCAGAAAATGGTGGATGATGCAGTGAATGAGTACATT
1274  M  I  R  E  T  R  K  R  Q  K  M  V  D  D  A  V  N  E  Y  I
                                                          XbaI
5821 GAGAAAGCAAACATCACCACAGATGACAAGACTCTTGACGAGGCGGAGAAGAGCCCTCTA
1294  E  K  A  N  I  T  T  D  D  K  T  L  D  E  A  E  K  S  P  L
                                                          SphI
5881 GAGACCAGCGGCGCCAGCACCGTTGGCTTTAGAGAGAGAACTCTCCCAGGTCAAAAGGCA
1314  E  T  S  G  A  S  T  V  G  F  R  E  R  T  L  P  G  Q  K  A
                                                             →
                                                          3B1
5941 TGCGATGACGTGAACTCCGAGCCTGCCCAACCTGTTGAGGAGCAACCACAAGCTGAAGGA
1334  C  D  D  V  N  S  E  P  A  Q  P  V  E  E  Q  P  Q  A  E  G

6001 CCCTACGCCGGACCACTCGAGCGTCAGAAACCTCTGAAAGTGAGAGCCAAGCTCCCACAG
1354  P  Y  A  G  P  L  E  R  Q  K  P  L  K  V  R  A  K  L  P  Q

          → 3B2
6061 CAGGAGGGGCCTTACGCTGGTCCGATGGAGAGACAAAAACCGCTAAAAGTGAAAGCAAAA
1374  Q  E  G  P  Y  A  G  P  M  E  R  Q  K  P  L  K  V  K  A  K

                          → 3B3
6121 GCCCCGGTCGTGAAGGAAGGACCTTACGAGGGACCGGTGAAGAAGCCTGTCGCTTTGAAA
1394  A  P  V  V  K  E  G  P  Y  E  G  P  V  K  K  P  V  A  L  K

                                   → 3Cpro
6181 GTGAAAGCTAAGAACCTGATTGTCACTGAGAGTGGTGCCCCACCGACCGACTTGCAAAAG
1414  V  K  A  K  N  L  I  V  T  E  S  G  A  P  P  T  D  L  Q  K


6241 ATGGTCATGGGCAACACAAAGCCTGTTGAGCTCATCCTTGACGGGAAGACGGTAGCCATC
1434  M  V  M  G  N  T  K  P  V  E  L  I  L  D  G  K  T  V  A  I
```

```
6301 TGCTGCGCTACTGGAGTGTTTGGCACTGCTTACCTCGTGCCTCGTCATCTCTTCGCAGAG
1454  C  C  A  T  G  V  F  G  T  A  Y  L  V  P  R  H  L  F  A  E

6361 AAGTATGACAAGATCATGGTGGACGGCAGAGCCATGACAGACAGTGACTACAGAGTGTTT
1474  K  Y  D  K  I  M  V  D  G  R  A  M  T  D  S  D  Y  R  V  F
6421 GAGTTTGAGATCAAAGTAAAAGGACAGGACATGCTCTCAGACGCCGCGCTCATGGTGCTC
1494  E  F  E  I  K  V  K  G  Q  D  M  L  S  D  A  A  L  M  V  L

6481 CACCGTGGGAACCGTGTGAGGGACATCACGAAGCACTTTCGTGACACAGCAAGAATGAAG
1514  H  R  G  N  R  V  R  D  I  T  K  H  F  R  D  T  A  R  M  K

6541 AAAGGCACCCCCGTTGTCGGTGTGATTAATAACGCCGATGTCGGGAGACTGATTTTCTCT
1534  K  G  T  P  V  V  G  V  I  N  N  A  D  V  G  R  L  I  F  S

6601 GGTGAGGCCCTTACTTACAAGGACATTGTGGTTTGCATGGACGGAGACACCATGCCTGGC
1554  G  E  A  L  T  Y  K  D  I  V  V  C  M  D  G  D  T  M  P  G

6661 CTCTTTGCCTACAGAGCCGCCACCAAGGCTGGTTACTGCGGAGGAGCCGTTCTTGCCAAA
1574  L  F  A  Y  R  A  A  T  K  A  G  Y  C  G  G  A  V  L  A  K
                                          PstI
6721 GACGGAGCTGACACTTTCATCGTCGGCACTCACT**CTGCAG**GAGGCAACGGAGTTGGATAC
1594  D  G  A  D  T  F  I  V  G  T  H  S  A  G  G  N  G  V  G  Y

6781 TGCTCATGCGTTTCCAGGTCCATGCTTCTTAAAATGAAGGCACACATTGACCCCGAACCA
1614  C  S  C  V  S  R  S  M  L  L  K  M  K  A  H  I  D  P  E  P

            → 3D^pol
6841 CACCACGAGGGGTTGATTGTGGACACCAGAGATGTGGAAGAGCGCGTTCACGTGATGCGC
1634  H  H  E  G  L  I  V  D  T  R  D  V  E  E  R  V  H  V  M  R
        HinDIII                                                ApaI
6901 AAAACC**AAGCTT**GCACCCACCGTTGCACACGGTGTGTTCAACCCCGAGTTT**GGGCCC**GCT
1654  K  T  K  L  A  P  T  V  A  H  G  V  F  N  P  E  F  G  P  A

6961 GCCTTGTCCAACAAGGACCCGCGTCTGAACGAGGGTGTTGTCCTCGACGAAGTCATCTTC
1674  A  L  S  N  K  D  P  R  L  N  E  G  V  V  L  D  E  V  I  F

7021 TCCAAACACAAGGGAGACACAAAGATGTCTGAGGAGGACAAAGCGCTGTTCCGCCGCTGC
1694  S  K  H  K  G  D  T  K  M  S  E  E  D  K  A  L  F  R  R  C
             MluI
7081 GCTGCTGACT**ACGCGT**CACGCTTGCACAGCGTGTTGGGCACAGCAAATGCCCCACTGAGC
1714  A  A  D  Y  A  S  R  L  H  S  V  L  G  T  A  N  A  P  L  S

7141 ATCTACGAGGCAATCAAGGGTGTCGACGGACTCGACGCCATGGAACCAGACACTGCGCCC
1734  I  Y  E  A  I  K  G  V  D  G  L  D  A  M  E  P  D  T  A  P
           ApaI
7201 GGCCTCCCCT**GGGCCC**TCCAGGGTAAACGCCGCGGCGCGCTCATCGACTTCGAGAACGGC
1754  G  L  P  W  A  L  Q  G  K  R  R  G  A  L  I  D  F  E  N  G

7261 ACGGTCGGACCCGAAGTTGAGGCTGCCCTGAAGCTCATGGAGAAGAGAGAATACAAATTT
1774  T  V  G  P  E  V  E  A  A  L  K  L  M  E  K  R  E  Y  K  F

7321 GTTTGTCAGACCTTCCTGAAGGACGAGATTCGCCCGTTGGAGAAAGTACGTGCCGGTAAG
1794  V  C  Q  T  F  L  K  D  E  I  R  P  L  E  K  V  R  A  G  K

7381 ACTCGCATTGTCGACGTCCTGCCCGTTGAGCACATTCTTTACACCAGGATGATGATTGGC
1814  T  R  I  V  D  V  L  P  V  E  H  I  L  Y  T  R  M  M  I  G

7441 AGATTTTGTGCACAGATGCACTCAAATAACGGACCGCAAATTGGCTCAGCGGTCGGTTGC
1834  R  F  C  A  Q  M  H  S  N  N  G  P  Q  I  G  S  A  V  G  C
```

```
7501 AACCCTGATGTTGATTGGCAGAGATTTGGCACACACTTCGCCCAGTACAGAAACGTGTGG
1854  N  P  D  V  D  W  Q  R  F  G  T  H  F  A  Q  Y  R  N  V  W

7561 GATGTGGACTATTCGGCCTTTGATGCTAATCACTGTAGTGATGCCATGAACATCATGTTT
1874  D  V  D  Y  S  A  F  D  A  N  H  C  S  D  A  M  N  I  M  F

                                                        BamHI
7621 GAGGAGGTGTTTCGCACGGAGTTCGGCTTCCACCCGAATGCTGAGTGGATCCTGAAGACT
1894  E  E  V  F  R  T  E  F  G  F  H  P  N  A  E  W  I  L  K  T

7681 CTTGTGAACACGGAACACGCCTATGAGAACAAACGCATCACTGTTGGAGGCGGAATGCCG
1914  L  V  N  T  E  H  A  Y  E  N  K  R  I  T  V  G  G  G  M  P

7741 TCTGGTTGCTCCGCAACAAGCATCATCAACACAATTTTGAACAACATCTACGTGCTCTAC
1934  S  G  C  S  A  T  S  I  I  N  T  I  L  N  N  I  Y  V  L  Y

7801 GCCCTGCGTAGACACTATGAGGGAGTTGAGCTGGACACATACACCATGATCTCCTACGGA
1954  A  L  R  R  H  Y  E  G  V  E  L  D  T  Y  T  M  I  S  Y  G

7861 GACGACATCGTGGTGGCAAGTGATTATGATTTGGACTTCGAGGCTCTCAAGCCCCACTTT
1974  D  D  I  V  V  A  S  D  Y  D  L  D  F  E  A  L  K  P  H  F
                              PvuII
7921 AAATCCCTTGGCCAAACCATCACTCCAGCTGACAAAAGCGACAAAGGTTTTGTTCTTGGT
1994  K  S  L  G  Q  T  I  T  P  A  D  K  S  D  K  G  F  V  L  G

7981 CACTCCATTACCGATGTCACTTTCCTCAAAAGGCACTTCCACATGGACTATGGAACTGGG
2014  H  S  I  T  D  V  T  F  L  K  R  H  F  H  M  D  Y  G  T  G

8041 TTTTACAAACCTGTGATGGCCTCAAAGACCCTTGAGGCTATCCTCTCCTTTGCACGCCGT
2034  F  Y  K  P  V  M  A  S  K  T  L  E  A  I  L  S  F  A  R  R

8101 GGGACCATACAGGAGAAGTTGATCTCCGTGGCAGGACTCGCCGTCCACTCTGGACCAGAC
2054  G  T  I  Q  E  K  L  I  S  V  A  G  L  A  V  H  S  G  P  D

8161 GAGTACCGGCGTCTCTTTGAGCCTTTCCAAGGTCTCTTTGAGATTCCAAGCTACAGATCA
2074  E  Y  R  R  L  F  E  P  F  Q  G  L  F  E  I  P  S  Y  R  S
                                          MluI
8221 CTTTACCTGCGTTGGGTGAACGCCGTGTGCGGTGACGCGTAATCCCTCAGAGGCCACGAC
2094  L  Y  L  R  W  V  N  A  V  C  G  D  A  *

8281 AGCCGGGCTCTGAGGCGTGCGACACCGTAGGAGTGAAAATCCCGAAAGGGTTTTTCCGCT
8341 TCCTTAATCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
8401 AAAAAAAAAAAAAAA                    HpaI
                    CCCCCCCCCCCCCCCCGTTAACGGCCCCCAAAAAACCCCCCTCCC
                                  SfiI
8461 TACTTCGGGAATACCTCAAAGGCCAAGGGGGCCTCGGCCTCTGCATAAATAAAAAAAATT
8521 ATTCACCCTTGGGGCGAAAAAGGGGCGGAATGGGGCGGATTTAGGGGCGGGAGGGGCGGA
8581 TTTAGGGGCGGAACTATGGTTGCGGACTAATTGAAATGCAGGCTTTGCTTACTTCGGCCT
8641 GCTGGGGGACCGGGGAACTTTC
```

A.3 pRbz_5nts replicon sequence

New FMDV Replicon: pRbz_5nts replicon (pCDNA3.1(+) backbone)

```
          SalI            BglII
1   GTCGACGGATCGGGAGATCTCCCGATCCCCTATGGTGCACTCTCAGTACAATCTGCTCTG

61  ATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGT

121 GCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATTGCATGAAGAATC

                                                    CMV Promoter ->
                                                          MluI
181 TGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGAC
               SpeI
241 ATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCAT

301 ATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACG

361 ACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTT

421 TCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAG
        NdeI
481 TGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGC
                    CMV primer 2
541 ATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAG

601 TCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGT

661 TTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC
                                                        CMV primer
721 ACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGG

                                              SacI    T7 Promoter->
781 GCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTAATACGACTCACTATAG

                    Hammerhead ribozyme              FMDV ->
                                                  SmaI  (Oligo=PMRfor17)
841 GGCCTTTCAACTGATGAGGCCGAAAGGCCGAAAACCCGGTATCCCGGGTTCTTGAAAGGG

901 GGCATTAGGGTCTCACCCCTAGTAAGCCAACGACAGTCCCTGCGTTGCACTCCACACTTA

   (Oligo=PMRrev1)
961 CGTTGTACACACGCGGGACCCGATGGGCTATCGTTCACCCACCTACAGCTGGACTCACGG

1021 CGCCGCGTGGCCATTTAGCTGGATTGTGCGGACGAACACGCTTGCGCACCTCGCGTGACC

          ScaI                              (Oligo=PMRrev3)
1081 GGTTAGTACTCTTACCACTCTCCGCCTACTTGGTCGTTAGCGCTGTCTTGGGCATTCCTG

1141 TGGGGGCCGTTCGACGCTCCACGGGAACTCTCCTGTGTGACATCTACGGTGATGGGGCCG
                                                          NheI
1201 TTTCGCGTGGGCTGGTCGTTTGGACTGCTTCGGCTGTCACCCGGCGCCCGCCTTTCAGCT

1261 AGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCTCCCCC
```

```
                                                  (Oligo=PMRrev2)      (Oligo=PMRfor1)
1321 CCCCCCAAGTTTTACCGTCGTTCCCGACGTAAAAGGGAGGTAACCAC**AAGCTT**GAAACCG

1381 TCCGGCCCGACGTAAAAGGGTGGTAACCAC**AAGCTT**ACTGCCGTCTTTCCCGACGTTAAA

1441 GGGATGAAACCACAAGACTTACCTTCGCTCGGAAGTAAAACGACAAACACACACAGTTTT
              BspHI
1501 GCCCGTTT**TCATGA**GAAATGGGACGTCTGCGCACGAAACGCGCCGTCGCTTGAGGAGGAC

1561 TTGTACAAACACGATCTATGCAGGTTTCCCCAACTGACACAAACCGTGCAACTTGAAACT
                         XbaI
1621 CCGCCTGGTCTTTCCAGG**TCTAGA**GGGGTAACATTTTGTACTGTGTTTGACTCCACGCTC

1681 GATCCACTAGCGAGTGTTAGTAGCGGTACTGCTGTCTCGTAGCGGAGCATGTTGGCCGTG

                                                  (Oligo=PMRrev7)      (Oligo=PMRfor2)
1741 GGAACACCTCCTTGGTAACAAGGACCCACGGGGCCGAAAGCCATGTCCTAACGGACCCAA

1801 CATGTGTGCAACCCCAGCACGGCAGCTTTACTGTGAAACCCACTTCAAGGTGACATTGAT
                                                        KpnI
1861 ACTGGTACTCAAACACTGGTGACAGGCTAAGGATGCCCTTCA**GGTACC**CCGAGGTAACAA

1921 GCGACACTCGGGATCTGAGAAGGGGACTGGGACTTC**TTTAAA**GTGCCCAGTTTAAA**AAGC**
                                    NaeI             PsiI
1981 **TT**CTACGCCTGAATAGGTGACCGGAG**GCCGGC**ACCTTTCCTT**TTATAA**CCACTGAACACA

     Lab^pro ->
2041 TGAATACAACTGACTGTTTTATCGCTTTGGTACAGGCTATCAGAGAGATTAAAGCACTTT
1    M  N  T  T  D  C  F  I  A  L  V  Q  A  I  R  E  I  K  A  L

                                Lb^pro ->
2101 TTCTATCACGCACCACAGGGAAAATGGAACTGACACTGTACAACGGTGAGAAGAAGACCT
21   F  L  S  R  T  T  G  K  M  E  L  T  L  Y  N  G  E  K  K  T

2161 TTTACTCCAGGCCCAACAACCACGACAACTGCTGGTTGAACGCCATCCTCCAGTTGTTCA
41   F  Y  S  R  P  N  N  H  D  N  C  W  L  N  A  I  L  Q  L  F

2221 GGTACGTTGAAGAACCATTCTTCGACTGGGTCTACAGTTCGCCTGAGAACCTCACGCTTG
61   R  Y  V  E  E  P  F  F  D  W  V  Y  S  S  P  E  N  L  T  L

2281 AAGCCATCAAGCAGTTGGAGGATCTCACAGGACTTGAACTGCATGAGGGTGGACCACCTG
81   E  A  I  K  Q  L  E  D  L  T  G  L  E  L  H  E  G  G  P  P
                                                              NruI
2341 CTCTCGTGATCTGGAACATCAAGCACTTGCTCCACACCGGCATCGGCACCGCC**TCGCGA**C
101  A  L  V  I  W  N  I  K  H  L  L  H  T  G  I  G  T  A  S  R
                                                  (Oligo=PMRfor3)
2401 CCAGCGAGGTGTGCATGGTGGATGGTACGGACATGTGCTTGGCTGATTTCCATGCTGGCA
121  P  S  E  V  C  M  V  D  G  T  D  M  C  L  A  D  F  H  A  G
          AflII
2461 TTTTC**CTTAAG**GGGCAAGAACACGCTGTGTTTGCGTGTGTCACCTCCAACGGGTGGTACG
141  I  F  L  K  G  Q  E  H  A  V  F  A  C  V  T  S  N  G  W  Y

2521 CGATTGACGATGAGGACTTCTACCCCTGGACGCCGGACCCGTCCGACGTTCTGGTGTTTG
161  A  I  D  D  E  D  F  Y  P  W  T  P  D  P  S  D  V  L  V  F
       BsiWI/SunI
2581 TCC**CGTACG**ATCAAGAACCACTCAACGGGGAATGGAAAGCCAAGGTTCAACGCAAGCTCA
181  V  P  Y  D  Q  E  P  L  N  G  E  W  K  A  K  V  Q  R  K  L
```

Δ1A -> NsiI/

2641 AAGGGGCTGGACAATCCAGTCCAGCGACCGGCTCGCAGAACCAATCTGGCAATACT**ATGC**
201 K G A G Q S S P A T G S Q N Q S G N T M

GFP ->

NdeI/EcoRV

2701 **ATATGGATATC**GTGTCCAAAGGGGAAGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCG
221 H M D I V S K G E E L F T G V V P I L V

2761 AGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
241 E L D G D V N G H K F S V S G E G E G D

2821 CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCT
261 A T Y G K L T L K F I C T T G K L P V P

eGFP active site residues

2881 GGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACC
281 W P T L V T T L T Y G V Q C F S R Y P D

2941 ACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA
301 H M K Q H D F F K S A M P E G Y V Q E R

3001 CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG
321 T I F F K D D G N Y K T R A E V K F E G

possible Lpro cleavage site

3061 ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
341 D T L V N R I E L K G I D F K E D G N I

3121 TGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGC
361 L G H K L E Y N Y N S H N V Y I M A D K

3181 AGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGC
381 Q K N G I K V N F K I R H N I E D G S V

3241 AGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
401 Q L A D H Y Q Q N T P I G D G P V L L P

3301 ACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATC
421 D N H Y L S T Q S A L S K D P N E K R D

3361 ACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGT
441 H M V L L E F V T A A G I T L G M D E L

PAC ->

ApaI

3421 ATAAG**GGGCCC**CACCACCACCACCACCACACCGAGTACAAGCCCACGGTGCGCCTCGCCA
461 Y K G P H H H H H H T E Y K P T V R L A

BsiWI

3481 CCCGCGACGACGTCCCCAGGGC**CGTACG**CACCCTCGCCGCCGCGTTCGCCGACTACCCCG
481 T R D D V P R A V R T L A A A F A D Y P

BspEI

3541 CCACGCGCCACACCGTCGA**TCCGGA**CCGCCACATCGAGCGGGTCACCGAGCTGCAAGAAC
501 A T R H T V D P D R H I E R V T E L Q E

3601 TCTTCCTCACGCGCGTCGGGCTCGACATCGGCAAGGTGTGGGTCGCGGACGACGGCCCCG
521 L F L T R V G L D I G K V W V A D D G P

3661 CGGTGGCGGTCTGGACCACGCCGGAGAGCGTCGAAGCGGGGGCGGTGTTCGCCGAGATCG
541 A V A V W T T P E S V E A G A V F A E I

```
                                                                  StuI
3721 GCCCGCGCATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCC
561  G  P  R  M  A  E  L  S  G  S  R  L  A  A  Q  Q  Q  M  E  G

3781 TCCTGGCGCCGCACCGGCCCAAGGAGCCCGCGTGGTTCCTGGCCACCGTCGGCGTCTCGC
581  L  L  A  P  H  R  P  K  E  P  A  W  F  L  A  T  V  G  V  S
                                                              EagI
3841 CCGACCACCAGGGCAAGGGTCTGGGCAGCGCCGTCGTGCTCCCCGGAGTGGAGGCGGCCG
601  P  D  H  Q  G  K  G  L  G  S  A  V  V  L  P  G  V  E  A  A

3901 AGCGCGCCGGGGTGCCCGCCTTCCTGGAGACCTCCGCGCCCCGCAACCTCCCCTTCTACG
621  E  R  A  G  V  P  A  F  L  E  T  S  A  P  R  N  L  P  F  Y

3961 AGCGGCTCGGCTTCACCGTCACCGCCGACGTCGAGGTGCCCGAAGGACCGCGCACCTGGT
641  E  R  L  G  F  T  V  T  A  D  V  E  V  P  E  G  P  R  T  W

                              (Δ1D ->)
                                 SmaI
4021 GCATGACCCGCAAGCCCGGTGCCACCCCGGGTCACCGAGTTGCTTTACCGGATGAAGAGGG
661  C  M  T  R  K  P  G  A  T  R  V  T  E  L  L  Y  R  M  K  R
                                                 (Oligo=PMRfor6)
4081 CCGAAACATACTGTCCAAGGCCCTTGCTGGCAATCCACCCAACTGAAGCCAGACACAAAC
681  A  E  T  Y  C  P  R  P  L  L  A  I  H  P  T  E  A  R  H  K
                                  2A ->
                    SgrAI
4141 AGAAAATTGTGGCACCGGTGAAACAGACTTTGAATTTTGACCTTCTCAAGTTGGCGGGAG
701  Q  K  I  V  A  P  V  K  Q  T  L  N  F  D  L  L  K  L  A  G

                               2B ->
                              ApaI
4201 ACGTCGAGTCCAACCCTGGGCCCTTCTTTTTTCTCCGACGTTAGGTCGAACTTCTCCAAAC
721  D  V  E  S  N  P  G  P  F  F  F  S  D  V  R  S  N  F  S  K

4261 TGGTGGAAACCATCAACCAGATGCAGGAGGACATGTCAACAAAACACGGGCCTGACTTTA
741  L  V  E  T  I  N  Q  M  Q  E  D  M  S  T  K  H  G  P  D  F

4321 ACCGGTTAGTGTCCGCATTTGAGGAGTTGGCCATTGGAGTGAAAGCCATCAGAACCGGTC
761  N  R  L  V  S  A  F  E  E  L  A  I  G  V  K  A  I  R  T  G

4381 TCGACGAAGCCAAACCCTGGTACAAGCTTATCAAGCTCCTAAGCCGCCTGTCGTGCATGG
781  L  D  E  A  K  P  W  Y  K  L  I  K  L  L  S  R  L  S  C  M

4441 CCGCTGTGGCAGCACGGTCCAAGGACCCAGTCCTTGTGGCCATCATGCTGGCCGACACCG
801  A  A  V  A  A  R  S  K  D  P  V  L  V  A  I  M  L  A  D  T
                                                BglII
4501 GTCTCGAGATTCTGGACAGCACCTTCGTCGTGAAGAAGATCTCCGACTCGCTCTCCAGTC
821  G  L  E  I  L  D  S  T  F  V  V  K  K  I  S  D  S  L  S  S

4561 TCTTTCACGTGCCGGCCCCCGTCTTCAGTTTCGGAGCACCGGTCCTGTTGGCCGGGTTGG
841  L  F  H  V  P  A  P  V  F  S  F  G  A  P  V  L  L  A  G  L
                                          (oligo=PMRfor7)
4621 TCAAAGTTGCCTCGAGTTTCTTCCGGTCCACACCCGAAGACCTTGAGAGAGCAGAGAAAC
861  V  K  V  A  S  S  F  F  R  S  T  P  E  D  L  E  R  A  E  K

       2C ->
4681 AGCTCAAAGCACGTGACATCAACGACATCTTCGCCATTCTCAAGAACGGCGAGTGGCTGG
881  Q  L  K  A  R  D  I  N  D  I  F  A  I  L  K  N  G  E  W  L

4741 TCAAACTGATCCTTGCCATCCGCGACTGGATTAAGGCTTGGATCGCCTCAGAAGAGAAGT
901  V  K  L  I  L  A  I  R  D  W  I  K  A  W  I  A  S  E  E  K
```

```
4801 TTGTCACCATGACAGACTTGGTGCCTGGCATCCTTGAAAAGCAGCGGGACCTGAACGACC
921  F  V  T  M  T  D  L  V  P  G  I  L  E  K  Q  R  D  L  N  D

4861 CGAGCAAGTACAAGGAAGCCAAGGAGTGGCTCGACAACGCGCGCCAAGCGTGTTTGAAGA
941  P  S  K  Y  K  E  A  K  E  W  L  D  N  A  R  Q  A  C  L  K

4921 GCGGGAACGTCCACATTGCCAACCTGTGCAAAGTGGTCGCACCAGCACCCAGCAAGTCGA
961  S  G  N  V  H  I  A  N  L  C  K  V  V  A  P  A  P  S  K  S

4981 GGCCCGAACCCGTGGTTGTTTGCCTCCGCGGCAAATCTGGCCAGGGCAAGAGCTTCCTTG
981  R  P  E  P  V  V  V  C  L  R  G  K  S  G  Q  G  K  S  F  L
                                                (Oligo=PMRfor8)
5041 CAAACGTGCTTGCACAGGCAATTTCCACCCACTTCACCGGCAGAATCGACTCAGTGTGGT
1001 A  N  V  L  A  Q  A  I  S  T  H  F  T  G  R  I  D  S  V  W

5101 ACTGCCCACCTGACCCTGACCACTTCGACGGTTACAACCAGCAAACCGTTGTTGTGATGG
1021 Y  C  P  P  D  P  D  H  F  D  G  Y  N  Q  Q  T  V  V  V  M

5161 ATGATTTGGGCCAGAACCCTGACGGCAAGGACTTCAAATACTTTGCCCAAATGGTCTCGA
1041 D  D  L  G  Q  N  P  D  G  K  D  F  K  Y  F  A  Q  M  V  S

5221 CCACAGGGTTTATCCCGCCCATGGCATCACTCGAGGACAAAGGTAAACCTTTCAACAGCA
1061 T  T  G  F  I  P  P  M  A  S  L  E  D  K  G  K  P  F  N  S
               NruI
5281 AAGTCATCA**TCGCGA**CCACCAACTTGTACTCGGGCTTCACCCCGAGGACCATGGTATGTC
1081 K  V  I  I  A  T  T  N  L  Y  S  G  F  T  P  R  T  M  V  C

5341 CCGACGCACTGAACCGGAGGTTTCACTTTGACATCGATGTGAGTGCTAAGGATGGGTACA
1101 P  D  A  L  N  R  R  F  H  F  D  I  D  V  S  A  K  D  G  Y
                                                (Oligo=PMRfor9)
5401 AAATTAACAGCAAATTGGACATTATCAAAGCACTCGAAGACACCCACGCCAACCCAGTGG
1121 K  I  N  S  K  L  D  I  I  K  A  L  E  D  T  H  A  N  P  V

5461 CAATGTTTCAATACGACTGTGCCCTTCTCAACGGCATGGCCGTTGAAATGAAGAGAATGC
1141 A  M  F  Q  Y  D  C  A  L  L  N  G  M  A  V  E  M  K  R  M

5521 AACAAGACATGTTCAAGCCTCAACCACCCCTCCAGAATGTGTACCAGCTTGTTCAGGAGG
1161 Q  Q  D  M  F  K  P  Q  P  P  L  Q  N  V  Y  Q  L  V  Q  E

                                                              3A ->
        PvuI       SacI                                        BglII
5581 TGAT**CGATCG**GGTC**GAGCTC**CACGAGAAAGTGTCGAGTCACCCGATCTTCAAGC**AGATCT**
1181 V  I  D  R  V  E  L  H  E  K  V  S  S  H  P  I  F  K  Q  I

5641 CAATTCCTTCTCAAAAATCTGTGTTGTACTTTCTCATTGAGAAGGGCCAACATGAGGCAG
1201 S  I  P  S  Q  K  S  V  L  Y  F  L  I  E  K  G  Q  H  E  A
     MfeI / EcoRI
5701 **CAATTGAATTC**TTTGAGGGCATGGTCCACGACTCCATCAAAGAGGAACTCCGACCCCTCA
1221 A  I  E  F  F  E  G  M  V  H  D  S  I  K  E  E  L  R  P  L

5761 TCCAACAAACTTCATTTGTGAAACGCGCTTTCAAGCGCCTGAAGGAAAATTTTGAGATTG
1241 I  Q  Q  T  S  F  V  K  R  A  F  K  R  L  K  E  N  F  E  I
                                                   BspHI
5821 TTGCTCTGTGTTTAACGCTTTTGGCAAACATTGTGA**TCATGA**TCCGTGAGACTCGCAAGA
1261 V  A  L  C  L  T  L  L  A  N  I  V  I  M  I  R  E  T  R  K
                  (Oligos = PCRAph2rev / PCRAph2for)(Oligo=PMRfor10)
5881 GGCAGAAAATGGTGGATGATGCAGTGAATGAGTACATTGAGAAAGCAAACATCACCACAG
1281 R  Q  K  M  V  D  D  A  V  N  E  Y  I  E  K  A  N  I  T  T
                                                XbaI
5941 ATGACAAGACTCTTGACGAGGCGGAGAAGAGCCC**TCTAGA**GACCAGCGGCGCCAGCACCG
1301 D  D  K  T  L  D  E  A  E  K  S  P  L  E  T  S  G  A  S  T
```

```
                                                SphI
6001 TTGGCTTTAGAGAGAGAACTCTCCCAGGTCAAAAGGCATGCGATGACGTGAACTCCGAGC
1321 V  G  F  R  E  R  T  L  P  G  Q  K  A  C  D  D  V  N  S  E

                                                3B1->
6061 CTGCCCAACCTGTTGAGGAGCAACCACAAGCTGAAGGACCCTACGCCGGACCACTCGAGC
1341 P  A  Q  P  V  E  E  Q  P  Q  A  E  G  P  Y  A  G  P  L  E

                                                      3B2->
6121 GTCAGAAACCTCTGAAAGTGAGAGCCAAGCTCCCACAGCAGGAGGGGCCTTACGCTGGTC
1361 R  Q  K  P  L  K  V  R  A  K  L  P  Q  Q  E  G  P  Y  A  G

                                                            3B3->
6181 CGATGGAGAGACAAAAACCGCTAAAAGTGAAAGCAAAAGCCCCGGTCGTGAAGGAAGGAC
1381 P  M  E  R  Q  K  P  L  K  V  K  A  K  A  P  V  V  K  E  G

6241 CTTACGAGGGACCGGTGAAGAAGCCTGTCGCTTTGAAAGTGAAAGCTAAGAACCTGATTG
1401 P  Y  E  G  P  V  K  K  P  V  A  L  K  V  K  A  K  N  L  I

                  3Cpro->                       (Oligo=PMRfor11)
6301 TCACTGAGAGTGGTGCCCCACCGACCGACTTGCAAAAGATGGTCATGGGCAACACAAAGC
1421 V  T  E  S  G  A  P  P  T  D  L  Q  K  M  V  M  G  N  T  K
          SacI
6361 CTGTTGAGCTCATCCTTGACGGGAAGACGGTAGCCATCTGCTGCGCTACTGGAGTGTTTG
1441 P  V  E  L  I  L  D  G  K  T  V  A  I  C  C  A  T  G  V  F

6421 GCACTGCTTACCTCGTGCCTCGTCATCTCTTCGCAGAGAAGTATGACAAGATCATGGTGG
1461 G  T  A  Y  L  V  P  R  H  L  F  A  E  K  Y  D  K  I  M  V

6481 ACGGCAGAGCCATGACAGACAGTGACTACAGAGTGTTTGAGTTTGAGATCAAAGTAAAAG
1481 D  G  R  A  M  T  D  S  D  Y  R  V  F  E  F  E  I  K  V  K

6541 GACAGGACATGCTCTCAGACGCCGCGCTCATGGTGCTCCACCGTGGGAACCGTGTGAGGG
1501 G  Q  D  M  L  S  D  A  A  L  M  V  L  H  R  G  N  R  V  R

6601 ACATCACGAAGCACTTTCGTGACACAGCAAGAATGAAGAAAGGCACCCCCGTTGTCGGTG
1521 D  I  T  K  H  F  R  D  T  A  R  M  K  K  G  T  P  V  V  G
        VspI
6661 TGATTAATAACGCCGATGTCGGGAGACTGATTTTCTCTGGTGAGGCCCTTACTTACAAGG
1541 V  I  N  N  A  D  V  G  R  L  I  F  S  G  E  A  L  T  Y  K
                                              (Oligo=PMRfor12)
6721 ACATTGTGGTTTGCATGGACGGAGACACCATGCCTGGCCTCTTTGCCTACAGAGCCGCCA
1561 D  I  V  V  C  M  D  G  D  T  M  P  G  L  F  A  Y  R  A  A

6781 CCAAGGCTGGTTACTGCGGAGGAGCCGTTCTTGCCAAAGACGGAGCTGACACTTTCATCG
1581 T  K  A  G  Y  C  G  G  A  V  L  A  K  D  G  A  D  T  F  I
                  PstI
6841 TCGGCACTCACTCTGCAGGAGGCAACGGAGTTGGATACTGCTCATGCGTTTCCAGGTCCA
1601 V  G  T  H  S  A  G  G  N  G  V  G  Y  C  S  C  V  S  R  S

                                                      3Dpol ->
6901 TGCTTCTTAAAATGAAGGCACACATTGACCCCGAACCACACCACGAGGGGTTGATTGTGG
1621 M  L  L  K  M  K  A  H  I  D  P  E  P  H  H  E  G  L  I  V

6961 ACACCAGAGATGTGGAAGAGCGCGTTCACGTGATGCGCAAAACCAAGCTTGCACCCACCG
1641 D  T  R  D  V  E  E  R  V  H  V  M  R  K  T  K  L  A  P  T
                                    ApaI
7021 TTGCACACGGTGTGTTCAACCCCGAGTTTGGGCCCGCTGCCTTGTCCAACAAGGACCCGC
1661 V  A  H  G  V  F  N  P  E  F  G  P  A  A  L  S  N  K  D  P
```

```
7081 GTCTGAACGAGGGTGTTGTCCTCGACGAAGTCATCTTCTCCAAACACAAGGGAGACACAA
1681 R  L  N  E  G  V  V  L  D  E  V  I  F  S  K  H  K  G  D  T

                                          (Oligo=PMRfor13)  MluI
7141 AGATGTCTGAGGAGGACAAAGCGCTGTTCCGCCGCTGCGCTGCTGACTACGCGTCACGCT
1701 K  M  S  E  E  D  K  A  L  F  R  R  C  A  A  D  Y  A  S  R

7201 TGCACAGCGTGTTGGGCACAGCAAATGCCCCACTGAGCATCTACGAGGCAATCAAGGGTG
1721 L  H  S  V  L  G  T  A  N  A  P  L  S  I  Y  E  A  I  K  G
                                                        ApaI
7261 TCGACGGACTCGACGCCATGGAACCAGACACTGCGCCCGGCCTCCCCTGGGCCCTCCAGG
1741 V  D  G  L  D  A  M  E  P  D  T  A  P  G  L  P  W  A  L  Q

7321 GTAAACGCCGCGGCGCGCTCATCGACTTCGAGAACGGCACGGTCGGACCCGAAGTTGAGG
1761 G  K  R  R  G  A  L  I  D  F  E  N  G  T  V  G  P  E  V  E

7381 CTGCCCTGAAGCTCATGGAGAAGAGAGAATACAAATTTGTTTGTCAGACCTTCCTGAAGG
1781 A  A  L  K  L  M  E  K  R  E  Y  K  F  V  C  Q  T  F  L  K

7441 ACGAGATTCGCCCGTTGGAGAAAGTACGTGCCGGTAAGACTCGCATTGTCGACGTCCTGC
1801 D  E  I  R  P  L  E  K  V  R  A  G  K  T  R  I  V  D  V  L
                                                    Alw44I
7501 CCGTTGAGCACATTCTTTACACCAGGATGATGATTGGCAGATTTTGTGCACAGATGCACT
1821 P  V  E  H  I  L  Y  T  R  M  M  I  G  R  F  C  A  Q  M  H

                              EspI          (Oligo=PMRfor14)
7561 CAAATAACGGACCGCAAATTGGCTCAGCGGTCGGTTGCAACCCTGATGTTGATTGGCAGA
1841 S  N  N  G  P  Q  I  G  S  A  V  G  C  N  P  D  V  D  W  Q

7621 GATTTGGCACACACTTCGCCCAGTACAGAAACGTGTGGGATGTGGACTATTCGGCCTTTG
1861 R  F  G  T  H  F  A  Q  Y  R  N  V  W  D  V  D  Y  S  A  F

7681 ATGCTAATCACTGTAGTGATGCCATGAACATCATGTTTGAGGAGGTGTTTCGCACGGAGT
1881 D  A  N  H  C  S  D  A  M  N  I  M  F  E  E  V  F  R  T  E
                                 BamHI
7741 TCGGCTTCCACCCGAATGCTGAGTGGATCCTGAAGACTCTTGTGAACACGGAACACGCCT
1901 F  G  F  H  P  N  A  E  W  I  L  K  T  L  V  N  T  E  H  A

7801 ATGAGAACAAACGCATCACTGTTGGAGGCGGAATGCCGTCTGGTTGCTCCGCAACAAGCA
1921 Y  E  N  K  R  I  T  V  G  G  G  M  P  S  G  C  S  A  T  S

7861 TCATCAACACAATTTTGAACAACATCTACGTGCTCTACGCCCTGCGTAGACACTATGAGG
1941 I  I  N  T  I  L  N  N  I  Y  V  L  Y  A  L  R  R  H  Y  E

7921 GAGTTGAGCTGGACACATACACCATGATCTCCTACGGAGACGACATCGTGGTGGCAAGTG
1961 G  V  E  L  D  T  Y  T  M  I  S  Y  G  D  D  I  V  V  A  S
                                                 (Oligo=PMRfor15)
7981 ATTATGATTTGGACTTCGAGGCTCTCAAGCCCCACTTTAAATCCCTTGGCCAAACCATCA
1981 D  Y  D  L  D  F  E  A  L  K  P  H  F  K  S  L  G  Q  T  I

8041 CTCCAGCTGACAAAAGCGACAAAGGTTTTGTTCTTGGTCACTCCATTACCGATGTCACTT
2001 T  P  A  D  K  S  D  K  G  F  V  L  G  H  S  I  T  D  V  T

8101 TCCTCAAAAGGCACTTCCACATGGACTATGGAACTGGGTTTTACAAACCTGTGATGGCCT
2021 F  L  K  R  H  F  H  M  D  Y  G  T  G  F  Y  K  P  V  M  A

8161 CAAAGACCCTTGAGGCTATCCTCTCCTTTGCACGCCGTGGGACCATACAGGAGAAGTTGA
2041 S  K  T  L  E  A  I  L  S  F  A  R  R  G  T  I  Q  E  K  L

8221 TCTCCGTGGCAGGACTCGCCGTCCACTCTGGACCAGACGAGTACCGGCGTCTCTTTGAGC
2061 I  S  V  A  G  L  A  V  H  S  G  P  D  E  Y  R  R  L  F  E
```

```
      EcoNI
8281 CTTTCCAAGGTCTCTTTGAGATTCCAAGCTACAGATCACTTTACCTGCGTTGGGTGAACG
2081 P  F  Q  G  L  F  E  I  P  S  Y  R  S  L  Y  L  R  W  V  N

                MluI                          (Oligo=PMRfor16)
8341 CCGTGTGCGGTGACGCGTAATCCCTCAGAGGCCACGACAGCCGGGCTCTGAGGCGTGCGA
2101 A  V  C  G  D  A  *

8401 CACCGTAGGAGTGAAAATCCCGAAAGGGTTTTTCCCGCTTCCTTAATCCAAAAAAAAAAAA

8461 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGGTCGGC

                          Hepatitis Delta Ribozyme
8521 ATGGCATCTCCACCTCCTCGCGGTCCGACCTGGGCATCCGAAGGAGGACGCACGTCCACTC

                    5nts   AscI
8581 GGATGGCTAAGGGAGAGCCAGGCGCGCCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGC

                EspI                          T7Φ terminator
8641 TGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGG
                             (BspEI /XmaI)
8701 GGTTTTTTGCTGAAAGGAGGAACTATATCCGGGAGCTTGTATATCCATTTTCGGATCTGAT

                                    Neomycin
8761 CAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCT
8821 GGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTC
8881 TGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGA
8941 CCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCAC
9001 GACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCT
9061 GCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAA
9121 AGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCC
9181 ATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCT
9241 TGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGC
9301 CAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTG
9361 CTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCT
9421 GGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCT
9481 TGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCGATTCGCA
                                                          BstBI
9541 GCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAA
9601 ATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTC
9661 TATGAAAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGC

                                                  SV40 polyA
9721 GGGGATCTCATGCTGGAGTTCTTCGCCCACCCCAACTTGTTTATTGCAGCTTATAATGGT
9781 TACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCT
                                                        SalI
9841 AGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCT
9901 AGCTAGAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTC
9961 ACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGA
10021 GTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTG
10081 TCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGG
10141 CGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCG
10201 GTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGA

                    pUC Ori
10261 AAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTG
10321 GCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAG
10381 AGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTC
10441 GTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCG
10501 GGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTT
10561 CGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCC
10621 GGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCC
```

```
10681 ACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGG
10741 TGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCA
10801 GTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGC
10861 GGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCT
10921 TTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTG
10981 GTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTT
                                                          Amp
11041 AAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGT
11101 GAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTC
11161 GTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCG
11221 CGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCC
11281 GAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGG
11341 GAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACA
11401 GGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGA
11461 TCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCT
11521 CCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTG
11581 CATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCA
11641 ACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATA
11701 CGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCT
11761 TCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACT
11821 CGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAA
11881 ACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTC
                     SspI
11941 ATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGA
12001 TACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGA
12061 AAAGTGCCACCTGAC
```

A.4 Accession numbers of 3D^{pol} sequences of the genus type species of the twenty six picornavirus genera

NCBI Accession number	*Type species*
EU142040	Aquamavirus A
JQ316470	Pasivirus A
L02971	Human parechovirus
DQ226541	Duck hepatitis A virus
KC465954	Avisivirus A
AJ225173	Avian encephalomyelitis virus
M14707	Hepatitis A virus
JF973686	Rosavirus A
JN819202	Cadicivirus A
HM751199	Melegrivirus A
GQ179640	Salivirus A
GU182408	Oscivirus A
AB040749	Aichivirus A
GU182406	Passerivirus A
JQ691613	Gallivirus A
AF406813	Porcine sapelovirus
J02281	Enterovirus C
JF973687	Mosavirus A
JQ941880	Hunnivirus A
AF231769	Porcine teschovirus
X96871	Equine rhinitis B virus
AAT01781	Foot-and-mouth disease virus_SAT 1
AAT01791	Foot-and-mouth disease virus_SAT 2
AAT01793	Foot-and-mouth disease virus_SAT 3
AY593829	Foot-and-mouth disease virus_O
PO33062	Foot-and-mouth disease virus_A
AAT01747	Foot-and-mouth disease virus_C
ABM66096	Foot-and-mouth disease virus_Asia 1
FJ438902	Cosavirus A
JQ814851	Mischivirus A
DQ641257	Seneca Valley virus
M81861	Encephalomyocarditis virus