IBV: Potential as a vaccine vector and identification of a novel subgenomic mRNA

by

Kirsten Bentley

A thesis submitted for the degree of Doctor of Philosophy

University of Warwick

School of Life Sciences

and

The Institute for Animal Health

September 2012

Contents

List of Tables

List of Figures

Acknowledgements

I would first like to express my thanks to Prof Paul Britton for support and guidance throughout this PhD and for allowing me the freedom to expand the project from its original objective. I would also like to thank Dr Maria Armesto for her supervisory assistance, as well as Prof Andrew Easton at the University of Warwick. I am grateful to the BBSRC and IAH for funding. I would particularly like to thank my colleague in the Coronavirus group at the IAH, Sarah Keep. Sarah was responsible for completing the reverse genetics process for generation of rIBV BeauR-L-CTTAACAT used in this study, as well as for titrating each of the TRS mutant rIBVs, and her assistance was invaluable in ensuring the completion of this project on time. I would also like to thank Dr Erica Bickerton for help in proofreading this thesis, and all other past and present members of the Coronavirus lab for their general help and assistance.

I am grateful for the help of members of the IAH Microbiological Services Department, particularly Ruth Hennion and Gillian Hill, for provision of cell cultures. Special thanks go to Dr Nicolas Eterradossi and Dr Olivier Guionie of the French Agency for Food, Environmental and Occupational Heath Safety, Ploufragan-Plouzane Laboratory, for providing the turkey coronavirus RNA samples used in this study.

Finally, I would like to thank my family and friends for their support, encouragement and inspiration over the last four years.

Declaration

I declare that the work presented in this thesis was performed solely by the author and has not been presented previously in support of another degree.

Summary

Using an IBV reverse genetics system a series of recombinant viruses were generated to investigate the potential for utilizing IBV as a vaccine vector. Through the replacement of non-essential regions of the IBV genome, with eGFP or hRluc, factors influencing the stability of recombinant viruses expressing heterologous genes were determined. Expression of heterologous proteins was possible from a variety of virus constructs. The stability of recombinant viruses varied depending on the genome location of the heterologous gene, with replacement of Gene 5 proving to be most stable following passage in cell culture. Stability was strongly influenced by the MOI at which viruses were passaged, with low MOIs resulting in increased stability. The replacement of Gene 5 with a heterologous virus gene may be a suitable target for development of a bivalent vaccine capable of protecting against IBV and a second avian viral disease.

Analysis of recombinant IBV mRNA expression profiles led to an investigation into an uncharacterized RNA species, and its link to the IBV intergenic region. A novel subgenomic mRNA was identified associated with the intergenic region that was shown to be transcribed via a non-canonical transcription regulatory sequence. In contradiction to the current model of coronavirus transcription this mRNA has a transcription regulatory sequence derived mainly from the leader, and not the body, transcription regulatory sequence. The non-canonical sequence was shown to be responsible for reduced transcription levels of the intergenic region mRNA. This project proposes the presence of an additional IBV subgenomic mRNA, transcribed via a non-canonical mechanism, and encoding a novel 5th accessory protein of IBV and closely related gammacoronaviruses.

List of Abbreviations

ADRP	ADP-ribose-1"-monophosphatase
BCA	Bicinchoninic acid
BES	N,N-bis[2-hydroxethyl]-2-Aminoethanesulfonic acid
BHK	Baby hamster kidney
BSA	Bovine serum albumin
CAT	Chloramphenicol acetyltransferase
cDNA	Copy DNA
CEF	Chicken embryo fibroblasts
CK	Chick kidney
CPE	Cytopathic effect
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
D-RNA	Defective RNA
E	Envelope
E. coli	Escherichia coli
e/GFP	Enhanced/green fluorescent protein
EAV	Equine arteritis virus
EDTA	Ethylenediaminetetraacetic acid
EMEM	Eagles minimal essential medium
ERGIC	ER-Golgi intermediate compartment
FCoV	Feline coronavirus
FCS	Fetal calf serum
fLuc	Firefly luciferase

FPV	Feline peritonitis virus
G-MEM	Glasgows minimal essential medium
GPT	Guanine-xanthine phosphoribosyltransferase
hpi	Hours post infection
hRluc	Humanized *Renilla* luciferase
IAH	Institute for Animal Health
IBV	Infectious bronchitis virus
IR	Intergenic region
IRES	Internal ribosome entry site
kDa	Kilo Daltons
LB	Luria broth
M	Membrane
MDV	Mareks disease virus
MHV	Murine hepatitis virus
MOI	Multiplicity of infection
MPA	Mycophenolic acid
mRNA	Messenger RNA
N	Nucleoprotein
NDV	Newcastles disease virus
nsp	Non-structural protein
nt	Nucleotide
ORF	Open reading frame
P	Passage
PBS	Phosphate buffered saline
pfu	Plaque forming unit

PL^{pro}	Papain-like protease
RdRp	RNA dependent RNA polymerase
RFS	Ribosome frameshift
rIBV	Recombinant infectious bronchitis virus
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RTC	Replication transcription complex
rVV	Recombinant vaccinia virus
S	Spike
SARS-CoV	Severe acute respiratory syndrome coronavirus
sg	Subgenomic
SPF	Specific pathogen free
TBE	Tris-borate-EDTA
TCoV	Turkey coronavirus
TDS	Transient dominant selection
TE	Tris-EDTA
TGEV	Transmissible gastroenteritis virus
TM	Transmembrane
TRS	Transcription regulatory sequence
TRS-B	Transcription regulatory sequence – Body
TRS-L	Transcription regulatory sequence – Leader
VLP	Virus-like particle

Chapter 1: Introduction

1.1 Infectious bronchitis virus

The avian coronavirus infectious bronchitis virus (IBV) is a highly pathogenic and contagious virus of domestic fowl that causes disease in chickens of all ages. First described in the United States in the 1930's, infectious bronchitis is now prevalent everywhere there is poultry farming and can be considered a global endemic disease. Primarily causing a respiratory disease, IBV initially infects tissues of the upper respiratory tract although the virus is able to replicate within, and cause disease in, other tissues such as the kidney, oviduct and testes and many parts of the alimentary tract (reviewed in (Cavanagh 2007, Raj and Jones 1997)). Uncomplicated infections in juvenile and mature birds have low mortality rates but other effects of infection, such as growth retardation, are observed. In egg laying birds infection can result in diminished egg production and egg quality (Crinion 1972). There is evidence to suggest that IBV infection can lead to reduced fertility in roosters resulting in reduced breeding capabilities (Boltz*, et al.* 2007, Boltz*, et al.* 2004). As a result of these effects IBV is responsible for huge economic loss within the global poultry industry.

Control of IBV by vaccination, using both live attenuated and inactivated vaccines, has been employed for many years in both commercial and smallholding settings. However, IBV is highly antigenically variable with over twenty different serotypes currently circulating worldwide and low levels of cross-protection between serotypes. Current vaccine strategy for egg laying birds is two, or more, vaccinations

with live vaccine starting at 2-3 weeks of age, followed by vaccination with inactivated vaccine shortly before birds come into lay. Broilers, those birds utilized for meat production, are typically vaccinated at one day of age with live vaccines, and as required subsequently if new serotypes are in circulation, or a threat to a flock (Cavanagh 2007). Production of a protective immune response varies depending on the similarity of the vaccine serotype to that of the challenge serotype and often no protective effect is seen from existing commercially available vaccines against new serotypes. In order to achieve control of the virus, vaccines must therefore be continually developed based on currently circulating strains (Ladman, *et al.* 2002, Lin, *et al.* 2005, Liu, *et al.* 2009).

Infectious bronchitis virus is classified in the genus *gammacoronavirus* within the family *Coronaviridae*, a member of the order *Nidovirales* (Table 1.1). Infectious bronchitis virus particles contain positive-sense, single stranded RNA genomes of approximately 27.6Kb (Lai and Holmes 2001) that are both 5' capped and 3' polyadenylated. Coronaviruses are unique amongst the positive-sense RNA viruses in that they have helical nucleocapsids, formed by the association of the genome with the nucleocapsid (N) protein. These helical nucleocapsids are enclosed within a lipid envelope consisting of the spike (S), membrane (M) and envelope (E) proteins, giving rise to pleiomorphic particles approximately 120nm in diameter with petal-shaped spikes projecting from the surface (Fig. 1.1).

TABLE 1.1: Classification of the *Coronaviridae*

Genus	Species*	
Alphacoronavirus	*Alphacoronavirus 1*	Feline coronavirus (FCoV)
		Porcine transmissible gastroenteritis virus (TGEV)
	Human coronavirus 229E (HCoV-229E)	
	Human coronavirus NL63 (HCoV-NL63)	
	Miniopterus bat coronavirus 1	
	Miniopterus bat coronavirus HKU8	
	Porcine epidemic diarrhoea virus (PEDV)	
	Rhinolophus bat coronavirus HKU2	
	Scotophilus bat coronavirus 512	
Betacoronavirus	*Betacoronavirus 1*	Bovine coronavirus (BCoV)
		Human coronavirus OC43 (HCoV-OC43)
		Human enteric coronavirus (HECoV)
		Porcine haemagglutinating encephalomyelitis virus (HEV)
		Equine coronavirus
	Human coronavirus HKU1 (HCoV-HKU1)	
	Murine coronavirus	Murine hepatitis virus (MHV)
		Puffinosis coronavirus
		Rat coronavirus (RtCoV)
	Pipistrellus bat coronavirus HKU5	
	Rousettus bat coronavirus HKU9	
	Severe acute respiratory syndrome-related coronavirus (SARS-CoV)	
	Tylonycteris bat coronavirus HKU4	
Gammacoronavirus	*Avian coronavirus*	Infectious bronchitis virus (IBV)
		Turkey coronavirus (TCoV)
		Pheasant coronavirus (PhCoV)
		Goose coronavirus
		Pigeon coronavirus
		Duck coronavirus
	Beluga whale coronavirus SW1 (BeCoV)	
Deltacoronavirus	*Bulbul coronavirus HKU11 (BuCoV)*	
	Munia coronavirus HKU13 (MunCoV)	
	Thrush coronavirus HKU 12(ThCoV)	

* The coronavirus species are listed as according to the International Committee on Taxonomy of Viruses – 2011 Release, for family *Coronaviridae.*

A **B**

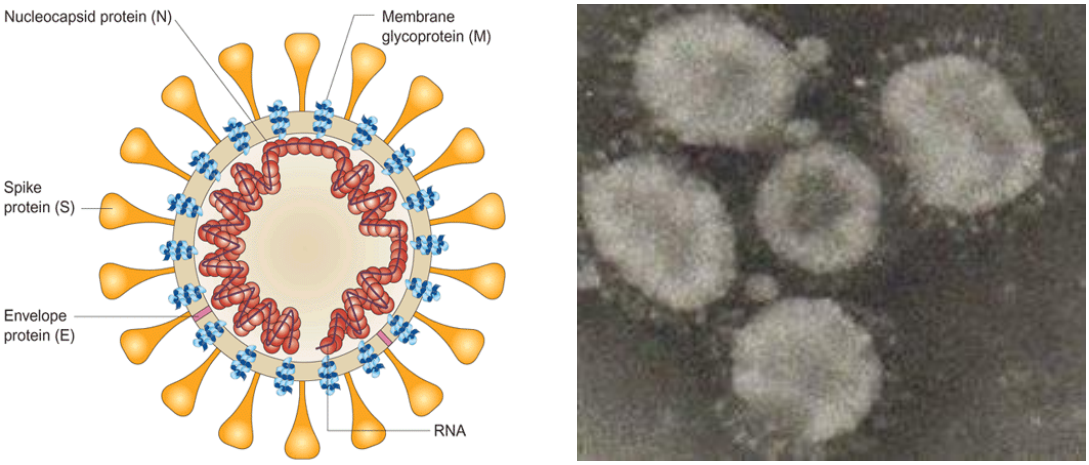


Fig. 1.1. Infectious bronchitis virus particles. (A) Schematic representation of a coronavirus particle identifying locations of the structural proteins (Stadler, *et al.* 2003). (B) Electron micrograph of IBV particles (sourced from www.infectious-bronchitis.com).

1.2 Genome organisation and replication

The IBV genome contains multiple open reading frames (ORFs) as shown in Fig. 1.2. The 5'-most two-thirds of the genome encodes the non-structural proteins involved in virus transcription and replication while the structural proteins S, E, M and N, are encoded within the 3'-most third of the genome, along with the group-specific accessory proteins 3a/3b and 5a/5b. At the very 5'-end of the genome is a ~60 nucleotide (nt) region known as the Leader, a feature common to all coronaviruses that will be discussed in the context of subgenomic (sg) mRNA transcription.

Fig. 1.2. Schematic of IBV genome. The replicase gene is shown at the 5'-end of the genome and the junction at which the ribosome frameshift occurs is highlighted (RFS). Structural and accessory proteins are shown at the 3'-end. Leader sequence (L) is shown upstream of ORF1a and uncharacterized intergenic region (IR) is located between the M gene and Gene 5.

1.2.1 Attachment and entry

The life cycle of IBV is similar to all other coronaviruses and is shown schematically in Fig. 1.3. Infectious bronchitis virus initially infects cells of the upper respiratory tract and can subsequently spread to other tissues. However, the cellular receptor to which IBV binds has not yet been identified; although there is evidence to suggest that sialic acid, specifically α2,3-linked sialic acid, plays an important role in infection (Winter, *et al.* 2008, Winter, *et al.* 2006). Binding to this receptor alone does not explain the tissue tropism of different IBV strains however, and it is likely that another receptor or protein is also involved in initiating infection.

The S protein mediates attachment to the host cell and entry has been demonstrated to involve a low-pH-dependent fusion activation event (Chu, *et al.* 2006). The exact process of virus entry is currently unclear but results in the release of the nucleocapsid into the cytoplasm, where host cell ribosomes translate genomic RNAs to produce the non-structural proteins (nsps), encoded in Gene 1, necessary for replication of the viral genome.

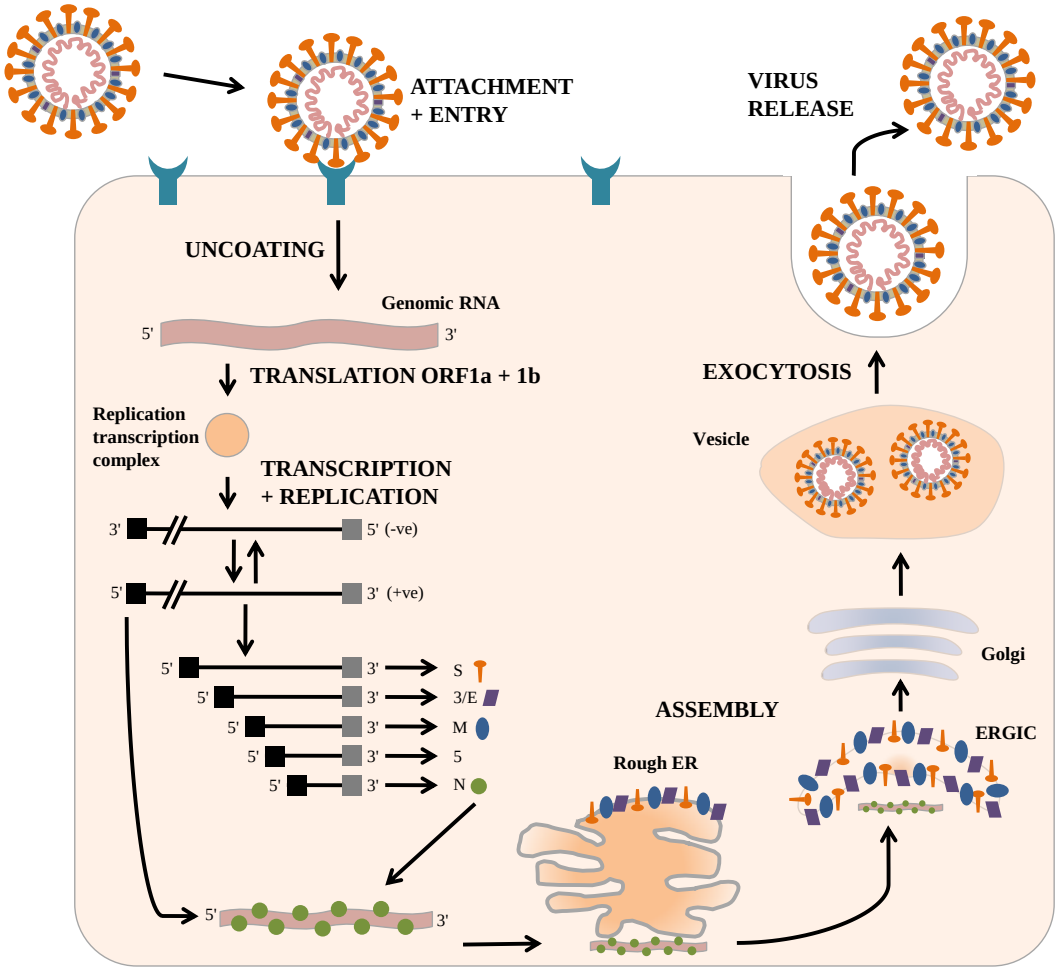


Fig. 1.3. Schematic representation of the coronavirus life cycle. The main steps of the coronavirus life cycle are shown. Attachment of the virus particle to the host cell is mediated by the virus spike protein. The virus entry pathway, which may involve a process of endocytosis or direct membrane fusion, has not been fully elucidated but results in the genomic RNA being deposited into the cytoplasm. Translation of ORF1a and ORF1b yields the nsps that form the replication transcription complex and genome replication and sg mRNA synthesis proceeds. Progeny virus particles are assembled at the ER-Golgi Intermediate Compartment (ERGIC) and are released from the cell via exocytosis.

1.2.2 The non-structural proteins

Gene 1 consists of two ORFs designated 1a and 1b. ORF1a and ORF1b briefly overlap and it is at this junction that a -1 ribosomal frameshift event occurs resulting in production of the replicase polyproteins pp1a and pp1ab (Brierley, *et al.* 1987). Virus-encoded proteases are responsible for the cleavage of the 15 nsps (nsps 2-16)

encoded for by the IBV replicase polyproteins (Fig. 1.4). The number of nsps of IBV differs from the majority of other coronaviruses, which encode 16 nsps. The nsps encoded by ORF1b are produced 60-80% less than proteins of ORF1a indicating the frameshift event to be a method of regulating protein production.

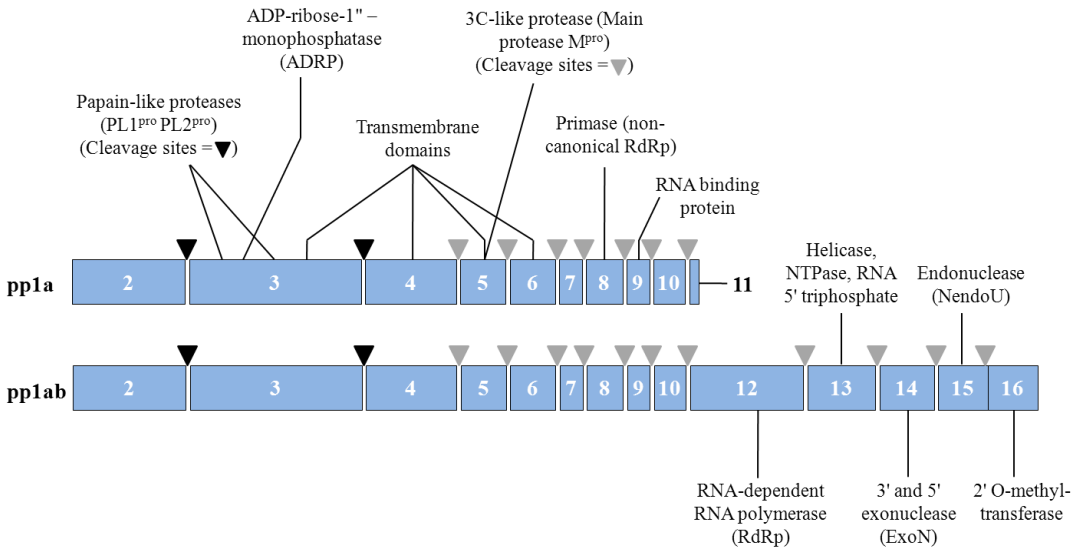


Fig. 1.4. Domain organisation of IBV pp1a and pp1ab. Non-structural proteins are numbered from 2 to 16 with known functions labelled. Nsp1 is not present in IBV. The coding sequence of nsp11 overlaps with elements involved in ribosomal frameshifting and is therefore not shown in pp1ab. Cleavage sites for PL2^{pro} and M^{pro} are indicated by black and grey arrowheads respectively. The PL1^{pro} domain is proteolytically inactive in IBV.

As indicated in Fig. 1.4 the nsps have been shown to play a number of roles in virus replication and transcription, although not all have been fully characterized. One major difference is observed between the gammacoronaviruses and other coronavirus genera and relates to nsp1 and the PL1^{pro} domain of nsp3. The nsp1 ORF, suggested to be involved in controlling host mRNA synthesis (Huang, *et al.* 2011, Tanaka, *et al.* 2012) is not present in IBV. Additionally, the PL1^{pro} domain of

nsp3, responsible for nsp1/nsp2 cleavage and missing in SARS-CoV (Rota, *et al.* 2003), is considered proteolytically inactive due to the absence of some sequences essential for proteolytic activity. It is possible that this is a result of co-evolution; the lack of nsp1 in IBV means that $PL1^{pro}$ is dispensable, as the main substrate of $PL1^{pro}$, the nsp1/nsp2 cleavage site, does not exist.

As well as the papain-like proteases nsp3 contains an ADP-ribose-1"-monophosphatase (ADRP) domain that dephosphorylates ADP-ribose-1"-monophosphate, a side product of cellular tRNA splicing, to ADP-ribose (Putics, *et al.* 2005). This may function to control the cellular concentrations of certain intracellular metabolites and thereby regulate cell metabolism during infection. The presence of the transmembrane (TM) domains of nsp3, 4 and 6 suggests a role for these proteins in determining membrane rearrangements, an essential function of replication for not only coronaviruses but also many other positive-sense RNA viruses (Baliji, *et al.* 2009, Kanjanahaluethai, *et al.* 2007, Oostra, *et al.* 2007).

While nsp12 was identified as the main viral primer-dependent RNA-dependent RNA polymerase (RdRp) in early studies of IBV (Gorbalenya, *et al.* 1989), nsp8 has more recently been proposed as a primase with RdRp activity capable of *de novo* initiation and primer extension. Interestingly nsp8 has only ever been crystallized in a complex with nsp7; while for SARS-CoV a hexadecameric structure consisting of 8 copies of each nsp was demonstrated, for FCoV a heterotrimer of 2 copies of nsp7 and 1 copy of nsp8 was observed, and the precise nature of this relationship is therefore still unclear (te Velthuis, *et al.* 2012, Xiao, *et al.* 2012). The function of nsp9 is unclear but has been shown to have RNA binding properties and is known to

form a dimer that is essential to virus replication (Miknis, *et al.* 2009, Sutton, *et al.* 2004). Studies of the SARS-CoV nsp13 helicase demonstrated both RNA and DNA duplex unwinding activities, (deoxy) nucleoside triphosphatase (dNTPase) activity and a RNA 5'-triphosphatase activity that might play a role in the formation of the 5' cap structure of viral RNAs (Ivanov, *et al.* 2004).

The finding that nsp14 includes 3'- to 5'-exonuclease (ExoN) motifs suggested a proofreading mechanism for coronaviruses that may help explain how the large RNA genomes maintain fidelity during replication (reviewed in (Denison, *et al.* 2011)). It has been shown that interaction of nsp14 with nsp10 forms a complex that is involved in 3'-end mismatch excision and results in a >35-fold more potent exoribonuclease (Bouvet, *et al.* 2012). Additionally, nsp14 has been characterized as a novel guanine-N7-methyltransferase (N7-MTase) involved in the capping of viral mRNAs (Chen, *et al.* 2009). The nsp15 is highly conserved across the *Nidovirales* and has been demonstrated in a number of coronaviruses, including IBV, to be a manganese-dependent endoribonuclease (termed NendoU) that can cleave both single- and double-stranded RNA and may play a role in cellular RNA processing during infection (Bhardwaj, *et al.* 2004). Finally, nsp16 has been demonstrated to be a 2'*O*-methyltransferase that binds capped RNAs carrying a methyl group at the N7 position of the guanosine cap, thus functioning in conjunction with nsp14 as part of the coronavirus RNA capping mechanism (Decroly, *et al.* 2008).

Once all of the replicase proteins have been produced, virus replication continues with the transcription of negative-sense copies of the genomic RNA that are subsequently used as templates for transcription of positive-sense copies of the virus

genome. Additionally, a nested set of mRNAs is produced for the structural and group-specific genes via a process of discontinuous transcription as described below.

1.2.3 Subgenomic mRNA synthesis

During the coronavirus replication cycle only the nsps from Gene 1 are translated directly from the genomic RNA, with all other proteins expressed from a 3'- and 5'-co-terminal nested set of sg mRNAs. A model proposed by Sawicki and Sawicki (Sawicki and Sawicki 1995) has led to the general acceptance that these sg mRNAs are synthesised via a process of discontinuous transcription during negative-strand synthesis (Fig. 1.5).

Located at the distal end of the 5' Leader sequence, all coronaviruses have a conserved sequence known as a transcription regulatory sequence (TRS). This short AU rich sequence, approximately 5 to 8 nucleotides in length depending on the coronavirus, is an important aspect of sg mRNA transcription. An exact copy of the Leader TRS (TRS-L), or a highly similar sequence, is also located upstream of each structural and group-specific gene of the virus and is referred to as a body TRS (TRS-B). For IBV the TRS is recognized to be the highly conserved sequence CUUAACAA. However, there are exceptions to this such as the S and Gene 3 TRSs of Beaudette strains, which have the sequence CUGAACAA, and the S and Gene 5 TRSs of the M41 strain, which have the sequences CUGAACAA and CUUAAUAA respectively. This demonstrates that there is some flexibility in the TRS-B requirements for different genes although all are still easily recognisable as the TRS for the associated gene.

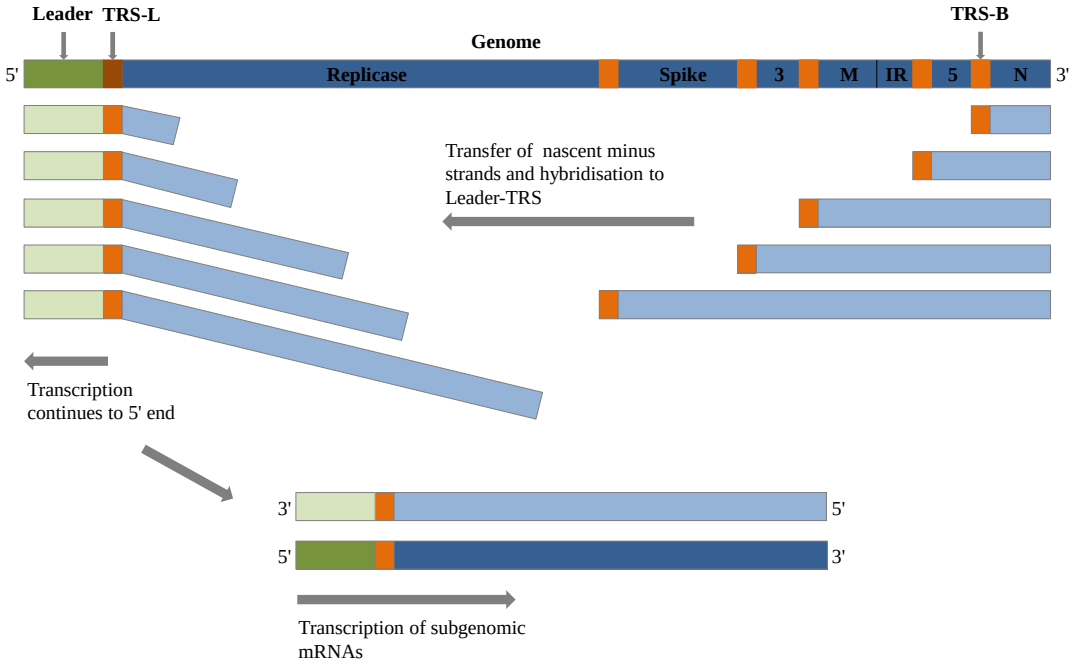


Fig. 1.5. Schematic representation of IBV sg mRNA synthesis. Following entry ORF1a/b is translated to yield the replicase proteins. These replicase proteins use the genome as a template to generate full-length negative-sense RNAs that are then used as templates to generate new genomes. A process of discontinuous transcription yields each of the structural and accessory proteins as a nested set of sg RNAs. Transcription is initiated at the 3'-end of the genome and continues until a transcriptional regulatory sequence is reached. The nascent minus-strand is then translocated to the now complementary 5' TRS-L where it hybridizes and transcription continues to the 5'-end of the genome. Each sg RNA then acts as a template for sg mRNA synthesis.

Synthesis of negative-sense sg RNAs is initiated at the 3'-end of the genome and proceeds until a TRS-B is encountered. The TRS-B is proposed to act as a pause/stop signal for the replication transcription complex (RTC) and once copied one of two possible outcomes will arise; either transcription will continue to the next TRS-B or, the now complementary TRS-B of the nascent negative-strand RNA will hybridize to the TRS-L and transcription will continue to the very 5'-end of the genome, adding the Leader sequence to each sg RNA. The negative-sense sg RNAs, with co-terminal 5'- and 3'-ends, are then transcribed into a nested set of sg mRNAs.

The precise mechanisms of sg mRNA synthesis are unknown, and while the TRS is known to be important (Pasternak, *et al.* 2001, Pasternak, *et al.* 2003, van Marle, *et al.* 1999, Zuniga, *et al.* 2004), there is increasing evidence to suggest a number of additional factors might also be involved. These may include specific sequence elements and RNA-RNA interactions (Mateos-Gomez, *et al.* 2011, Moreno, *et al.* 2008), the formation of secondary RNA structures such as stem-loops (Dufour, *et al.* 2011, Ozdarendeli, *et al.* 2001, Yang, *et al.* 2011) or RNA-protein interactions (Baric, *et al.* 1988, Keane, *et al.* 2012, Sola, *et al.* 2011, Zhang and Lai 1995, Zuniga, *et al.* 2010).

Once synthesised the sg mRNAs are translated by host cell ribosomes to produce the structural and group-specific accessory proteins. With the exception of the smallest sg mRNA (N), all other sg mRNAs contain two or more ORFs however, generally only the 5'-most ORF is subsequently translated to yield the associated protein.

1.2.4 Structural proteins

Infectious bronchitis virus encodes four structural proteins, S, E, M and N, as shown in Fig. 1.1. The largest of these, S, is a multifunctional type 1 glycoprotein involved in binding to the host cell receptor and mediating virus-cell and cell-cell fusion. Mature S proteins assemble into a trimer to form the spikes seen on the surface of the virus (Delmas and Laude 1990) and are incorporated into virus particles at the ER-Golgi Intermediate Compartment (ERGIC), the site of coronavirus budding (Klumperman, *et al.* 1994). The S1 subunit of the protein mediates receptor binding and a conformational change in the protein is triggered. The S2 subunit, containing two heptad repeat regions, is responsible for the class I virus fusion protein activity

of the spike (Bosch, *et al.* 2003, Yamada, *et al.* 2009) and is therefore thought to be involved in initial fusion of virus to the cell membrane following attachment, as well as playing a role in cell-cell spread.

As well as a role in attachment the S1 subunit is involved in the induction of a protective immune response (Cavanagh, *et al.* 1986, Ignjatovic and Galli 1994). However, S1 sequences are highly variable and this has led to extensive variation. As a result numerous serotypes exist in the field which are generally poorly cross-protective (Cavanagh, *et al.* 1997).

The coronavirus E protein is a small integral membrane protein (76-109 amino acids) and, in the case of IBV, is translated from the third ORF (ORF 3c) of a tri-cistronic Gene 3 sg mRNA via an internal ribosome entry site (IRES) (Liu and Inglis 1992). Although a minor structural component of the viral envelope, the E protein is thought to play a role in inducing membrane curvature and, if co-expressed with the M protein leads to the formation of virus-like particles (VLPs) (Baudoux, *et al.* 1998, Corse and Machamer 2000, Raamsman, *et al.* 2000). The requirement for E in a productive infection appears to vary amongst coronaviruses; proving essential for the production of recombinant infectious TGEV virions (Curtis, *et al.* 2002), while for both MHV and SARS-CoV recombinant viruses lacking E were shown to replicate in cell culture, albeit to reduced levels compared to recombinant wild type viruses (DeDiego, *et al.* 2007, Kuo and Masters 2003). There is also evidence that the E protein has ion channel activity, although the exact role of this activity during infection is currently unclear (Wilson, *et al.* 2006).

The M protein, a type III membrane protein, is the most abundant protein in the viral envelope, estimated to account for ~40% of the mass of IBV particles (de Haan, *et al.* 1998, Stern, *et al.* 1982) and plays a major role in virus assembly and budding. As mentioned above, M interacts with E leading to the formation of VLPs with all domains of M shown to be important for VLP formation (Corse and Machamer 2003, de Haan, *et al.* 1998). Additionally, M-M interactions, particularly at TM domains, have been shown to be essential for envelope formation (de Haan, *et al.* 2000). Studies carried out with MHV have also shown that the M protein interacts with the N protein, in conjunction with genome length RNAs, suggesting a mechanism by which genomes are specifically packaged into virus particles during infection (Narayanan, *et al.* 2000).

The nucleocapsid protein, N, is a multifunctional phosphoprotein (Calvo, *et al.* 2005) involved in packaging the genomic RNA into a helical ribonucleoprotein, a unique occurrence amongst positive-strand RNA viruses. A number of additional roles of the N protein have also been demonstrated including regulation of viral RNA synthesis (Baric, *et al.* 1988, Grossoehme, *et al.* 2009, Keane, *et al.* 2012, Nelson, *et al.* 2000), likely mediated by the known RNA-binding domains of the protein (Chang, *et al.* 2006). The importance of the N protein was highlighted with the development of reverse genetics systems for coronaviruses when it was demonstrated that the N protein was either essential for the recovery of infectious virions, as in the case of IBV (Casais, *et al.* 2001), or resulted in substantially enhanced recovery of infectious virions (Almazan, *et al.* 2004, Yount, *et al.* 2000, Yount, *et al.* 2003).

1.2.5 The group-specific accessory proteins

In addition to genes for the structural proteins, all coronaviruses possess a varying number of group-specific genes at the 3'-end of the genomic RNA, interspersed between the structural genes. These group-specific genes encode a variety of proteins generally referred to as accessory proteins due to the fact that they can be deleted with limited effects observed on virus replication in cell culture (Casais, *et al.* 2005, de Haan, *et al.* 2002, Frieman, *et al.* 2006, Hodgson, *et al.* 2006, Ortego, *et al.* 2003, Youn, *et al.* 2005, Yount, *et al.* 2005). Although relatively few coronavirus accessory proteins have been characterized to date, the evidence suggests that these proteins have immunomodulatory roles thereby influencing virus pathogenicity and virulence (Cruz, *et al.* 2011, de Haan, *et al.* 2002, Freundt, *et al.* 2010, Frieman, *et al.* 2007, Koetzner, *et al.* 2010, Varshney, *et al.* 2012).

Infectious bronchitis virus possesses two group-specific genes, Gene 3 and Gene 5, each of which encode two accessory proteins designated 3a and 3b, and 5a and 5b respectively (Liu, *et al.* 1991, Liu and Inglis 1992). At present no functions have been determined for any of the four IBV accessory proteins although all have been shown to be dispensable for virus replication in cell culture (Casais, *et al.* 2005, Hodgson, *et al.* 2006, Youn, *et al.* 2005).

1.2.6 The intergenic region

In addition to the structural and group-specific genes IBV, and closely related gammacoronaviruses such as turkey coronavirus (TCoV), also contain an uncharacterized region located between the M gene and Gene 5, referred to variably as the IR, ORF4b, or ORF X (Armesto, *et al.* 2009, Cao, *et al.* 2008, Gomaa, *et al.*

2008, Hewson, *et al.* 2011). With the exception of some vaccine and attenuated laboratory strains of IBV this region, of ~350nts, contains a putative ORF with the potential to code for a protein of 94 amino acids. The AUG initiation codon of this potential protein is located directly downstream of the M gene stop codon.

The chick kidney (CK) cell adapted IBV strain, Beau-CK, from which the Beau-R cDNA clone used in the Coronavirus Group at the Institute of Animal Health (IAH) was derived, contains a large deletion of ~40 amino acids from the 3'-end of the potential coding sequence. Smaller truncations are observed for some other strains of IBV making it likely that any potential protein from this ORF is inactive in these strains (Fig. 1.6). Such deletions and truncations also indirectly demonstrate that the IR is dispensable for virus replication.

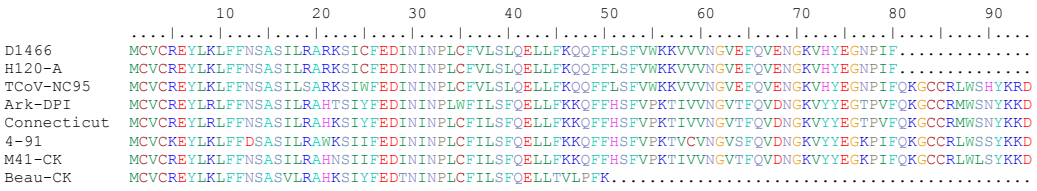


Fig. 1.6. Amino acid alignments of gammacoronavirus IR ORF. The amino acid alignments are shown for the putative 94 amino acid protein coding sequence of the IR. A selection of strains including Beau-CK and M41-CK, as used in the Coronavirus Group at the IAH, are shown to demonstrate the deletions and truncations present in some strains.

Despite the presence of the IR ORF the function of this region has always been debated due to the lack of identification of a TRS upstream of the IR. The presence of a TRS is believed to be a fundamental requirement for the expression of sg mRNAs and therefore subsequent protein expression. For IBV the TRS-Bs of the

known genes are located ~20-100nts upstream of the AUG for the gene and despite extensive scans of the IBV genome no sequence closely matching the required CUUAACAA TRS has been identified upstream of the IR. The identification of the IR in TCoV (as ORF X) was accompanied by a proposed TRS-B of GUCAACAA located 288nts upstream of the initiation codon (Gomaa, *et al.* 2008). However, there was no experimental data to back-up this proposal and the sequence mis-matches and distance from the AUG make it an unlikely target. The function of this region therefore remains unclear.

1.2.7 Virus assembly and release

Coronaviruses, as with many positive strand RNA viruses, are known to induce a variety of membrane rearrangements to support virus replication complexes including double-membrane vesicles and convoluted membranes (Knoops, *et al.* 2008, Ulasli, *et al.* 2010). As well as providing a scaffold for replication these membrane rearrangements may offer a mechanism by which the replicating viruses can evade host immune responses by shielding immunogenic molecules, such as double-stranded RNA, from the host immune system. The association of replication with membranes may also be beneficial for assembly as all the components of progeny virions are concentrated together ready for incorporation into new virions, although it is not clear how each of the structural proteins are associated with the membranes.

Assembly begins with the encapsidation of newly synthesised genomes, by numerous copies of the N protein, to form helical nucleocapsids. Interactions between N, M and the genomic RNA, as previously described, at the endoplasmic

reticulum and Golgi body mediate the selective packaging of new genomes into virions. The final stages of assembly and budding of new virions has been demonstrated to occur at cytoplasmic surfaces of the ERGIC (Klumperman, *et al.* 1994) with virions subsequently transported in membrane-bound vesicles to the plasma membrane where they are released from the cell by a process of exocytosis (Tooze, *et al.* 1987).

1.3 Viral-vector vaccines

For many years viruses, particularly members of the *Adenoviridae*, *Poxviridae* and *Retroviridae* families, have been studied for their potential to be utilized as gene delivery systems, for example in the treatment of certain types of cancer, or as viral-vector vaccines for protection against human and veterinary infectious diseases (reviewed in (Draper and Heeney 2010)). As well as mammalian viruses, some avian viruses such as the paramyxovirus newcastle disease virus (NDV) and the avipoxvirus fowlpox virus (FPV) have been evaluated for use as vectors, not only as vaccines for veterinary use (Ge, *et al.* 2011, Shi, *et al.* 2011) but also for vaccine and gene therapy use in humans ((DiNapoli, *et al.* 2007, Nakaya, *et al.* 2001, Tsang, *et al.* 2005) and reviewed in (Skinner, *et al.* 2005, Weli and Tryland 2011)).

Often the virus backbone simply provides a delivery mechanism for the gene of interest. However, particularly in veterinary situations, the desire is to generate a viral-vector vaccine that is bi- or polyvalent and can be used to vaccinate against two or more diseases at the same time. This desire is particularly true in the field of avian diseases and commercial poultry enterprises. Birds in commercial flocks may require

vaccination against up to 6 or 7 different viral diseases, often with multiple doses required, and any ability to reduce this number through combining vaccines will have enormous benefits for both animal welfare and commercial interests. There has already been success in this field with the production of several vector vaccines including Merial's VAXXITEK® HVT+IBD vaccine, which vaccinates against both Mareks disease virus (MDV) and infectious bursal disease virus, and the Vectormune® HVT-NDV vaccine from Ceva, which protects against both MDV and NDV. The MDV HVT vaccine offers many possibilities for use as a vector to also immunize against additional avian diseases and research is ongoing in this area (Iqbal 2012).

1.3.1 Reporter genes

To be successful, viral-vector vaccines must be capable of stably expressing foreign genes and this ability can be investigated by using recombinant viruses expressing reporter genes. When expressed from an organism a reporter gene allows for a measurable output of gene expression, with a wide variety available to suit different needs. A number of bacterial genes such as the *E. coli* enzyme beta-glucoronidase (GUS), the chloramphenicol resistance gene chloramphenicol acetyltransferase (CAT) and the *E. coli* lacZ gene were developed into successful reporter gene systems. However, these systems have been overshadowed in the study of viral vectors by the popularity of fluorescent and luminescent proteins.

The fluorescent protein green fluorescent protein (GFP), first isolated in 1962 from the jellyfish *Aequorea victoria* (Shimomura, *et al.* 1962), has become one of the most widely used markers for gene expression. The *A. victoria* GFP is a 27kDa

protein of 238 amino acids that absorbs blue light (maximally at 395nm) and emits green light with a peak emission of 509nm (Prasher, *et al.* 1992). This process can be reproduced experimentally by exposure to a source of UV light, such as from a fluorescence microscope. One of the advantages of GFP is that to achieve fluorescence no additional substrates or cofactors are required, thus allowing detection of expression in living cells and tissues. However, GFP is also species specific and it was not until the development of 'humanised' and 'enhanced' GFPs (hGFP and eGFP), which were modified for increased translation in mammalian cells, that GFP could be fully exploited (Zhang, *et al.* 1996, Zolotukhin, *et al.* 1996).

Luminescent proteins are enzymes, referred to as a luciferase, which catalyse a light-emitting reaction of its associated substrate, known as luciferin. Isolated from the sea pansy *Renilla reniformis*, the *Renilla* luciferase is a protein of 311 amino acids that catalyses the oxidation of coelenterazine to yield blue light of 480nm, which can then be detected using a luminometer (Lorenz, *et al.* 1991). *Renilla* luciferase, with a molecular weight of 37kDa, originally showed some limitation in use compared to other luciferases, such as the larger (~60kDa) firefly luciferase (fLuc), due to low levels of autoluminescence, which reduced assay sensitivity. However, as with GFP, *Renilla* luciferase has also been modified for improved expression in mammalian cells in an effort to increase the reliability as a genetic reporter and reduce levels of autoluminescence. This was achieved by redesigning the gene sequence using preferred mammalian codons and generating a synthetic *Renilla* gene, hRluc (Zhuang, *et al.* 2001). There have also been improvements to substrates for *Renilla* luciferase allowing it to also be used in living cells and tissues.

1.4 Coronaviruses and the expression of heterologous genes

The large size of coronavirus genomes, combined with the possibility of expressing heterologous genes via the generation of novel sg mRNAs, has meant that coronaviruses have long been attractive targets for use as viral-vector vaccines. However, the expression of heterologous genes requires considerable manipulation of the virus genome and thus requires the development of a reverse genetics system for the virus. Initially the development of coronavirus reverse genetics systems were hampered; partly due to the large size of coronavirus genomes but also because of the instability of certain regions of the replicase gene when expressed from bacteria. In the absence of reverse genetics systems expression of heterologous genes from coronaviruses was limited to artificial constructs such as defective RNAs (D-RNAs) (Stirrups, *et al.* 2000), replicon particles (Curtis, *et al.* 2002) or expression vectors (Thiel, *et al.* 2003), which although useful as tools to study aspects of the coronavirus life cycle, were not always suitable for use as vaccine vectors.

In recent years the problems associated with coronavirus reverse genetics systems have been overcome and a number of strategies have been successfully developed to produce full-length cDNAs from several coronaviruses including TGEV, human coronavirus 229E, SARS-CoV and human coronavirus NL63 (Almazan, *et al.* 2000, Donaldson, *et al.* 2008, Thiel, *et al.* 2001, Yount, *et al.* 2000, Yount, *et al.* 2003). A reverse genetics system for IBV, utilizing vaccinia virus, has also been established and so made it possible to investigate the use of recombinant IBVs (rIBVs) for vaccine development (Casais, *et al.* 2001). Using vaccinia virus as a vector for IBV cDNA it is possible to make large-, or small-scale modifications to the IBV cDNA

directly by utilizing the natural ability of vaccinia virus to undergo homologous recombination (see Chapter 2 section 2.10 for more details) (Britton, *et al.* 2005). These advances in reverse genetics have made it possible to realize the potential of using coronaviruses as vaccine vectors, as well as gaining the ability to generate recombinant viruses that can be used as tools to study various aspects of the coronavirus life cycle through the expression of reporter genes.

To investigate the potential use of coronaviruses as vaccine vectors, and as tools to investigate virus replication, studies have often focussed on how to generate recombinant viruses that can stably express heterologous genes. Utilizing reporter genes such as GFP/eGFP, fLuc or hRluc allows for visual analysis and determination of the factors that influence the stability of recombinant viruses. Of particular interest is the position at which heterologous genes are inserted into the virus genome. Due to the nature of coronavirus sg mRNA transcription two possibilities present themselves for the expression of heterologous genes: (1) expression from an additional, newly synthesised sg mRNA or, (2) expression from an existing sg mRNA via insertion of a gene cassette, replacement of existing genes or fusion to existing genes.

Expression of heterologous genes via the synthesis of additional sg mRNAs was demonstrated with the use of expression systems such as minigenomes and D-RNAs (Alonso, *et al.* 2002, Stirrups, *et al.* 2000). However, when using recombinant viruses consisting of full-length virus genomes stability was often greatest following the replacement of the non-essential accessory genes (de Haan, *et al.* 2003, Shen, *et al.* 2009, Tekes, *et al.* 2008, Youn, *et al.* 2005). Alterations to coronavirus structural

genes appear to be less well tolerated with reduced stability, or an inability to rescue infectious virus (Bosch, *et al.* 2004, Shen, *et al.* 2009). Differences in recombinant virus stability have also been identified depending on the particular coronavirus genomic background, as well as the type of heterologous gene being inserted (de Haan, *et al.* 2005).

In the case of the closely related arteriviruses the genomes do not contain genes that are dispensable for virus replication, as with the coronaviruses, and therefore gene replacement has not been a viable option for the expression of heterologous genes. However, the successful rescue of infectious recombinant viruses expressing heterologous genes has been achieved for the arterivirus equine arteritis virus (EAV) by synthesis of novel sg mRNAs inserted within the structural genes (de Vries, *et al.* 2001), or the replicase gene (van den Born, *et al.* 2007), with the latter proving to be the more stable construct.

Few studies have analysed the potential of coronavirus vectors beyond aspects of stability, although the generation of antigen-specific immune responses has been demonstrated, thus providing further evidence that coronaviruses have the potential to be utilized as vaccine vectors (Cervantes-Barragan, *et al.* 2010, Sola, *et al.* 2003). Taking the concept of coronavirus vaccine vectors one step further the generation of a TGEV-based vector expressing the rotavirus VP7 protein resulted in detection of rotavirus-specific antibodies, as well as demonstrating partial protection against rotavirus-induced disease when used to immunize mice against rotavirus challenge (Ribes, *et al.* 2011). One previous report also suggested the possibility of utilizing

the human coronavirus 229E as a multigene vaccine for HIV although the outcomes of this proposed work are unknown (Eriksson, *et al.* 2006).

1.5 Aims

The use of viral-vector vaccines in the poultry industry has already proved to be a possibility with successful vaccines from veterinary pharmaceutical companies on the market. However, the number of diseases for which poultry require vaccination against warrants the research and development of more bivalent, or polyvalent, viral-vector vaccines. Coronaviruses have been suggested as good candidates for the generation of vectors for the expression of heterologous genes. The avian coronavirus IBV may be suitable for the generation of viral-vector vaccines to protect against not only IBV but also additional avian diseases based on the heterologous genes expressed. The initial aim of this thesis therefore, was to investigate the potential for using IBV as a vector for the expression of heterologous genes. This was to be achieved through the generation of a series of recombinant viruses, expressing reporter genes, which could be used to test the following four hypotheses:

1) Expression of heterologous genes from different locations within the IBV genome would result in recombinant viruses with varying degrees of genome stability.
2) Expression of heterologous genes of different sizes, and with different characteristics, would alter the stability of the resultant recombinant viruses.
3) Codon optimization of the heterologous gene sequence would improve the stability of the recombinant viruses.

4) Fusion of the heterologous gene to an existing virus gene would result in improved stability of recombinant viruses.

Due to observations made during the analysis of rIBVs for the vaccine vector project, relating to the uncharacterized intergenic region of IBV, a second project was initiated to investigate the findings. This led to additional aims being defined for this thesis:

1) To confirm that a novel sg mRNA corresponding to the IBV intergenic region is synthesised during infection.
2) To determine the transcription mechanism for this novel sg mRNA.

Chapter 2: Materials and Methods

2.1 Cells

All cell cultures were prepared, and stocks maintained, by the Microbiological Services department of the Institute for Animal Health, Compton. Cell cultures were maintained at 37°C, 5% CO_2.

CK cells: Chick kidney (CK) cells are a primary cell line generated by the removal and trypsinization of kidneys from 2- to 3-week old specific-pathogen-free (SPF) Rhode Island Red chickens.

Vero cells: Vero cells are a continuous cell line originally derived from kidney epithelial cells of the African Green Monkey.

BHK-21 cells: Baby hamster kidney cells (BHK-21) cells are a continuous cell line originally derived from the kidneys of 5 1-day old Syrian golden hamsters.

CEF cells: Chicken embryo fibroblasts (CEF) are a primary cell line prepared from 9-day old embryonated eggs from SPF Rhode Island Red chickens. Embryos are decapitated, eviscerated and washed in PBS. Individual cells are generated using 0.25% trypsin in PBS. Cells are spun and resuspended in growth medium.

2.2 Cell culture media

A variety of cell media were used for maintenance of different cell lines and plaque assays as listed in Tables 2.1 – 2.4. All recipes are made to 100mL final volume with sterile water.

TABLE 2.1: 1X BES CK cell maintenance medium
2X BES IBV plaque assay medium

Ingredient	Volume (mL) (1X/2X)
10X EMEM (Sigma)	10/20
Tryptose Phosphate Broth	10/20
10% BSA (Sigma)	2/4
1M N,N-bis[2-hydroxethyl]-2-Aminoethanesulfonic acid (BES) (Sigma)	2/4
7.5% Sodium bicarbonate	2.8/5.6
L-glutamine (100X) (Life Technologies)	1/2
Nystatin (100,000 U/mL) (Sigma)	0.25/0.5
Penicillin/Streptomycin (100,000 U/mL) (Sigma)	0.1/0.2

TABLE 2.2: 2X EMEM vaccinia virus plaque assay medium

Ingredient	Volume (mL)
10X EMEM (Sigma)	20
FCS (Sigma)	10
7.5% Sodium bicarbonate	4.6
L-glutamine (100X) (Life Technologies)	2
Nystatin (100,000 U/ml) (Sigma)	0.5
Penicillin/Streptomycin (100,000 U/mL) (Sigma)	0.2

TABLE 2.3: 1X GMEM BHK-21 maintenance medium

Ingredient	Volume (mL)
G-MEM	88.4
FCS (Sigma)	1
Tryptose Phosphate Broth	10
Nystatin (100,000 U/mL) (Sigma)	0.5
Penicillin/Streptomycin (100,000 U/mL) (Sigma)	0.1

TABLE 2.4: 1X 199 CEF maintenance medium

Ingredient	Volume (mL)
10X 199 (Sigma)	10
Tryptose Phosphate Broth	10
7.5% Sodium bicarbonate	3
L-glutamine (100X) (Life Technologies)	1
Nystatin (100,000 U/mL) (Sigma)	2.5
Penicillin/Streptomycin (100,000 U/mL) (Sigma)	0.1
NBCS (New born calf serum) (Sigma)	2

2.3 Viruses

Beau-R: Beau-R is a molecular clone of Beaudette-CK (Beau-CK); a CK cell-adapted strain of Beaudette generated by multiple passages in embryonated hens' eggs and CK cells. Compared to Beau-CK, Beau-R contains two nucleotide substitutions: (1) C19666U in the replicase, resulting in a Ser to Leu change and (2) A27087G in the N gene, a silent mutation. Beau-R stocks were grown on CK cells, harvested 24 hours post-infection (hpi) and titrated on CK cells.

M41-CK: M41-CK is a pathogenic strain of IBV that has been adapted for growth on CK cells. Stocks were grown in 10-day old embryonated SPF Rhode Island Red hens' eggs and titrated on CK cells.

BeauRΔIR: BeauRΔIR is a recombinant of Beau-R in which the intergenic region has been deleted between nucleotides 25191 and 25450 inclusively. BeauRΔIR stocks were grown on CK cells, harvested 24 hpi and titrated on CK cells.

BeauRΔ5ab: BeauRΔ5ab is a recombinant of Beau-R in which all of ORF 5a and the first seven amino acids of ORF 5b has been deleted (nucleotides 25488-25702 inclusive). BeauRΔ5ab stocks were grown on CK cells, harvested 24 hpi and titrated on CK cells.

BeauRΔ3ab: BeauRΔ3ab is a recombinant of Beau-R in which all of ORF 3a, and all except the final 17 nucleotides of ORF 3b has been deleted (nucleotides 23857-24203 inclusive). BeauRΔ3ab stocks were grown on CK cells, harvested 24 hpi and titrated on CK cells.

Additional IBV field and vaccine strains: Field strains D1466, Italy-02 and QX, and vaccine strains CR88 and H120 were obtained from Prof. Richard Jones,

University of Liverpool for an unrelated project. Stocks were grown in 10-day old embryonated SPF Rhode Island Red hens' eggs.

vNotI/IBV$_{FL}$: vNotI/IBV$_{FL}$ is a recombinant vaccinia virus containing a full length cDNA clone of IBV Beau-R (Fig. 2.1). Stocks of vNotI/IBV$_{FL}$ were grown on BHK-21 cells and harvested 4 days post-infection.

rFPV-T7: rFPV-T7 is a recombinant fowlpox virus expressing the T7 bacteriophage RNA polymerase under the control of the vaccinia virus $P_{7.5}$ promoter. Stocks were grown on CEF cells, harvested 4 days post-infection and titrated on CEF cells.

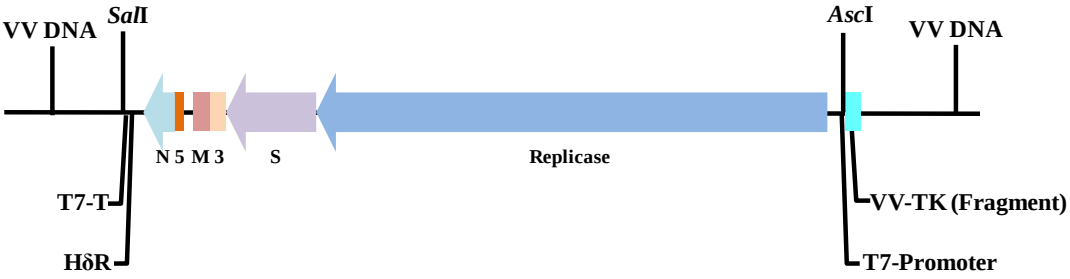


Fig. 2.1. Schematic of vNotI/IBV$_{FL}$. The IBV Beau-R genome is shown in a 3' to 5' orientation within the vaccinia virus genome. The IBV genes within the cDNA are shown. T7-Promoter, HδR and T7-T refer to the positions of the T7 promoter, hepatitis delta antigenome ribozyme and T7 terminator sequences respectively.

2.4 IBV plaque assay

Virus was 10-fold serially diluted in 1X BES medium and used to infect duplicate wells of 6-well plates of confluent CK cells. After 1 h incubation inoculum was removed and cells washed once with PBS. Cells were overlayed with a mixture of 50% 2X BES medium, 50% 2%-agar and incubated for a further 72 h. Cells were

fixed by the addition of 1mL/well of 10% formaldehyde in PBS for 1 h. Agar plugs were removed and cells stained with 0.1% (w/v) crystal violet solution for 20 min to visualize plaques.

2.5 Vaccinia virus plaque assay

Vaccinia virus plaque assays were carried out as part of the IBV reverse genetics system to generate rVVs (see section 2.10). Vaccinia viruses were subjected to three cycles of freeze-thawing, sonicated and 10-fold serially diluted in serum-free EMEM from 10^{-1} – 10^{-3} and used to infect duplicate wells of six-well plates of confluent Vero cells at 500µL/well. After 1 h incubation inoculum was removed and cells washed once with PBS. Cells were overlayed with a mixture of 50% 2X EMEM, 50% 2%-agar, with or without selection components as described in section 2.10, and incubated for a further 72 h. Cells were stained overnight by the addition of a further 2mL overlay medium containing 0.01% (w/v) neutral red. Well-isolated plaques were picked for each recombinant and placed in 400µL of serum-free EMEM.

2.6 Growth curves

Individual wells of 6-well plates of confluent CK cells were infected with $1x10^5$ pfu of virus and incubated for 1 h after which cells were washed twice with PBS and 2mL of fresh 1X BES medium added. Extracellular virus was harvested at 1, 8, 12, 24, 48 and 72 hpi and subsequently analysed in triplicate by plaque assay on CK cells for progeny virus.

2.7 Bacteriological methods

2.7.1 Transformation of TOP10 cells

All transformations were carried out using One Shot® TOP10 chemically competent *E. coli* (Life Technologies) according to manufacturer's instructions. Transformation reactions were plated out onto LB plus ampicillin (100μg/mL) agar plates and placed at 37°C overnight. SOC medium and LB plates were produced by the IAH Microbiological Services department.

2.7.2 Bacterial cultures

Transformed *E. coli* cells were cultured for mini- or maxi-prep of plasmid DNA in LB plus ampicillin (100μg/mL) in 5mL (mini-prep) or 100mL (maxi-prep) volumes overnight at 37°C, 1600 rpm.

2.8 RNA methods

2.8.1 RNA extraction

- Viral RNA was extracted from extracellular virus samples using the RNeasy Mini Kit (Qiagen) following the protocol for RNA Cleanup as per manufacturer's instructions.
- Total RNA was extracted from CK cells using the RNeasy Mini Kit (Qiagen) following the protocol for Animal Cells as per manufacturer's instructions with homogenization using the TissueLyser II (Qiagen) for 30 s at 25hz.
- Total RNA was extracted from excised chicken tracheas using the RNeasy Mini Kit (Qiagen) following the protocol for Animal Tissues as per manufacturer's

instructions with homogenization using the TissueLyser II (Qiagen) for 2 x 1 min at 30hz.

2.8.2 Reverse transcription

Viral RNA was reverse transcribed using SuperScript III (Life Technologies) with random primer (5'-GTTTCCCAGTCACGATCNNNNNNNNNNNNNNN-3') as follows:

RNA template	5μL	
dTNP mix (10mM each dNTP)	1μL	65°C – 5 min
Random primer (40μM)	1μL	
Water	6μL	

Incubate on ice for ≥ 1 min and add 7μL of SuperScript III mix:

5x first strand buffer	4.5μL	25°C – 5 min
0.1M DTT	1μL	50°C – 60 min
RNase OUT (Life Technologies)	1μL	70°C – 10 min
SuperScript III	0.5μL	

2.8.3 Northern blot analysis

Messenger RNA was purified from total RNA samples using the Poly(A)Purist™ MAG Kit (Ambion), as per manufacturer's instructions. Northern blot analysis was carried out with the NorthernMax®-Gly Kit (Ambion) with the following conditions. Briefly, viral mRNA transcripts were denatured in glyoxal load dye at 50°C for 30 min followed by separation on a 0.8% LE-agarose gel. RNA was transferred to BrightStar®-Plus positively charged nylon membrane (Ambion) using capillary action for 2 h, cross-linked by treatment with UV light using the auto-crosslink function on a Stratalinker UV Crosslinker (Stratagene) and prehybridised for 30 min with ULTRAhyb® buffer at 42°C. Blots were probed with a DNA probe specific to

the 3'-end of IBV (forward primer 5'-CAACAGCGCCCAAAGAAG-3' and reverse primer 5'-GCTCTAACTCTATACTAGCCT-3') that was labelled using the BrightStar® Psoralen-Biotin Nonisotopic Labeling Kit (Ambion). Blots were hybridised overnight at 42°C followed by washing and development with the BrightStar® BioDetect™ Kit (Ambion).

2.9 DNA methods

2.9.1 Plasmid mini-prep

Plasmid DNA was prepared from 5mL cultures of transformed *E. coli* using the QIAprep Spin Miniprep Kit (Qiagen) as per manufacturer's instructions. Plasmid DNA was eluted in 30μL water and quantified on a Nanodrop ND-1000 (Thermo Scientific).

2.9.2 Plasmid maxi-prep

Plasmid DNA was prepared from 100mL cultures of transformed *E. coli* using the QIAfilter Plasmid Maxi Kit (Qiagen) as per manufacturer's instructions. Plasmid DNA was eluted in 100μL water, quantified on a Nanodrop ND-1000 (Thermo Scientific) and diluted to a final concentration of 1μg/μL in water.

2.9.3 PCR

All reactions were carried out using an Applied Biosystems 2720 Thermal Cycler.

Overlapping PCR

All PCR reactions were carried out using Pfu Ultra II fusion hotstart DNA polymerase (Agilent Technologies) with the following reaction mix and cycle parameters:

Primer #1 (10μM)	1μL
Primer #2 (10μM)	1μL
dNTP mix (10mM each dNTP)	1μL
10x Pfu Ultra II reaction buffer	5μL
Pfu Ultra II polymerase	1μL
DNA template (100ng/μL)	1μL
Water	40μL

95°C – 2 min
95°C – 30 sec
55°C – 30 sec
72°C – 30 sec
} x 25
72°C – 5 min

rVV and rIBV rescue confirmation PCR

All PCRs were carried out with Taq polymerase (Life Technologies) using the following reaction mix and cycle parameters:

Primer #1 (10μM)	1μL
Primer #2 (10μM)	1μL
dNTP mix (10mM each dNTP)	1μL
$MgCl_2$ (50mM)	2μL
10x Reaction buffer	5μL
Taq polymerase	0.5μL
cDNA template	2μL
Water	37.5μL

94°C – 2 min
94°C – 30 sec
50°C – 30 sec
72°C – 30 sec
} x 25
72°C – 2 min

Leader-body junction PCR

All PCRs were carried out with Taq polymerase (Life Technologies) using the following reaction mix and cycle parameters (see Fig. 6.2 for additional details):

Primer #1 (10μM)	1μL
Primer #2 (10μM)	1μL
dNTP mix (10mM each dNTP)	1μL
$MgCl_2$ (50mM)	4μL
10x Reaction buffer	5μL
Taq polymerase	1μL
cDNA template	2μL
Water	35μL

94°C – 2 min
94°C – 30 sec
50°C – 30 sec
72°C – 30 sec
} x 25
72°C – 2 min

Site-directed mutagenesis PCR

The PCR mutagenesis was carried out using PFU Turbo high fidelity polymerase (Agilent Technologies) with the following reaction mix and cycle parameters:

Primer #1 (10mM)	0.5µL	
Primer #2 (10mM)	0.5µL	
dNTP mix (10mM each dNTP)	0.4µL	95°C – 30 sec
10X PFU reaction buffer	2µL	95°C – 30 sec
PFU Turbo polymerase	0.5µL	45°C – 45 sec] x 16
Water	15.1µL	72°C – 15 min
pGPT-Vaccinia/IBV DNA (300ng/µL)	1µL	

2.9.4 PCR cloning

PCR products amplified using Taq polymerase were cloned using the TOPO TA Cloning® Kit for Sequencing (Life Technologies) as per manufacturer's instructions with 4µL of PCR product per reaction. Cloning reactions were used to transform chemically competent TOP10 cells as in section 2.7.

2.9.5 DNA gel electrophoresis

DNA products were visualized on 1% agarose gels (0.5g agarose, 50mL 1X TBE buffer, 1µL ethidium bromide (10mg/mL)). Electrophoresis was carried out at 150V until products had run sufficiently far. Samples were run with loading buffer (70% water, 30% glycerol, 3.5% (w/v) bromophenol blue). Size marker for all gels was 1Kb plus ladder (Life Technologies) prepared in loading buffer.

2.9.6 PCR purification

PCR products were purified using the QIAquick PCR Purification Kit (Qiagen) as per manufacturer's instructions. Purified products were eluted in 30μL water.

2.9.7 Nucleotide removal

Digested DNA products were purified, and unwanted digested fragments removed, using the QIAquick Nucleotide Removal Kit (Qiagen) as per manufacturer's instructions. DNA was eluted in 50μL water.

2.9.8 Gel extraction

DNA products were extracted from agarose gels using the QIAquick Gel Extraction Kit (Qiagen) as per manufacturer's instructions. DNA was eluted in 30- or 50μL water depending on concentration of starting material.

2.9.9 Ligation

All ligations were carried out using the Quick Ligation^{TM} Kit (NEB) as per manufacturer's instructions.

2.9.10 Sequence analysis

All sequencing reactions were performed by DNA Sequencing and Services, an MRC unit at the University of Dundee. DNA templates were supplied ready mixed with appropriate primers at the requested concentrations depending on the source material to be sequenced. A complete list of primers is given in Table A1.5.

2.10 IBV reverse genetics and rescue of recombinant viruses

All rIBVs used in these studies were generated using the IBV reverse genetics system developed in the Coronavirus Group at IAH (Britton, *et al.* 2005, Casais, *et al.* 2001). The following sections describe the different stages, and the methods used, to generate rIBVs.

2.10.1 Generation of recombinant vaccinia virus by homologous recombination

The reverse genetics system for generating rIBVs is based on a process of transient dominant selection (TDS) (Falkner and Moss 1990). The initial stage involves a recombination event between an IBV cDNA clone in a recombinant vaccinia virus (rVV) and a plasmid containing a modified stretch of IBV genome. Each individual rIBV requires an appropriately modified region of the IBV cDNA genome inserted into selection plasmid pGPTNEB193 (Appendix 2), a modified version of pNEB193 that also contains the *E. coli Ecogpt* gene (GPT) that codes for guanine-xanthine phosphoribosyltransferase. Details are given in individual results chapters as to how the modified genome regions and GPT plasmids were constructed for each rIBV.

Each modified pGPT plasmid was used to generate a rVV containing a correctly modified IBV cDNA. Vero cells grown to 70% confluency in 6-well culture plates were infected with vNotI/IBV_{FL} at an MOI equivalent to 0.2 pfu per cell and incubated for 2 h. Culture medium was removed and cells transfected with 5μg of GPT plasmid DNA using 12 μL Lipofectin® (Life Technologies) as per manufacturer's instructions, in a final volume of 3mL OPTI-MEM (Life Technologies). Transfection medium was replaced after 1 h with 3mL of 1X BES

medium and incubated for a further 24 h to allow for the occurrence of a recombination event between the IBV cDNA clone in vNotI/IBV_{FL} and the modified IBV genome region in the GPT selection plasmid. A successful recombination event results in the production of an unstable virus intermediate containing the modified IBV sequence and the GPT gene (Fig. 2.2).

2.10.2 Transient dominant selection

The unstable virus intermediate was maintained by the addition of GPT selection components at 24 h consisting of 25μg/mL mycophenolic acid (MPA), 250μg/mL xanthine and 15μg/mL hypoxanthine. As MPA is an inhibitor of purine biosynthesis replication of vaccinia virus is blocked in its presence. The presence of the GPT gene provides an alternative pathway for purine biosynthesis when xanthine and hypoxanthine are also supplied.

Following a 48 h incubation in the presence of selection components cell-associated virus was harvested by centrifugation at 13K rpm, 4°C for 3 min (Biofuge fresco, Heraeus) and cell pellets resuspended in 500μL of EMEM. Recombinant vaccinia viruses expressing GPT were selected for by three rounds of plaque purification on Vero cells in the presence of selection components (see 2.5). This was followed by three rounds of plaque purification in the absence of selection components, which results in the rapid loss of GPT and one of two possible virus outcomes depending on a second recombination event (Fig. 2.2). Around 50% of viruses revert to the original virus while the other 50% will contain the desired modification to the IBV cDNA. Plaque purified rVVs were screened by sequence analysis to identify those rVVs containing the correct modification to the IBV cDNA, and were GPT-negative.

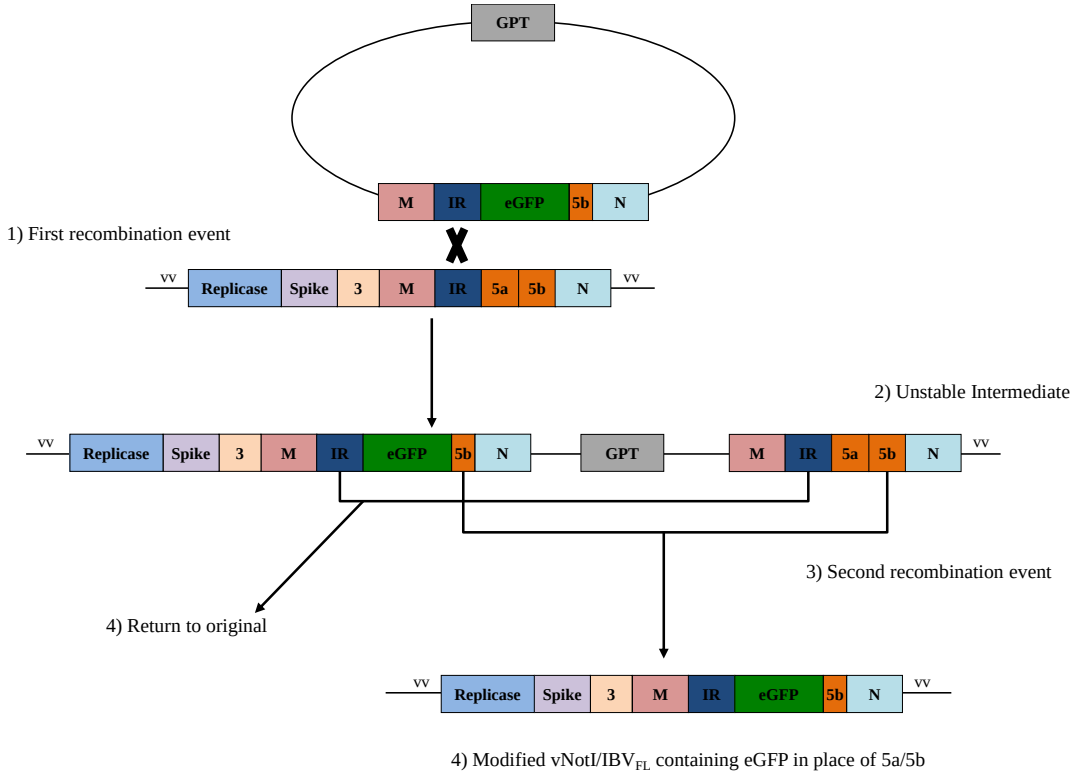


Fig. 2.2. Schematic of transient dominant selection. The schematic shows the four stages of the TDS process. (1) Recombination event between the modified genome fragment of the GPT selection plasmid and the IBV cDNA in vNotI/IBV$_{FL}$. (2) Maintenance of an unstable intermediate during plaque purification under positive selection. (3) Second recombination event following plaque purification in the absence of selection components. (4) The possible outcomes of the second recombination; reversion to wild type or incorporation of the genome modification.

2.10.3 Isolation of rVV DNA

Individual rVVs were grown in 11 T150 flasks of BHK-21 cells until cytopathic effect (CPE) was observed. Infected cells were harvested and resuspended in Tris-EDTA (TE) buffer, pH9 at a volume of 1mL/T150 flask of cells. Cells were lysed by three cycles of freeze-thawing, sonicated and cell nuclei pelleted by centrifugation at 1200 rpm, 4°C for 10 min (ALC, PK121R). Supernatant was transferred to a clean 50mL tube and volume adjusted to 13mL with TE buffer, pH9. Virus particles were partially purified by centrifugation through a 30% sucrose cushion at 14K rpm, 4°C

for 1 h (Sorvall WX Ultra-80, Surespin 630 Sorvall Rotor (Thermo Scientific)) and resuspended in 5mL of TE buffer, pH9.

Partially purified virus was incubated in an equal volume of 2X proteinase K buffer (400mM NaCl, 200mM Tris/HCl pH 7.5, 10mM EDTA, 0.4% SDS), containing 200μg/mL proteinase K for 2.5 h at 50°C. DNA was extracted by 2 x phenol-chloroform and 1 x chloroform extraction and DNA ethanol precipitated. DNA pellets were re-dissolved by the addition of 100μL water and incubation at 4°C overnight.

2.10.4 Pulsed field gel electrophoresis

Vaccinia virus DNA was examined by pulsed field electrophoresis. One μg of rVV DNA was restriction digested with *Sal*I for 3 h at 37°C in a 20μL final volume reaction. Loading buffer was added to DNA samples and heated for 10 min at 65°C. DNA was separated on a 0.8% agarose gel (Pulsed Field Certified Agarose, Bio-Rad Laboratories) prepared in 0.5X TBE buffer using a pulsed field gel electrophoresis tank. DNA samples were run against CHEF DNA Size Standards 8-48 Kb (Bio-Rad Laboratories). Conditions for running the gel were; initial pulse time of 0.1 sec, final pulse time of 1.0 sec, 6.0 V/cm for 12 h. The gel was stained in MQ water containing 0.1μg/mL ethidium bromide for 30 min. Visualization of a DNA product at 33.5Kb, corresponding to the IBV cDNA, was required to confirm that the rVV DNA was intact and suitable for rescue of rIBVs (Fig. 2.3).

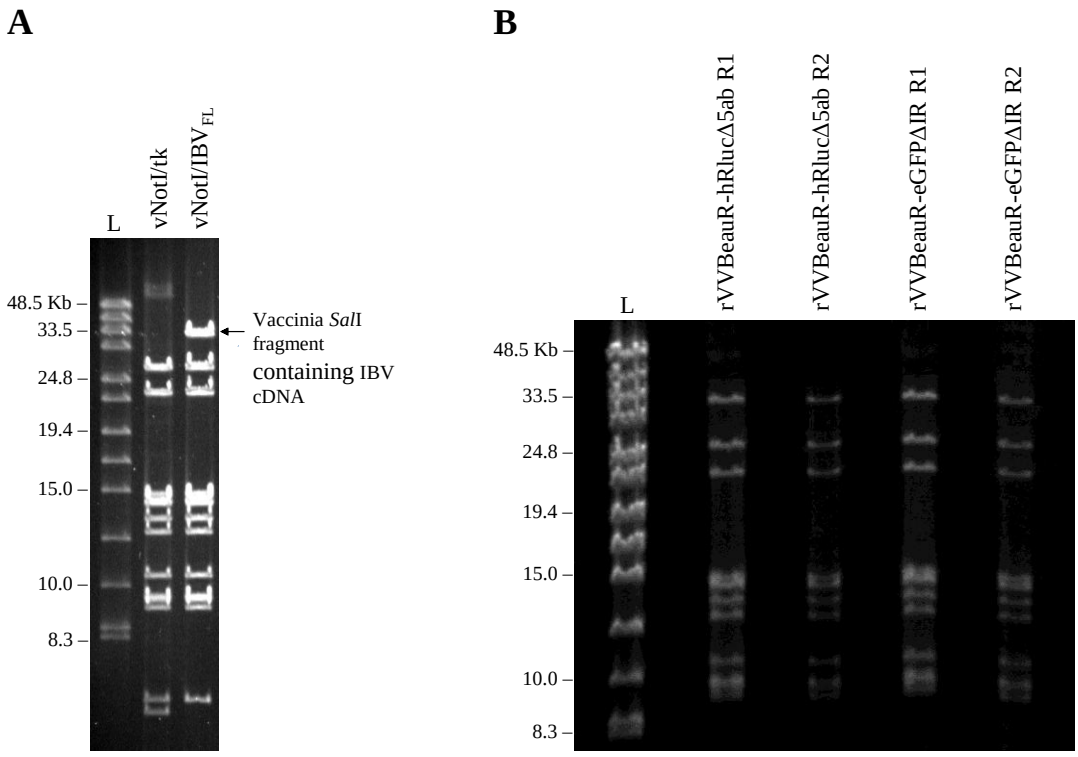


Fig. 2.3. Pulsed field gel analysis of rVV. (A) Pulsed field gel analysis comparing S*al*I digested vNotI/tk (without IBV cDNA) and vNotI/IBV$_{FL}$ (with IBV cDNA) (Casais, *et al.* 2001). (B) Example of pulsed field analysis of generated recombinant viruses rVVBeauR-hRlucΔ5ab and rVVBeauR-eGFPΔIR.

2.10.5 Rescue of recombinant IBVs

Recombinant IBVs were rescued on primary CK cells cultured in 6-well tissue culture plates. Synthesis of infectious IBV RNA, that is equivalent to genomic RNA, is driven by T7 polymerase expressed from recombinant fowlpox rFPV-T7. The infectious IBV RNAs are subsequently used as templates for further rounds of replication of the virus.

Confluent monolayers of CK cells were infected with rFPV-T7 at an MOI of 10 and incubated for 1 h. Virus inoculum was removed and cells washed once with OPTI-MEM (Life Technologies). Cells were transfected with 10μg of rVV DNA plus 5μg

of pCi-Nuc, a plasmid expressing the IBV N protein and essential for rescue, in a total volume of 3mL of OPTI-MEM containing 30μL Lipofectin (Life Technologies). Cells were incubated overnight, transfection medium removed and 5mL of 1X BES added. Following a 48 h incubation the cell supernatant was harvested and passed through 0.22μm filters to remove any fowlpox virus. Cell supernatants containing any potential rescued rIBVs were subsequently passaged on CK cells. One mL of supernatant was used to inoculate 1x T25 flask of confluent CK cells and following a 1 h incubation inoculum was removed and 3mL of 1X BES medium added to flask. Cells were incubated for 3-4 days, or until IBV-induced CPE was observed, and supernatants harvested. Supernatants were passaged a total of 3 times on CK cells before further analysis.

2.11 Protein methods

2.11.1 Western blot analysis

Confluent monolayers of CK cells were infected with $1x10^5$ pfu virus. At 20 hpi cells were washed twice with PBS and lysed with 20mM Tris, 150mM NaCl, 5mM EDTA, 0.3% Triton X100 and 1X protease inhibitor cocktail (Sigma) for 15 min on ice, and clarified by centrifugation for 5 min at 4°C (Biofuge fresco, Heraeus). Cell lysates in 13μL sample volumes were heated at 70°C for 10 min with 5μL NuPage Sample Buffer (Life Technologies) and 2μL 25% β-mercaptoethanol and separated on 10% Bis-Tris precast gels (Life Technologies). Samples were transferred to nitrocellulose membrane (Amersham Hybond ECL nitrocellulose membrane) and probed with antibody. Primary antibodies were diluted in PBS containing 10% Tween 20; (1) anti-IBV polyclonal antibody (non-commercial) at a dilution of

1/2500, (2) anti-beta Actin polyclonal antibody (Abcam) at a dilution of 1/500, (3) anti-GFP polyclonal antibody (Abcam) at a dilution of 1/1000, (4) anti-*Renilla* Luciferase polyclonal antibody (MBL International) at a dilution of 1/1000. Bound antibody was visualized using an enhanced-chemiluminescence detection system (ECL) (Millipore).

2.11.2 Luciferase assay

Cell lysates were prepared from IBV-infected CK cells and 20μL samples analysed for luminescence using the *Renilla* Luciferase Assay System (Promega) as per manufacturer's instructions. Luminescence values were obtained using a GloMax® 20/20 luminometer (Promega) with integration over 10-seconds with a 2-second delay.

2.11.3 BCA protein assay

Protein concentrations were calculated from cell lysates generated for *Renilla* luciferase assays using a BCA protein assay (Pearce). Due to incompatibilities with the *Renilla* Luciferase Assay System lysis buffer, samples were first acetone precipitated to remove traces of buffer. Cell lysates were clarified by centrifugation and to each sample was added 4 volumes of -20°C acetone. Samples were incubated at -20°C for ≥1 h followed by centrifugation for 10 min at 13K rpm (Biofuge Pico, Heraeus). Supernatants were removed and protein pellets left to air dry before being resuspended in a volume of sterile water equal to the original sample volume.

BCA assays were carried out using the protocol for microtitre well plates as per manufacturer's instructions, with protein samples diluted 1/10 in sterile water and readings taken at 540nm.

2.12 Bright field and fluorescence microscopy

All bright field images were captured on an EVOS *xl* digital inverted microscope.
All fluorescence images were captured on a Leica DM IRB inverted microscope attached to a camera.

2.13 Confocal microscopy

Monolayers of Vero cells at 70% confluency and grown on coverslips in 24-well tissue culture plates were infected with $1x10^5$ pfu of rIBV. At 18 hpi cells were washed twice with PBS and fixed with 4% paraformaldehyde for 10 min. Cells were washed once with PBS and permeabilised with PBS containing 0.5% Triton X100 for 15 min with shaking. Cells were blocked with PBS containing 0.5% BSA for 1 h with shaking. Primary antibodies were prepared in PBS containing 0.5% BSA and added for 1 h with shaking: (1) anti-IBV polyclonal antibody (non-commercial) at a dilution of 1/1000, (2) anti-*Renilla* Luciferase polyclonal antibody (MBL International) at a dilution of 1/500. Cells were washed 3 x 5 min with PBS. Secondary antibodies were prepared in PBS containing 0.5% BSA and added at a 1/200 dilution for 1 h with shaking: (1) Alexa-Fluor 568-goat anti-rabbit for detection of anti-IBV binding (Life Technologies), (2) Alexa-Fluor 488-goat anti-rabbit for detection of anti-*Renilla* Luciferase binding (Life Technologies). Cells

were washed 3 x 5 min in PBS. DAPI stain (Life Technologies) was added at 1/20,000 dilution in water for 10 min with shaking. DAPI stain was removed and coverslips mounted on slides. All images were taken using a Leica TCS SP5DM6000 confocal microscope and analysed using Leica Microsystems LAS AF software.

2.14 Quantitative PCR

Confluent monolayers of CK cells were infected with $1x10^5$ pfu of Beau-R, BeauR-L-CTGAACAA, BeauR-L-CTTAACAT or BeauR-IR-CTGAACAA and intracellular RNA harvested at 18 hpi. From each infection the IR and Gene 3 mRNAs were amplified for analysis along with 28S rRNA to act as a housekeeping gene.

Reverse transcription

RNA was reverse transcribed using TaqMan® Reverse Transcription Reagents as supplied (Applied Biosystems). To ensure only mRNAs, and not negative sense RNAs, were reverse transcribed only reverse primers were used for the generation of cDNA (Table 2.5):

10x Reaction buffer	1μL	
$MgCl_2$	2.2μL	
dNTP	2μL	
Reverse Primer (2μM)	1μL	48°C – 30 min
RNAsin	0.2μL	95°C – 5 min
Reverse Transcriptase (RT)	0.25μL	
RNAse free water	1.35μL	
RNA (250ng/μL)	2μL	

For each cDNA generated a no-RT reaction was included to control against non-target amplification during the PCR.

Standard curves

To quantify the mRNA levels for each virus standard curves were generated for each of the three genes to be amplified. The 28S standard curve was generated using Beau-R cDNA that was 10-fold serially diluted and PCRs run on 10^{-1}–10^{-8} samples. For the Gene 3 and IR standard curves the relevant sequences were amplified by PCR and cloned into pcDNA3.12(-) (Life Technologies) using restriction sites 5'-*Xho*I and 3'-*Hin*dIII (Table 2.5). Plasmid DNA was 10-fold serially diluted from a starting concentration of 100pg/μL.

qPCR

The synthesised cDNAs were subsequently used as templates for specific PCR amplification. Trial experiments demonstrated that all test cDNAs be diluted 10^{-4} to give Ct values that fit to the relevant standard curve. All primer and probe sequences are shown in Table 2.5. Samples were run using Applied Biosystems TaqMan Fast Universal PCR Master Mix and the following conditions:

TaqMan Fast Universal PCR Master Mix (2x)	5μL	
Probe (5μM)	0.25μL	
Primers (F and R) (Final 1μM)	0.2μL	95°C – 20 sec
Water	2.05μL	95°C – 3 sec ⎫ x 40
cDNA	2.5μL	60°C – 30 sec ⎭

For each sample a no-template control containing water instead of cDNA was included to control against sample contamination.

Data analysis

Amplification plots were analysed using Applied Biosystems Sequence Detection Software version 1.3.1.21 and the average Ct values for each sample calculated. Standard curves were generated using all points falling within the linear range from the relevant dilution series, and these were then used to calculate the amount of DNA in each sample based on the Ct value. All Ct values were normalized to 28S to standardize each sample. The IR values were additionally normalized against the Gene 3 Ct values for each virus to account for any differences between virus replication rates of the different rIBVs. Finally IR mRNA levels of each rIBV were calculated as a fold-change compared to wild type Beau-R levels.

TABLE 2.5: qRT-PCR Primer and Probe Sequences

Primer/Probe		Sequence
Primer	qPCR Cloning (F)	ATCGCTCGAGCTAGCCTTGCGCTAGATTTTTAAC
	IR Cloning (R)	GCATAAGCTTGCATAGACAAACGTAGCAAACCTTT
	Gene 3 Cloning (R)	GCATAAGCTTCGTGGGACTTTGGATCATCAAAC
	28S (F)	GGGGAAGCCAGAGGAAACT
	28S (R)	GACGACCGATTTGCACGTC
	IBV (F)	CTAGCCTTGCGCTAGATTTTTAACT
	Gene 3 (R)	TGGGACTTTGGATCATCAAACA
	IR (R)	GCATAGACAAACGTAGCAAACCTTT
Probe	28S	6FAM-AGGACCGCTACGGACCTCCACCA-TAM
	Gene 3	6FAM-CAATACAGACCTAAAAAGT-MGB
	IR	6FAM-ACAAAGCGGAAATAA-MGB

Chapter 3: Generation of recombinant IBVs to investigate the potential of IBV as a vaccine vector

3.1 Introduction

The initial aim of this project was to investigate the potential for using IBV as a vaccine vector. This was achieved through the generation of a series of rIBVs, expressing a heterologous protein, and the subsequent analysis of these recombinants. The use of coronaviruses as vaccine vectors has been investigated with a number of different viruses. These studies showed that the expression of heterologous proteins is dependent on a number of factors, as outlined by de Haan *et al.* (de Haan, *et al.* 2005, de Haan, *et al.* 2003). These factors include the genetic background of the virus, the location at which the heterologous gene is inserted and, if replacing genes, how essential these genes are to the replication of the coronavirus.

Using these criteria as a guideline this project aimed to identify the best rIBV construct for use as a vaccine vector. Recombinant viruses were designed with the aim of answering four fundamental questions regarding the expression of heterologous genes and the stability of the resultant rIBVs.

1. Does the position of the heterologous gene within the IBV genome affect gene expression and virus stability?

Previous work by the IAH Coronavirus group showed that the group-specific genes encoding accessory proteins 3a/3b and 5a/5b could be deleted, or expression

prevented, without affecting replication of the virus in cell culture (Casais, *et al.* 2005, Hodgson, *et al.* 2006). Additionally, the region referred to as the IR, located between the M gene and Gene 5, had also been shown to be dispensable for virus replication in cell culture (unpublished observations). Recombinant viruses were designed to express heterologous genes, individually from each of the three locations, and to subsequently identify the best location in terms of heterologous gene expression and virus stability.

2. Does the size, or type, of heterologous protein inserted affect gene expression or stability?

For each location of the genome to be replaced two rIBVs were designed with the insertion of one of two possible reporter genes, eGFP and hRluc. By generating isogenic pairs of rIBVs, where only the heterologous gene is altered, it is possible to identify any differences in protein expression or virus stability that are influenced by the protein itself. The decision to use reporter genes eGFP and hRluc was based on the ability to have a visual output of gene expression in the form of fluorescing (eGFP), or light emitting (hRluc) cells, thus aiding in analysis of protein expression. Additionally, these proteins differ in size by approximately 200bp thus allowing a limited investigation into whether virus stability is affected by the size of the gene inserted. The fLuc gene would have allowed for a greater comparison of gene sizes than hRluc, as it is over twice the size of eGFP. However, previous attempts to use fLuc in IBV D-RNA constructs did not lead to stable D-RNA production (Stirrups, *et al.* 2000). Therefore, hRluc was chosen over fLuc to maximise the chances of producing viable rIBVs.

3. Does codon optimization of the heterologous gene affect virus stability?

Both reporter genes to be used in this study have been optimized for expression in mammalian systems through manipulation of the amino acid, and nucleotide, sequence of the gene (Zhang, *et al.* 1996, Zhuang, *et al.* 2001). It was hypothesised that altering the nucleotide sequence of each gene, so that codon usage better represents that utilized by IBV and avian cells, virus stability would be improved. To limit the number of rIBVs to be made, comparison of optimized and non-optimized genes was planned with rIBVs replacing Gene 5 only.

4. Is recombinant virus stability improved by fusion of the heterologous protein to an existing virus protein?

It was hypothesised that by coupling expression of the heterologous protein to expression of an existing virus protein the resulting recombinant virus would be more stable. Due to the overlapping nature of many of the IBV genes there were limited options to test this hypothesis, with the most suitable being the fusion of the reporter gene to the C-terminus of accessory protein 5a. Only eGFP was tested in this construct to keep the fusion protein as small as possible and again, limit the number of rIBVs that needed to be made.

Based on the questions outlined above the generation of nine rIBVs was planned to investigate the potential for using IBV as a vaccine vector (Fig. 3.1). This chapter details the construction of each of these recombinants while Chapters 4 and 5 discuss the subsequent analysis of the rIBVs in terms of growth and protein expression

(Chapter 4) and virus stability (Chapter 5). Due to the inter-linked nature of the three chapters all of the data will be discussed together at the end of Chapter 5.

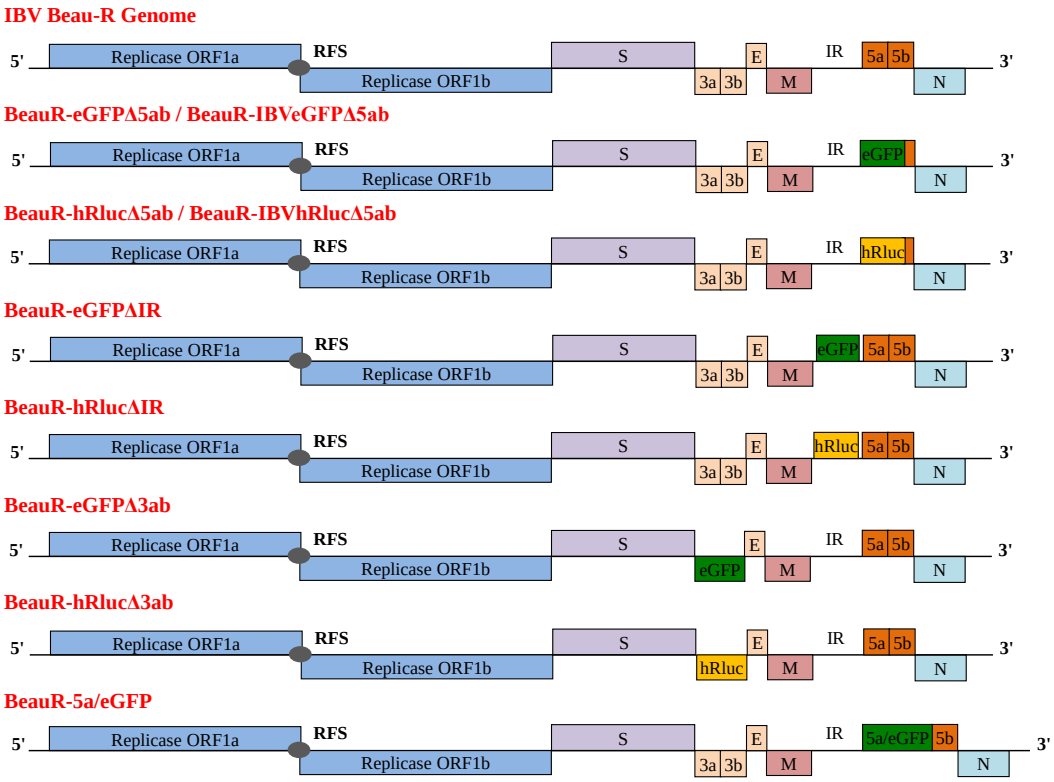


Fig. 3.1. Schematic of recombinant IBVs. The genome of wild type Beau-R is shown with each of the recombinant virus genomes below. Each rIBV genome shows the replacement of a particular genome region with either eGFP or hRluc.

3.2 Generation of heterologous gene constructs

The first step of the IBV reverse genetics system (described in full in section 2.10) was the generation of a modified IBV genome fragment, containing the desired changes to the IBV genome for each different rIBV, and subsequent cloning of this fragment into plasmid pGPTNEB193. These modified plasmids were then used in the transient dominant selection stage of the reverse genetics system. Different

strategies were used to generate appropriate plasmids for each of the different recombinant viruses.

3.2.1 Generation of pGPT-eGFPΔ5ab and pGPT-hRlucΔ5ab

Generation of plasmids for recombinant viruses BeauR-eGFPΔ5ab and BeauR-hRlucΔ5ab utilized a series of overlapping PCRs to obtain genome fragments, containing reporter genes, that could subsequently be cloned into plasmid pGPTNEB193. The sequence of Beau-R Gene 5 to be replaced by the reporter genes was determined by the previously successful deletion of Gene 5 between nucleotides 25488 and 25702 inclusive (unpublished data from IAH Coronavirus group), which resulted in the entire deletion of ORF 5a and deletion of the first 7 amino acids of ORF 5b. It had not been possible in this construct to delete the entire of ORF 5b as this ORF overlaps with the N gene ORF at the 3'-end and additionally contains the N gene TRS required for transcription. As the N protein is essential to replication the majority of the ORF 5b sequence therefore needed to be retained. Genome fragments were designed with approximately 400bp of IBV sequence either side of the reporter gene to allow for homologous recombination events in the vaccinia clone. Genome fragment sequences can be found in Appendix 3.

Genome fragments were split into three sections (A, B and C) and PCR primers designed to amplify each of the sections with an approximate 20bp overhang of the neighbouring section (Table A1.1; Fig. 3.2,). IBV sequences were amplified from vNotI/IBV_{FL} while eGFP and hRluc were amplified from plasmids pEGFP-C1 (Clontech) and pGL4.75hRluc/CMV (Promega) respectively. To facilitate cloning of

the completed fragment, restriction sites were added to the 5'-end of section A (*Asc*I) and to the 3'-end of section C (*Pac*I).

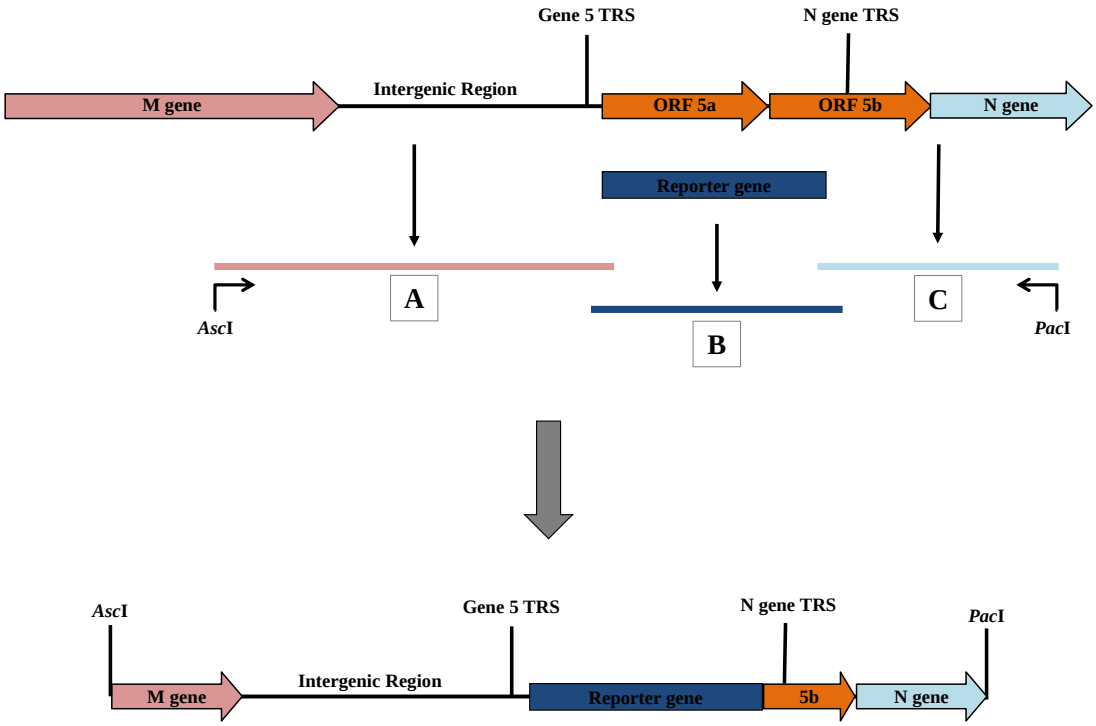


Fig. 3.2. Schematic of overlapping PCR. Sections A, B and C were amplified by PCR from IBV or plasmid templates. The three sections were joined by further rounds of PCR to generate the modified genome fragment.

Modified genome fragments were amplified by PCR (see section 2.9) and cloned into plasmid pGPTNEB193, as outlined in Fig. 3.3, to yield plasmids pGPT-eGFPΔ5ab and pGPT-hRlucΔ5ab (Fig. 3.4).

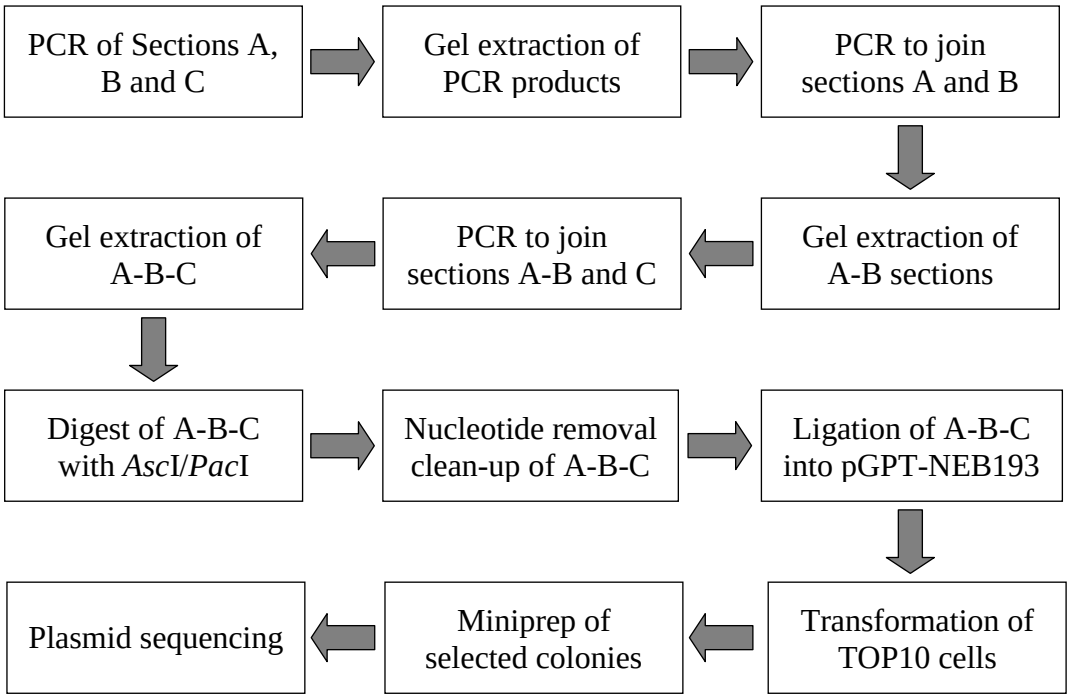


Fig. 3.3. Overlapping PCR process. Flowchart outlining the steps taken to generate plasmids pGPT-eGFPΔ5ab and pGPT-hRlucΔ5ab by overlapping PCR.

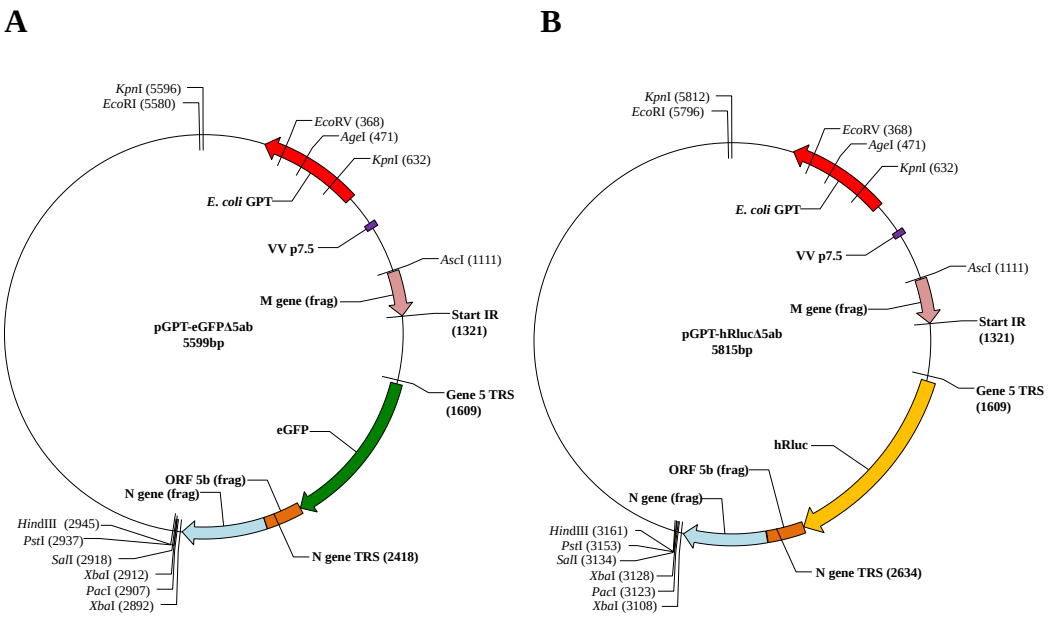


Fig. 3.4. Plasmid maps Δ5ab constructs. Plasmid maps for (A) BeauR-eGFPΔ5ab and (B) BeauR-hRlucΔ5ab showing the modified IBV genome fragment within plasmid pGPTNEB193.

3.2.2 Generation of pGPT-eGFPΔIR and pGPT-hRlucΔIR

To generate plasmids pGPT-eGFPΔIR and pGPT-hRlucΔIR, plasmid pGPTΔIR previously made in the IAH Coronavirus group was utilized. This plasmid contains a 268bp deletion in the Beau-R intergenic region (nucleotides 25192 to 25459 inclusive) and an inserted multiple cloning site adapter (see Appendix 2 for plasmid map). Restriction sites *Nhe*I and *Xma*I were added by PCR to the 5'- and 3'-end of each reporter gene respectively, using overlapping PCR cycle parameters. Primer sequences can be found in Table A1.2. Amplified reporter gene sequences were *Nhe*I and *Xma*I digested, processed through a nucleotide removal column and subsequently ligated into plasmid pGPTΔIR to generate plasmids pGPT-eGFPΔIR and pGPT-hRlucΔIR with a complete genome fragment encompassing ORF 3c to the N gene (Fig. 3.5) (see Appendix 3 for sequences).

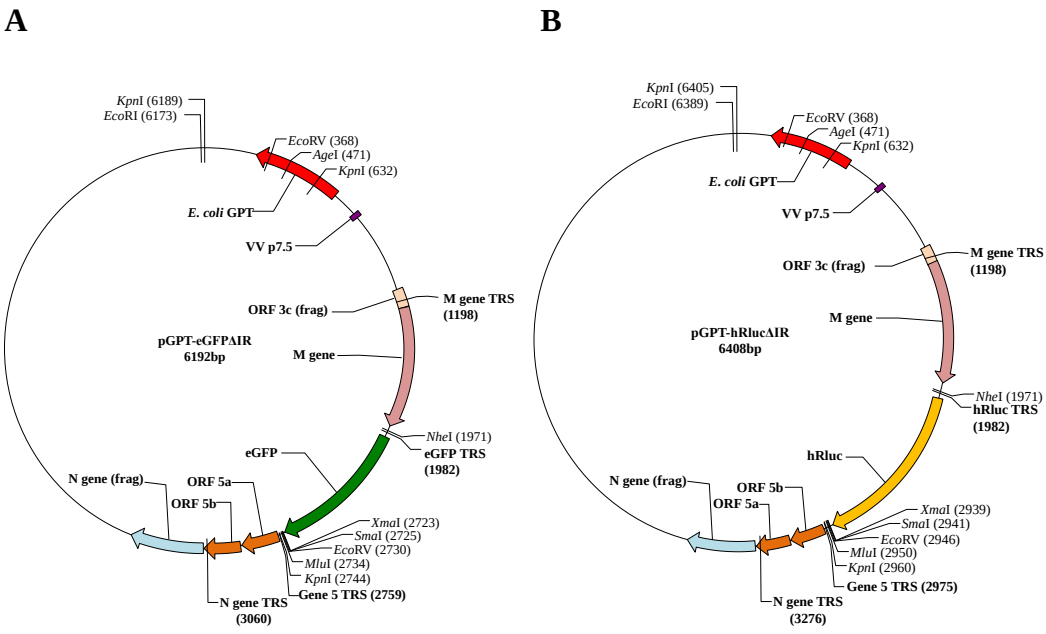


Fig. 3.5. Plasmid maps ΔIR constructs. Plasmid maps for (A) BeauR-eGFPΔIR and (B) BeauR-hRlucΔIR showing the modified IBV genome fragment within plasmid pGPTΔIR.

3.2.3 Generation of pGPT-IBVeGFPΔ5ab and pGPT-IBVhRlucΔ5ab

The first stage in generation of plasmids pGPT-IBVeGFPΔ5ab and pGPT-IBVhRlucΔ5ab was codon optimization of reporter genes eGFP and hRluc to be more IBV-like. An IBV codon usage table was generated using the Codon Usage Database program hosted at www.kazusa.or.jp/codon/ (Table 3.1).

TABLE 3.1: IBV Codon Usage

```
 */Infectious bronchitis virus /[gbvrl]: 785 CDS's (286941 codons)*
-----------------------------------------------------------------------
fields: [triplet] [frequency: *per thousand*] ([number])
-----------------------------------------------------------------------

  UUU 42.3(12136)  UCU 22.2(6359)  UAU 29.9( 8580)  UGU 22.3( 6395)
  UUC 10.0( 2859)  UCC  3.1( 879)  UAC 12.2( 3487)  UGC  5.8( 1653)
  UUA 18.6( 5339)  UCA 18.3(5254)  UAA  1.3(  376)  UGA  1.0(  292)
  UUG 14.3( 4101)  UCG  3.0( 871)  UAG  0.4(  113)  UGG 14.8( 4247)

  CUU 24.9( 7142)  CCU 18.5(5312)  CAU  8.7( 2500)  CGU 12.2( 3501)
  CUC  5.4( 1541)  CCC  5.0(1428)  CAC  5.3( 1533)  CGC  5.6( 1612)
  CUA 11.4( 3266)  CCA 20.0(5733)  CAA 24.2( 6937)  CGA  2.5(  708)
  CUG  7.2( 2057)  CCG  4.0(1138)  CAG 18.8( 5381)  CGG  1.5(  424)

  AUU 27.2( 7813)  ACU 26.5(7618)  AAU 44.0(12624)  AGU 21.6( 6186)
  AUC  5.6( 1596)  ACC  4.8(1379)  AAC 12.8( 3671)  AGC  5.9( 1690)
  AUA 19.7( 5643)  ACA 20.9(6004)  AAA 31.1( 8924)  AGA 16.0( 4605)
  AUG 15.5( 4434)  ACG  4.9(1415)  AAG 30.6( 8769)  AGG  8.2( 2341)

  GUU 33.3( 9556)  GCU 25.7(7371)  GAU 38.3(10980)  GGU 39.4(11297)
  GUC  7.7( 2198)  GCC  8.0(2302)  GAC 15.6( 4482)  GGC  8.1( 2332)
  GUA 18.8( 5383)  GCA 27.4(7874)  GAA 26.2( 7506)  GGA 18.6( 5347)
  GUG 14.6( 4199)  GCG  6.7(1916)  GAG 17.4( 4979)  GGG  4.7( 1353)

-----------------------------------------------------------------------
Coding GC 39.32% 1st letter GC 48.54% 2nd letter GC 40.72% 3rd letter
GC 28.71%
-----------------------------------------------------------------------
```

The codon usage table was input, along with amino acid sequences for eGFP and hRluc, into the EMBOSS backtranseq application (details at http://emboss.open-bio.org/). Nucleotide sequences were derived for both eGFP and hRluc where codon usage mirrored the predominant codon usage of IBV (see Appendix 4). Additionally, to confirm the results of the backtranseq programme the eGFP sequence was

manually altered and found to be identical to the computer generated sequence. The codon optimized sequences were placed in context of the larger genome fragment used for pGPT-eGFPΔ5ab and pGPT-hRlucΔ5ab to give the complete sequence for cloning into pGPTNEB193 (see Appendix 3). The entire sequences, with *Asc*I and *Pac*I restriction sites at the 5'- and 3'-end respectively, were sent to GeneArt® for gene synthesis. Once synthesised GeneArt® were additionally tasked with cloning the sequences into pGPTNEB193 to give plasmids pGPT-IBVeGFPΔ5ab and pGPT-IBVhRlucΔ5ab. Plasmid maps are identical to those for pGPT-eGFPΔ5ab and pGPT-hRlucΔ5ab (see Fig. 3.4).

3.2.4 Generation of pGPT-eGFPΔ3ab and pGPT-hRlucΔ3ab

The initial step in constructing plasmids pGPT-eGFPΔ3ab and pGPT-hRlucΔ3ab was to determine the sequence of Gene 3 that could be deleted whilst maintaining the E protein (ORF 3c) that is also translated from the Gene 3 mRNA. It was also necessary to take into account the overlapping spike stop and ORF 3a initiation codons. It was therefore decided to replace the Beau-R genome from position 23857 to 24203 inclusive, with eGFP or hRluc. In order to ensure transcription of a mRNA for ORF 3c a single copy of the Gene 5 TRS, along with flanking nucleotides, was inserted between the reporter gene stop and the E protein start codons (see Appendix 3). Due to the complexity of the sequences, gene synthesis and cloning into pGPTNEB193 was again contracted to GeneArt® for construction of pGPT-eGFPΔ3ab and pGPT-hRlucΔ3ab (Fig. 3.6).

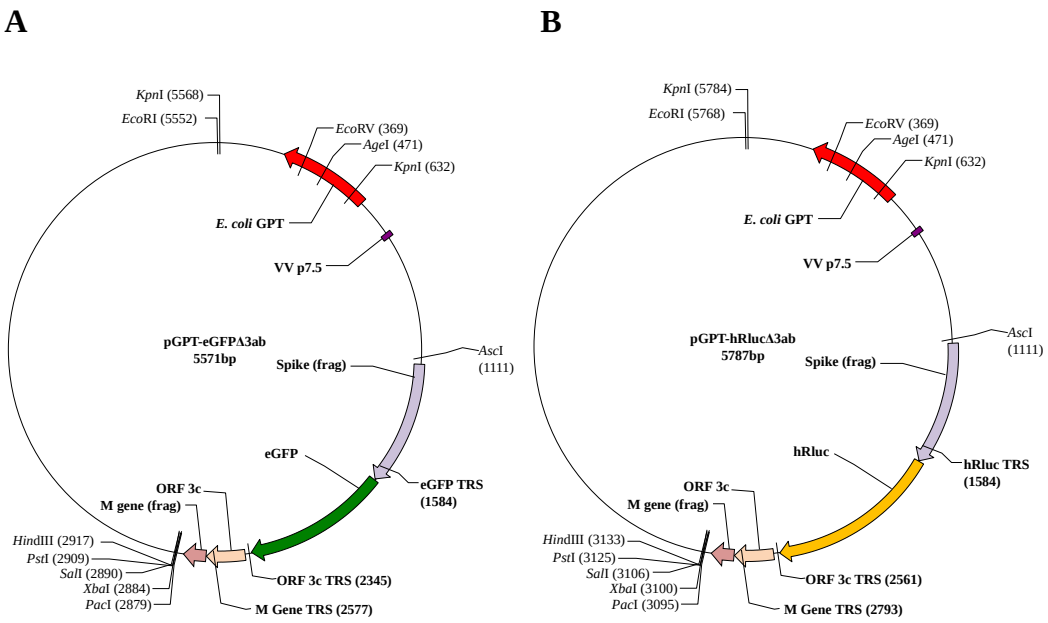


Fig. 3.6. Plasmid maps Δ3ab constructs. Plasmid maps for (A) BeauR-eGFPΔ3ab and (B) BeauR-hRlucΔ3ab showing the modified IBV genome fragment within plasmid pGPTNEB193.

3.2.5 Generation of pGPT-ORF5a/eGFP

Generation of plasmid pGPT-ORF5a/eGFP, for the fusion protein recombinant virus, also required gene synthesis and cloning by GeneArt®. To create the fusion protein, the eGFP sequence, minus the initiation codon, was added to the 3'-end of ORF 5a with the stop codon deleted. The sequence then continued with ORF 5b as for wild type, with the complete genome fragment consisting of the M gene to the N gene (see Appendix 3). The genome fragment was cloned into pGPTNEB193 using *Asc*I and *Pac*I restriction sites (Fig. 3.7).

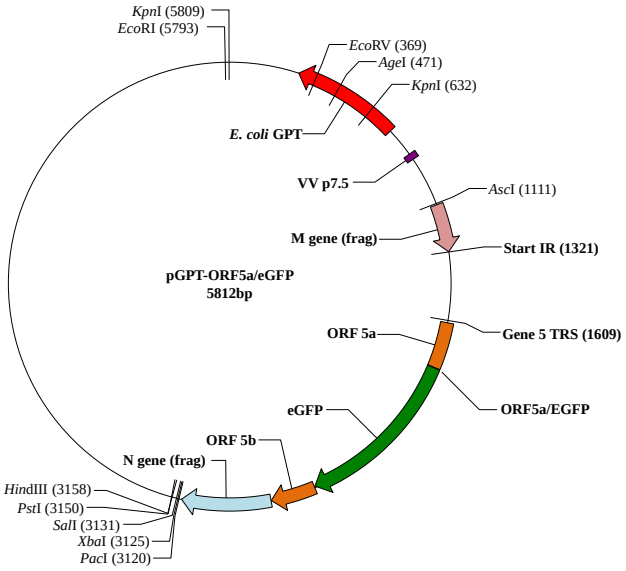


Fig. 3.7. Plasmid map 5a fusion. Plasmid map for pGPT-ORF5a/eGFP showing the modified IBV genome fragment within plasmid pGPTNEB193.

3.3 Generation of recombinant IBVs expressing heterologous genes

The successful rescue of recombinant viruses was initially characterized by progress through four main steps:

1. Generation of a pGPT plasmid containing the correctly modified IBV genome sequence as confirmed by sequence analysis.
2. Generation of two recombinant vaccinia viruses containing the correctly modified full-length IBV cDNA clone. Vaccinia viruses were taken from as distinct plaque lineages as possible and confirmed by sequence analysis of the modified region.
3. Rescue of rIBV from vaccinia viruses as confirmed by observation of CPE in CK cells by passage (P) 3.
4. Confirmation of presence of reporter gene, by RT-PCR, in two replicates (R1 and R2) of rIBV.

The PCR process used to amplify virus regions for sequence analysis of the modified IBV cDNA clones (step 2), and the RT-PCR process used to confirm the successful rescue of each rIBV (step 4), used specific primer pairs (Table A1.4) to amplify a region of the IBV genome containing the inserted reporter gene. PCR products were subsequently analysed by gel electrophoresis. The expected PCR products for each recombinant are listed in Table 3.2.

TABLE 3.2: Confirmation PCR Products

Virus	Primer Set	PCR Product (bp)
BeauR-eGFPΔ5ab/ BeauR-IBVeGFPΔ5ab	5abS + BG147	963
BeauR-hRlucΔ5ab/ BeauR-IBVhRlucΔ5ab	5abS + BG147	1179
BeauR-eGFPΔIR	BG52 + BG146	1087
BeauR-hRlucΔIR	BG52 + BG146	1303
BeauR-eGFPΔ3ab	BG49 + BG144	1700
BeauR-hRlucΔ3ab	BG49 + BG144	1916
BeauR-5a/eGFP	5abS + BG148	1475

The success, or failure, of each rescue is summarized in Table 3.3. As shown three of the planned viruses, BeauR-IBVeGFPΔ5ab, BeauR-eGFPΔ3ab and BeauR-5a/eGFP, did not result in a rescued rIBV. For all other viruses two replicates of each were successfully rescued, as determined by RT-PCR analysis described above, and were taken forward for further analysis (see Chapters 4 and 5).

TABLE 3.3: Rescue of Recombinant IBVs

Virus	pGPT Plasmid	rVV Clones	rIBV CPE	rIBV PCR
BeauR-eGFPΔ5ab	✓	✓	✓	✓
BeauR-hRlucΔ5ab	✓	✓	✓	✓
BeauR-IBVeGFPΔ5ab	×	N/A	N/A	N/A
BeauR-IBVhRlucΔ5ab	✓	✓	✓	✓
BeauR-eGFPΔIR	✓	✓	✓	✓
BeauR-hRlucΔIR	✓	✓	✓	✓
BeauR-eGFPΔ3ab	✓	✓	×	N/A
BeauR-hRlucΔ3ab	✓	✓	✓	✓
BeauR-5a/eGFP	✓	✓	✓	×

The three viruses that were not rescued each failed at different stages of the reverse genetics process as described below.

3.3.1 Investigation of BeauR-IBVeGFPΔ5ab

Attempts to generate a recombinant virus containing a codon optimized version of eGFP in place of Gene 5 did not go beyond initial attempts to clone the modified genome fragment into plasmid pGPTNEB193. The genome fragment was correctly synthesised by GeneArt®, as confirmed by sequence analysis, however they were unable to clone the fragment into plasmid pGPTNEB193. Analysis of the genome fragment revealed no obvious reasons for the failure, i.e. toxic sequences, and several competent cell lines were tried without success (personal communication from GeneArt®). The synthesised genome fragment was provided in the form of a

PCR product. Several further attempts were made by myself to clone this fragment though none were successful and therefore this construct was taken no further.

3.3.2 Investigation of BeauR-eGFPΔ3ab

Two correctly sequenced vaccinia viruses were obtained for BeauR-eGFPΔ3ab however, no rIBVs could be rescued from these viruses. The rescue process in CK cells was carried out a total of three times using fresh preparations of vaccinia virus DNA each time. At each attempt no CPE could be seen in any cells up to, and including, P3 of the virus. To confirm that the lack of CPE was due to a lack of virus and not due to potentially altered growth characteristics, RT-PCRs were carried out at P3 of each attempt to look for virus RNA. This RT-PCR was carried out using primers BG56 and 93/100, which cover the 3' UTR of the virus. As an essential region of the virus a positive RT-PCR result would demonstrate that the virus is present in some form. All RT-PCRs showed negative results confirming that this virus did not rescue in CK cells.

Previously in the IAH Coronavirus group some rIBVs had been shown to rescue only when passaged in embryonated eggs, believed to be a consequence of the increased efficiency of virus replication in eggs. An attempt was therefore made to rescue BeauR-eGFPΔ3ab by passage in embryonated hens' eggs. Passage 0 was carried out in CK cells as normal but any potential virus harvested in the cell culture medium was immediately inoculated into 10-day old Rhode Island Red embryonated hens' eggs for P1. Allantoic fluid harvested from these eggs, at 20 h post inoculation, was subsequently passaged twice more in eggs and then screened by RT-PCR as

above to check for the presence of virus. Again, all RT-PCRs were negative confirming that this virus could not be rescued. The lack of rescued virus meant that this construct could also not be studied any further.

3.3.3 Investigation of BeauR-5a/eGFP

Recombinant virus BeauR-5a/eGFP was identified as a successful rescue as defined by the observation of CPE by P3 in CK cells. However, when RT-PCRs were carried out to confirm the presence of eGFP unexpected results were observed. Using primers 5abS and BG148, which flank the modified region, a PCR product of 1475bp was expected, but instead a product of between 650bp and 850bp was observed (Fig. 3.8).

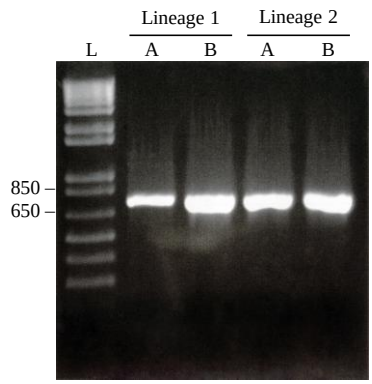


Fig. 3.8. RT-PCR confirmation of BeauR-5a/eGFP rescue. Presence of eGFP was investigated by RT-PCR of P3 rescued virus. Two rescued replicates (A and B) from each vaccinia virus lineage were tested and a PCR product of between 650bp and 850bp was confirmed in all four viruses. Lane L; DNA marker.

The size of the PCR product observed is between 625bp and 825bp smaller than that expected and suggests the deletion of the eGFP gene, the size of which falls into this range. To confirm if this was the case each rescued virus replicate was sequenced from the IR to the end of ORF 5b to capture any deletion (Fig. 3.9).

Fig. 3.9. BeauR-5a/eGFP sequence alignment. Each replicate of BeauR-5a/eGFP rescued was analysed by RT-PCR with primers 5abS and BG148. Shown are partial alignments of each rescue with the wild type sequence. The complete eGFP sequence was found to be deleted in all four replicates. The starts of eGFP and 5b are highlighted green, the ends of 5a and eGFP are highlighted red.

5a End/eGFP Start

```
BeauR-5a/eGFP: ACTTAGAGTGTTAGATAGGTTAATTTTAGATCACGGACTACTACGCGTTTTAACGTGTAGTAGGCGCGTGCTTTTAGTTCAATTAGATTTAGTTTATAGGTTGGCGTATACGCCCACCCAATCGCTGGCAGTGAGCAAGGGCGAGGAGCT
1-A          : ACTTAGAGTGTTAGATAGGTTAATTTTAGATCACGGACTACTACGCGTTTTAACGTGTAGTAGGCGCGTGCTTTTAGTTCAATTAGATTTAGTTTATAGGTTGGCGTATACGCCCACCCAATCGCTGGC----------------------
1-B          : ACTTAGAGTGTTAGATAGGTTAATTTTAGATCACGGACTACTACGCGTTTTAACGTGTAGTAGGCGCGTGCTTTTAGTTCAATTAGATTTAGTTTATAGGTTGGCGTATACGCCCACCCAATCGCTGGC----------------------
2-A          : ACTTAGAGTGTTAGATAGGTTAATTTTAGATCACGGACTACTACGCGTTTTAACGTGTAGTAGGCGCGTGCTTTTAGTTCAATTAGATTTAGTTTATAGGTTGGCGTATACGCCCACCCAATCGCTGGC----------------------
2-B          : ACTTAGAGTGTTAGATAGGTTAATTTTAGATCACGGACTACTACGCGTTTTAACGTGTAGTAGGCGCGTGCTTTTAGTTCAATTAGATTTAGTTTATAGGTTGGCGTATACGCCCACCCAATCGCTGGC----------------------

BeauR-5a/eGFP: GTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTG
1-A          : ------------------------------------------------------------------------------------------------------------------------------------------------
1-B          : ------------------------------------------------------------------------------------------------------------------------------------------------
2-A          : ------------------------------------------------------------------------------------------------------------------------------------------------
2-B          : ------------------------------------------------------------------------------------------------------------------------------------------------

BeauR-5a/eGFP: GCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAA
1-A          : ------------------------------------------------------------------------------------------------------------------------------------------------
1-B          : ------------------------------------------------------------------------------------------------------------------------------------------------
2-A          : ------------------------------------------------------------------------------------------------------------------------------------------------
2-B          : ------------------------------------------------------------------------------------------------------------------------------------------------

BeauR-5a/eGFP: GACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCA
1-A          : ------------------------------------------------------------------------------------------------------------------------------------------------
1-B          : ------------------------------------------------------------------------------------------------------------------------------------------------
2-A          : ------------------------------------------------------------------------------------------------------------------------------------------------
2-B          : ------------------------------------------------------------------------------------------------------------------------------------------------

BeauR-5a/eGFP: GAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCT
1-A          : ------------------------------------------------------------------------------------------------------------------------------------------------
1-B          : ------------------------------------------------------------------------------------------------------------------------------------------------
2-A          : ------------------------------------------------------------------------------------------------------------------------------------------------
2-B          : ------------------------------------------------------------------------------------------------------------------------------------------------
```

eGFP Stop/5b Start

```
BeauR-5a/eGFP: GAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAATAAATGAATAATAGTAAAGATAATCCTTTTCGCGGAGCAATAGCAAGAAAAGC
1-A          : ----------------------------------------------------------------------------------------------------------------------------ATGAATAATAGTAAAGATAATCCTTTTCGCGGAGCAATAGCAAGAAAAGC
1-B          : ----------------------------------------------------------------------------------------------------------------------------ATGAATAATAGTAAAGATAATCCTTTTCGCGGAGCAATAGCAAGAAAAGC
2-A          : ----------------------------------------------------------------------------------------------------------------------------ATGAATAATAGTAAAGATAATCCTTTTCGCGGAGCAATAGCAAGAAAAGC
2-B          : ----------------------------------------------------------------------------------------------------------------------------ATGAATAATAGTAAAGATAATCCTTTTCGCGGAGCAATAGCAAGAAAAGC
```

The sequence alignment shows the precise deletion of the eGFP sequence from each virus replicate. As the recombinant vaccinia viruses from which the rIBVs were rescued had been shown by sequence analysis to contain the correctly modified sequence it was clear that the deletion of eGFP had occurred during passage on CK cells as part of the rescue process. The presence of the same deletion in all four replicates suggests the sequence construction of this virus was not conducive to production of the desired fusion protein. No further analysis of this virus was carried out and given timescales it was not possible to alter the virus sequence and attempt a different construct for generating the fusion protein recombinant.

3.4 Summary

- Virus was successfully rescued for six of the nine planned recombinants based on the presence of the reporter gene by RT-PCR: BeauR-eGFPΔ5ab, BeauR-hRlucΔ5ab, BeauR-IBVhRlucΔ5ab, BeauR-eGFPΔIR, BeauR-hRlucΔIR and BeauR-hRlucΔ3ab.
- BeauR-IBVeGFPΔ5ab could not be rescued, as the modified genome fragment could not successfully be cloned into plasmid pGPTNEB193 for reverse genetics.
- BeauR-eGFPΔ3ab failed to rescue from correctly modified vaccinia virus clones for unknown reasons.
- BeauR-5a/eGFP did not successfully rescue. Although virus was rescued the eGFP sequence was deleted in all replicates obtained.

Results from this chapter will be discussed in detail in an overall discussion of Chapters 3-5 at the end of Chapter 5.

Chapter 4: Analysis of rIBV growth characteristics and heterologous gene expression

4.1 Introduction

This chapter details the analyses carried out for each of the successfully rescued rIBVs. Each virus was assessed on a number of criteria to establish which constructs are the most suitable to act as vaccine vectors. Viruses were assessed for growth on CK cells, transcription of the reporter gene and expression of the reporter protein. In some of the analyses rIBVs were compared to a parental virus as well as the wild type Beau-R. These parental viruses, BeauRΔ5ab, BeauRΔIR and BeauRΔ3ab, contain deletions in the Beau-R genome corresponding to the regions replaced by the reporter genes in the rIBVs being analysed, and were used to confirm that any differences between rIBVs and Beau-R were not simply due to the loss of virus genes. All analyses were carried out using P4 stocks of rIBVs.

4.2 Characterization of rIBVs

All rIBVs that rescued were initially assessed for the formation of plaques on CK cells to highlight any differences between wild type, or parental viruses, and the rIBVs (Table 4.1; Fig 4.1).

Fig. 4.1. Plaque morphology of rIBVs. Viruses were titrated on CK cells by plaque assay in 6-well culture plates. Images were taken of individual wells for each virus in which plaques could be counted and measured. Recombinants were compared to wild type Beau-R and parental viruses BeauRΔ5ab, BeauRΔIR and BeauRΔ3ab.

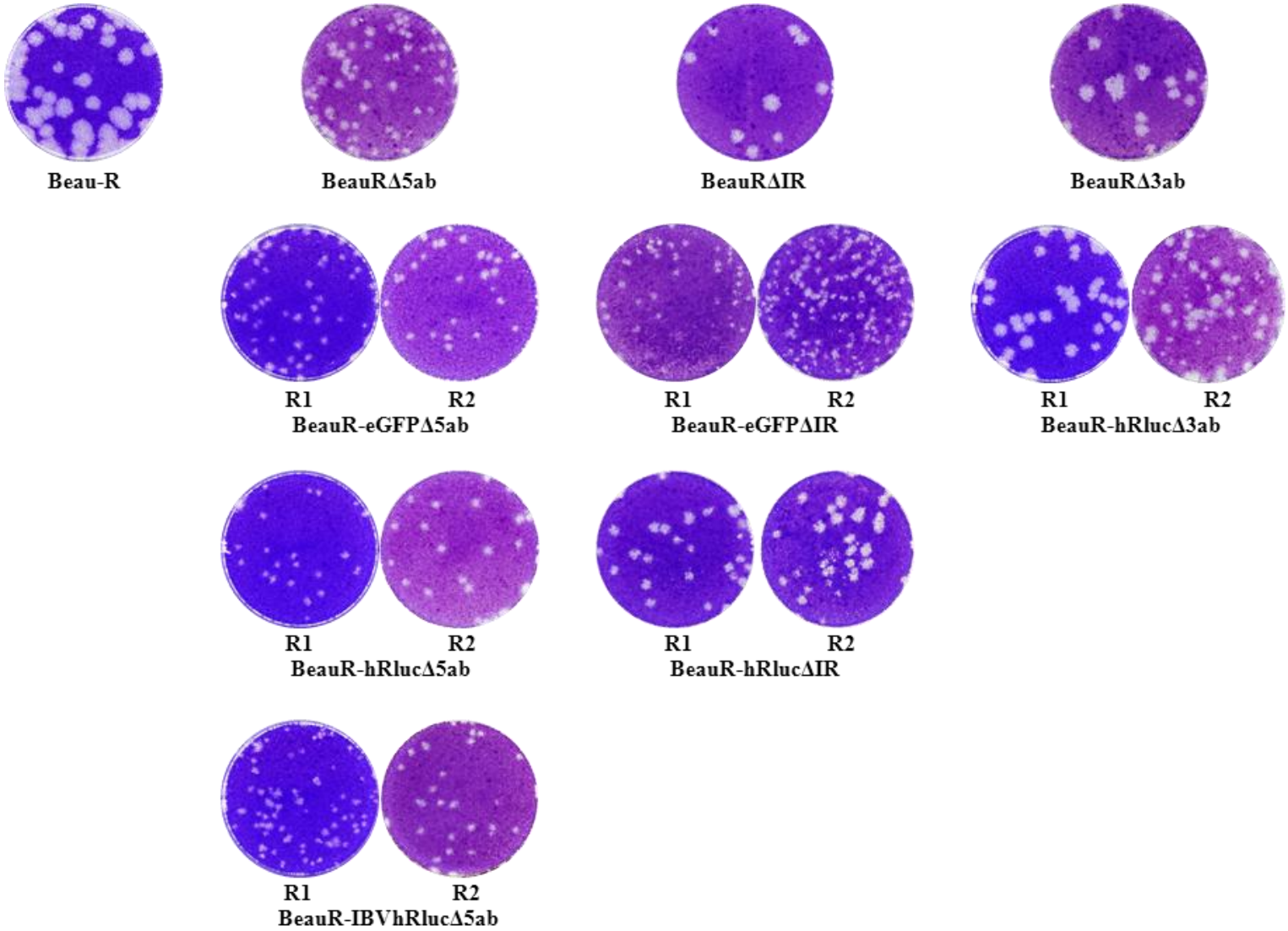

Beau-R
BeauRΔ5ab
BeauRΔIR
BeauRΔ3ab
R1
R2
BeauR-eGFPΔ5ab
R1
R2
BeauR-eGFPΔIR
R1
R2
BeauR-hRlucΔ3ab
R1
R2
BeauR-hRlucΔ5ab
R1
R2
BeauR-hRlucΔIR
R1
R2
BeauR-IBVhRlucΔ5ab

TABLE 4.1: Recombinant IBV Plaque Size

Virus	Plaque Size (mm)[a]	
	R1	R2
Beau-R	3.37 ± 0.43	-
BeauRΔ5ab	1.50 ± 0.33	-
BeauRΔIR	2.85 ± 0.47	-
BeauRΔ3ab	2.65 ± 0.47	-
BeauR-eGFPΔ5ab	1.22 ± 0.34	1.32 ± 0.40
BeauR-hRlucΔ5ab	1.28 ± 0.34	1.42 ± 0.32
BeauR-IBVhRlucΔ5ab	1.35 ± 0.35	1.15 ± 0.30
BeauR-eGFPΔIR	1.15 ± 0.30	1.25 ± 0.34
BeauR-hRlucΔIR	2.23 ± 0.39	2.35 ± 0.42
BeauR-hRlucΔ3ab	2.12 ± 0.31	1.72 ± 0.41

[a] Average of 30 plaques from triplicate titrations ± standard deviation(SD)

All viruses were found to form plaques on CK cells as expected with similar results for the two replicates. However, it is clear that plaques for all rIBVs, including parental viruses BeauRΔ5ab, BeauRΔIR and BeauRΔ3ab were smaller than wild type Beau-R. Recombinant viruses in which Gene 5 was deleted, or replaced, displayed the smallest plaque sizes overall and were approximately 2mm smaller in diameter than wild type. This suggested that the loss of Gene 5 results in some form of growth defect for the virus, although the reason for this cannot be deduced from plaque data alone.

Parental virus BeauRΔIR showed an approximate 0.5mm reduction in diameter size compared to wild type indicating a minimal effect on virus growth when the IR is deleted. Interestingly, while BeauR-hRlucΔIR only showed a further size reduction of approximately 0.5mm for both replicates, BeauR-eGFPΔIR showed a further decrease of approximately 1.5mm suggesting that eGFP is less well tolerated than hRluc at this position with regard to growth on CK cells. The unsuccessful rescues of

BeauR-IBVeGFPΔ5ab and BeauR-eGFPΔ3ab meant that it was not possible to fully compare eGFP constructs to see if eGFP is generally less well tolerated, or if this effect is only observed following replacement of the IR.

The deletion of ORFs 3a and 3b (BeauRΔ3ab) resulted in a similar reduction in plaque size to the deletion of the IR, again suggesting a minimal effect on virus growth on CK cells. The presence of hRluc resulted in a further small decrease in plaque diameter similar to the effect observed with the ΔIR recombinants. However, while similar sized plaques are seen between replicates for the Δ5ab and ΔIR recombinants this is not the case for BeauR-hRlucΔ3ab where a mixture of plaque sizes is visible for R2. Varied plaque sizes are often the result of a mixed virus population, and suggested that BeauR-hRlucΔ3ab had not rescued as a single population of virus (discussed further in 4.3).

To further investigate the growth characteristics of the rIBVs replication kinetics were compared to wild type and parental viruses using multistep growth curves. Initially growth curves were carried out to assay R1 only for each rIBV. Comparing the replacement of Gene 5 with eGFP, hRluc or IBVhRluc all three rIBVs replicated as observed for the parental virus BeauRΔ5ab (Fig. 4.2). However, when compared to Beau-R two differences in kinetics were observed. First, all Δ5ab rIBVs replicated to approximately 1-$\log_{10}$ lower peak titres at 24 h than wild type suggesting an impairment of virus growth with the deletion of Gene 5. Second, there is a delay in replication at 8 h compared to wild type. A 2-$\log_{10}$ increase in titre is noted for Beau-R from 1 to 8 h post infection compared to a <0.5-$\log_{10}$ increase for the rIBVs.

Together with the smaller plaque sizes these results suggest that the loss of Gene 5 has a small detrimental effect on virus replication.

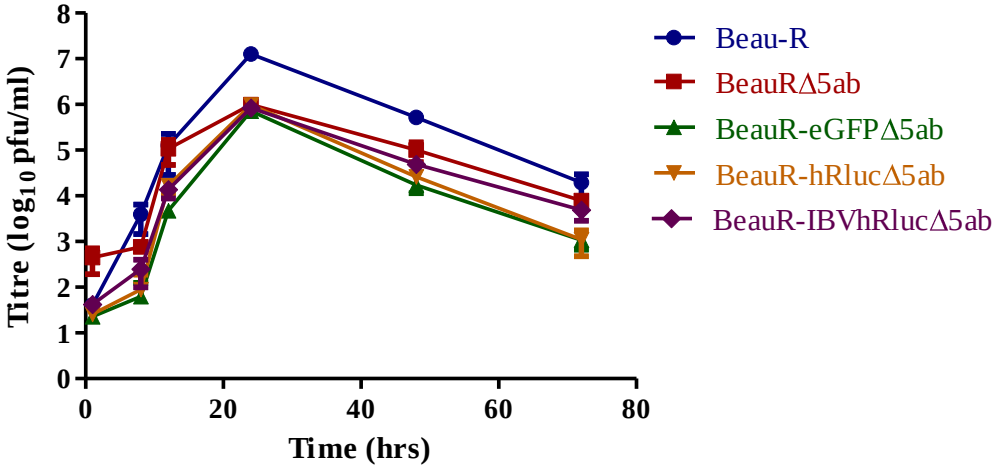


Fig. 4.2. Replication kinetics of 5ab rIBVs. Confluent monolayers of CK cells were infected with $1x10^5$ pfu virus and extracellular virus harvested at various time points for plaque assay. Error bars represent SD of results from three independent experiments each assayed in triplicate.

The replication kinetics of rIBVs replacing the IR were also investigated (Fig. 4.3). No difference in replication was observed between Beau-R and parental virus BeauRΔIR with only an approximate 0.5-log_{10} drop in peak titre of BeauR-hRlucΔIR. However, a reduction in peak titre of almost 2-log_{10} was observed for BeauR-eGFPΔIR.

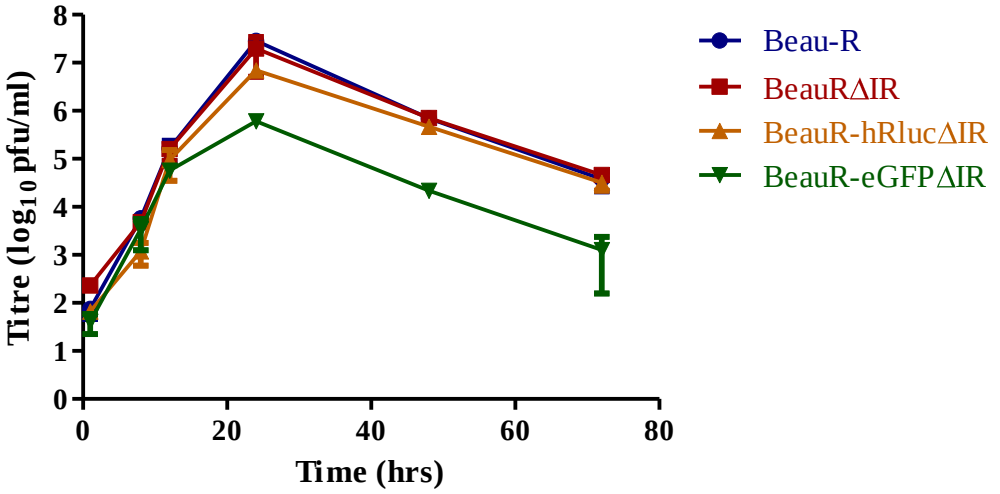


Fig. 4.3. Replication kinetics of IR rIBVs. Confluent monolayers of CK cells were infected with $1x10^5$ pfu virus and extracellular virus harvested at various time points for plaque assay. Error bars represent SD of results from three independent experiments each assayed in triplicate.

Although this result was consistent with the plaque sizes observed for these viruses it was decided to sequence the ΔIR rIBVs at this stage to ensure that no additional, unintentional, changes were present and to confirm that all effects observed were due to the presence of eGFP. Sequence analysis identified that R1 of BeauR-eGFPΔIR contained an additional change when compared to the sequences of each of the other ΔIR rIBVs. This consisted of a double T→A mutation, at nucleotides 25770 and 25771, directly upstream of the N gene TRS resulting in amino acid change F30Stop within ORF 5b (Fig. 4.4A). Comparing this to R2 of BeauR-eGFPΔIR the presence of the same two mutations was observed, but as a minor population (Fig. 4.4B). Additional sequence analysis of a later passage of R2 (P7) showed the continuing presence of this double mutation as a minor virus population (Fig. 4.4C). Later passage data was not available to investigate whether these mutations would eventually become the dominant population in both replicates.

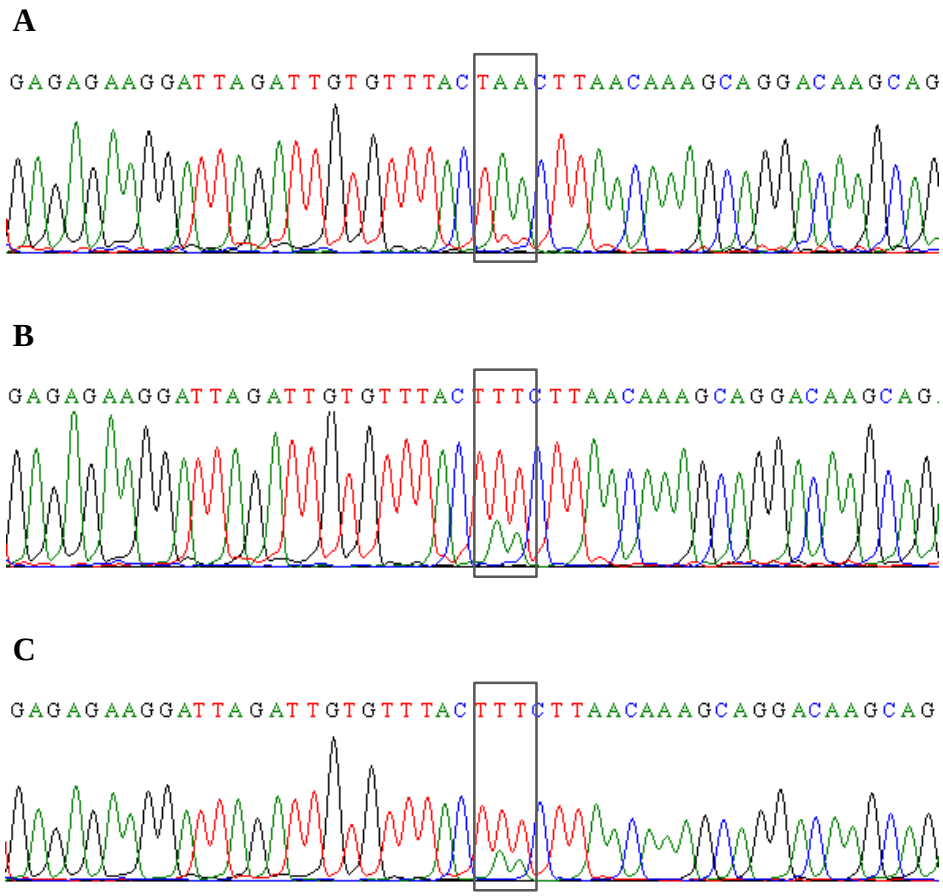


Fig. 4.4. Sequence traces of BeauR-eGFPΔIR 5b mutations. Viruses were sequenced with primer BG148 and sequence traces compared. (A) Passage 4 BeauR-eGFPΔIR R1. (B) Passage 4 BeauR-eGFPΔIR R2. (C) Passage 7 BeauR-eGFPΔIR R2. Grey box highlights position of mutations.

The result of this double mutation is the introduction of a premature stop codon within ORF 5b and therefore loss of expression of this protein. To determine whether the additional loss of 5b in R1 of BeauR-eGFPΔIR was responsible for the lower titres of this virus compared to BeauRΔIR and BeauR-hRlucΔIR a second growth curve was carried out to compare replication kinetics of the two BeauR-eGFPΔIR replicates, R1 and R2 (Fig. 4.5). As the double mutation is present only as a minor population in R2 any additional effect of the loss of 5b should be noticeable between the two replicates.

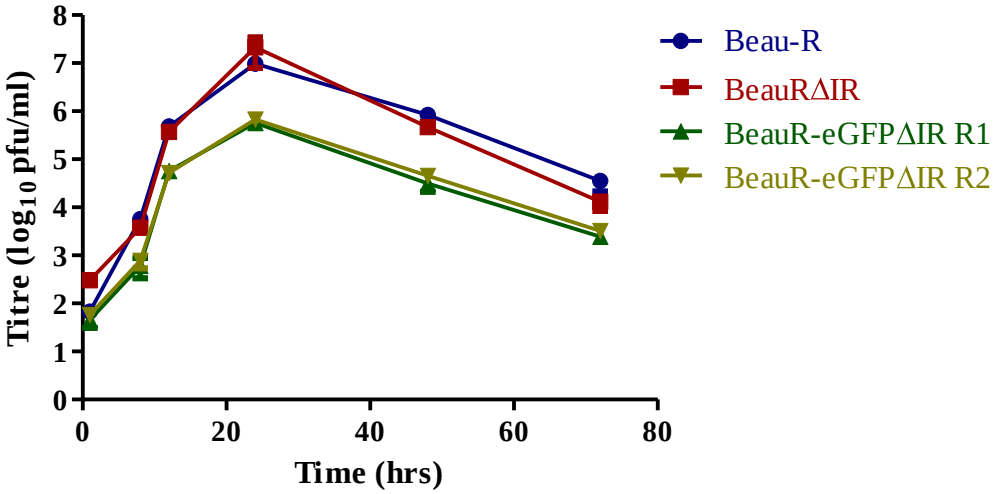


Fig. 4.5. Replication kinetics of BeauR-eGFPΔIR replicates. Confluent monolayers of CK cells were infected with $1x10^5$ pfu virus and extracellular virus harvested at various time points for plaque assay. Error bars represent SD of results from three independent experiments each assayed in triplicate.

Both R1 and R2 of BeauR-eGFPΔIR displayed the same replication kinetics indicating that the additional loss of 5b was not responsible for the decrease in replication fitness of R1 compared to wild type and parental virus. Growth differences are therefore due to the insertion of eGFP and suggest that eGFP is not as well tolerated as hRluc at this location.

The final growth curve showed that the replication kinetics of rIBV BeauR-hRlucΔ3ab were similar to those observed for parental virus BeauRΔ3ab, with both viruses reaching peak titres at 24 h that were approximately 1-log_{10} lower than wild type virus (Fig. 4.6). As with the deletion of Gene 5, the lower peak titres suggest a minor impact on the replication fitness of viruses lacking ORFs 3a and 3b, but no additional effect is observed by the insertion of hRluc.

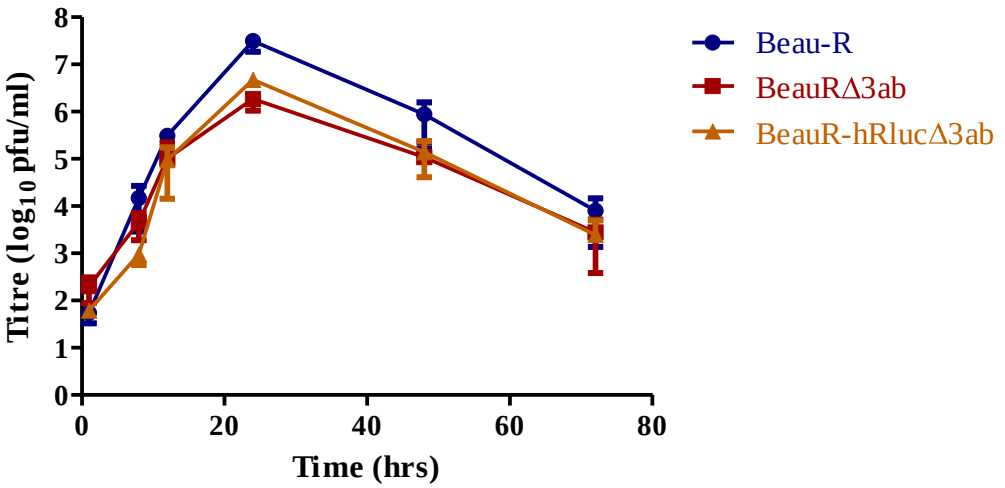


Fig. 4.6. Replication kinetics of 3ab rIBVs. Confluent monolayers of CK cells were infected with $1x10^5$ pfu virus and extracellular virus harvested at various time points for plaque assay. Error bars represent SD of results from three independent experiments each assayed in triplicate.

All of the recombinant viruses were able to grow on CK cells although peak titres were between approximately 1- and 2-log_{10} lower than wild type Beau-R. This was considered acceptable based on the modifications made to the virus and analysis was continued to investigate heterologous gene expression from each of the recombinants.

4.3 Recombinant IBV subgenomic mRNA expression

The mechanism of transcription for coronavirus structural and accessory genes yields a nested set of sg mRNAs that can be studied by northern blot analysis to show changes in the number, or size, of mRNAs produced as well as changes in mRNA expression levels as a result of alteration to the virus. A typical northern blot for the structural and accessory genes of Beau-R shows the presence of five mRNAs

plus one unidentified RNA species as originally described by Stern & Kennedy (Stern and Kennedy 1980) (Fig. 4.7).

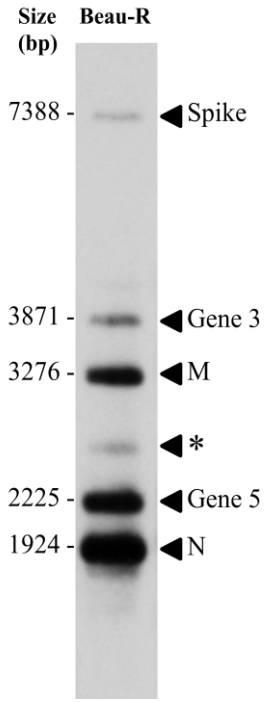


Fig. 4.7. Northern blot analysis of Beau-R. Intracellular RNA was harvested 24 hpi and northern blot analysis carried out. Left hand markers denote mRNA sizes of Beau-R structural and accessory genes. Right hand markers denote which genes are encoded on each mRNA. The * represents the unidentified RNA species of IBV.

To confirm the generation of an IBV sg mRNA capable of expressing eGFP or hRluc from each of the rIBVs, northern blot analyses of the rIBV sg mRNAs were carried out (Fig. 4.8). For each of the rIBVs the new expected sizes for each sg mRNA are summarized in Table 4.2.

Fig. 4.8. Characterization of rIBV mRNA Expression. Northern blot analysis was carried out 24 hpi with (A) BeauR-eGFPΔ5ab, (B) BeauR-hRlucΔ5ab and BeauR-IBVhRlucΔ5ab, (C) BeauR-eGFPΔIR, (D) BeauR-hRlucΔIR and (E) BeauR-hRlucΔ3ab. ◄ denotes wild type mRNA sizes. ◅ denotes altered mRNA sizes in rIBVs. * denotes unidentified RNA.

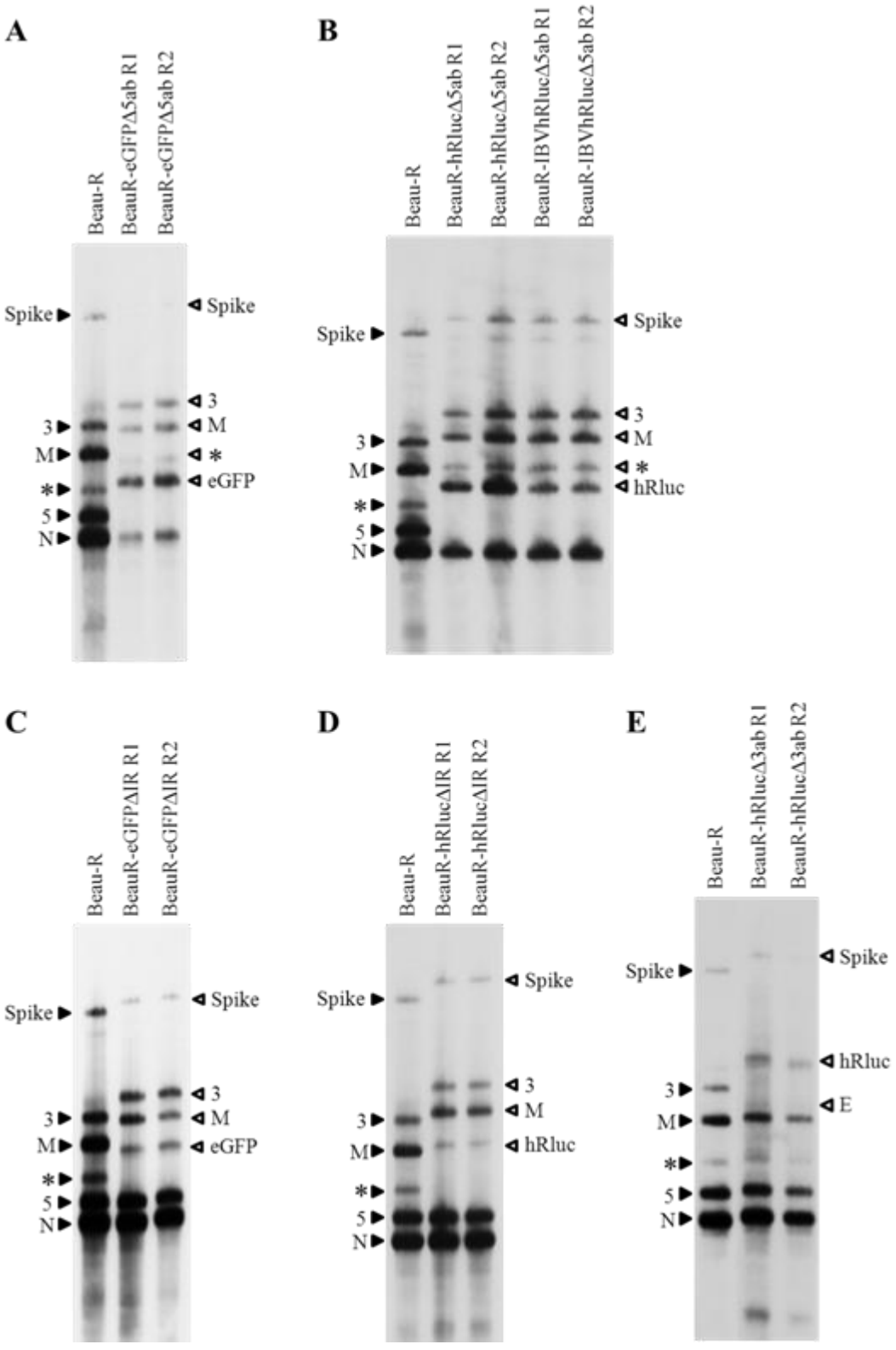

A
Beau-R
BeauR-eGFPΔ5ab R1
BeauR-eGFPΔ5ab R2
Spike
3
M
*
5
N
eGFP
B
Beau-R
BeauR-hRlucΔ5ab R1
BeauR-hRlucΔ5ab R2
BeauR-IBVhRlucΔ5ab R1
BeauR-IBVhRlucΔ5ab R2
Spike
3
M
*
5
N
hRluc
C
Beau-R
BeauR-eGFPΔIR R1
BeauR-eGFPΔIR R2
Spike
3
M
*
5
N
eGFP
D
Beau-R
BeauR-hRlucΔIR R1
BeauR-hRlucΔIR R2
Spike
3
M
*
5
N
hRluc
E
Beau-R
BeauR-hRlucΔ3ab R1
BeauR-hRlucΔ3ab R2
Spike
hRluc
E
3
M
*
5
N

TABLE 4.2: Recombinant IBV mRNA Sizes

rIBV	mRNA	Expected Size (nts)
BeauR-eGFPΔ5ab/ BeauR-IBVeGFPΔ5ab	*N*	1924
	eGFP	2733
	M	3784
	Gene 3	4379
	Spike	7896
BeauR-hRlucΔ5ab/ BeauR-IBVhRlucΔ5ab	*N*	1924
	hRluc	2949
	M	4000
	Gene 3	4595
	Spike	8112
BeauR-eGFPΔIR	*N*	1924
	Gene 5	2225
	eGFP	3005
	M	3786
	Gene 3	4381
	Spike	7898
BeauR-hRlucΔIR	*N*	1924
	Gene 5	2225
	hRluc	3221
	M	4002
	Gene 3	4597
	Spike	8114
BeauR-hRlucΔ3ab	*N*	1924
	Gene 5	2225
	M	3276
	E	3512
	hRluc	4489
	Spike	8006

The replacement of Gene 5 with eGFP or hRluc was expected to increase the size of the Gene 5 mRNA by 508nts or 724nts, respectively. This is shown in blots A and B, along with the subsequent increase in size of all upstream mRNAs as expected. Only the N mRNA remains unchanged for these rIBVs, as transcription of this mRNA is not affected by insertion of genes upstream. The increases in mRNA sizes confirmed the correct transcription of the reporter gene mRNAs from the Gene 5 location.

Recombinant viruses BeauR-eGFPΔIR and BeauR-hRlucΔIR utilized the insertion of a new TRS in order to obtain transcription of the reporter gene as an additional mRNA. Based on the insertion site of the reporter genes this new transcript was expected to be 3005nts for BeauR-eGFPΔIR and 3221nts for BeauR-hRlucΔIR. Northern blots C and D showed the presence of mRNA transcripts at the sizes expected for eGFP and hRluc and again the subsequent increase in size of all upstream mRNAs. It is interesting to note from these two blots that the band representing the unidentified RNA species (*) was no longer present for these rIBVs. Re-examination of blots A and B showed that this band also shifted in size with the mRNAs upstream of Gene 5 (as labelled). These observations suggested that the unidentified RNA species might be a mRNA linked to the IR. Investigation of these observations forms the work detailed in Chapters 6 and 7 of this thesis.

Analysis of rIBV BeauR-hRlucΔ3ab is shown in blot E. The replacement of ORFs 3a and 3b with hRluc was expected to yield a mRNA, representing the hRluc gene, of 4489nts. As described in section 3.2.4, when constructing this recombinant virus an additional TRS was inserted upstream of ORF 3c (E) to ensure transcription of a mRNA for the envelope protein, normally translated from the Gene 3 mRNA via an IRES. This was expected to generate a new mRNA transcript of 3512nts, in addition to the hRluc mRNA.

Both replicates showed the presence of a band representing hRluc although for R2 the mRNA appeared to be slightly smaller than for R1 suggesting a possible deletion in the hRluc mRNA. To investigate whether this observation was a result of a problem during the northern blot process or if there is a deletion in this replicate both

replicates were sequenced over the region of the hRluc insert. Sequence analysis showed that R2 contains a ~40nt region at the 3'-end of the hRluc gene (also covering the inserted TRS for transcription of E) that did not yield clean sequence traces, and showed a number of potential virus populations, as well as a potential deletion in the inserted E TRS (Fig. 4.9A).

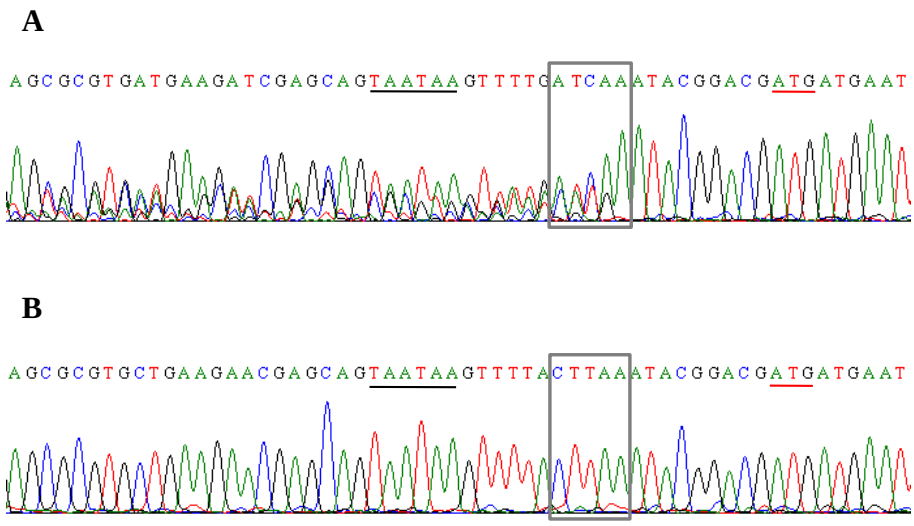


Fig. 4.9. Analysis of BeauR-hRlucΔ3ab. Passage 4 stocks of (A) BeauR-hRlucΔ3ab R2 and (B) BeauR-hRlucΔ3ab R1 were sequenced with primer BG143. Grey box highlights sequence that should correlate to E TRS. Underlined in black are the hRluc stop codons. Underlined in red is the E initiation codon.

The presence of mixed virus populations in this region may be enough to account for the apparent decrease in hRluc mRNA size for R2 if the majority of these populations contain deletions in the hRluc gene. This result could be linked back to the mixed plaque sizes observed for R2 in Fig. 4.1 and confirmed that the cause of the mixed virus populations is a region located at the 3'-end of the inserted hRluc gene. Furthermore, sequence analysis of R1 showed a 3nt deletion of CAA within the inserted TRS for E (Fig. 4.9B). A question was raised over whether transcripts for E were being generated by BeauR-hRlucΔ3ab as both replicates contained

sequence changes over the inserted TRS (Fig. 4.9). Close examination of the BeauR-hRlucΔ3ab northern blot (Fig. 4.8E) showed the faint detection of a mRNA for R1 at the size expected for the E mRNA. This mRNA was not observed for R2, although this may have been a result of the overall lower levels of mRNA detection for R2 due to differences in the amount of RNA loaded onto the gel for each replicate.

With no definitive proof of the E mRNA for either replicate it was decided to investigate this further via leader-body junction PCRs carried out on intracellular RNA extracted from CK cells infected with either R1 or R2 of BeauR-hRlucΔ3ab (see section 2.9 and Fig. 6.2 for PCR methodology). A PCR product relating to the E mRNA could not be amplified for R2, possibly due to the sequencing errors previously identified. For R1 a PCR product relating to the E mRNA was amplified, cloned and two colonies sequenced. The mRNA sequence obtained confirmed that E was being transcribed from R1 however, the TRS of the mRNA was non-canonical and its entire derivation was unclear (Fig. 4.10).

```
E-mRNA (1)          :ACACTAGCCTTGCGCTAGATTTTTAACTTACTTAAATACGGACGATGATGAATTT:85
E-mRNA (2)          :ACACTAGCCTTGCGCTAGATTTTTAACTTACTTAAATACGGACGATGATGAATTT:85
BeauR-hRlucΔ3ab R1:GTGCTGAAGAACGAGCAGTAATAAGTTTTACTTAAATACGGACGATGATGAATTT:24836
rVVBeauR-hRlucΔ3ab:GTGCTGAAGAACGAGCAGTAATAAGTTTTACTTAACAAATACGGACGATGATGAA:24836
```

Fig. 4.10. Messenger RNA sequence of E transcript. Two mRNAs representing the E ORF of BeauR-hRlucΔ3ab R1 were sequenced using primers Leader1 and BG143. rVVBeauR-hRlucΔ3ab shows the presence of the complete TRS-B for ORF E in the vaccinia virus IBV cDNA sequence. The genome sequence of BeauR-hRlucΔ3ab R1 shows the deletion in the TRS-B. The mRNA Leader sequence is highlighted in blue. Aligned sequences are highlighted in black with the ATG of E highlighted in green. Region of rVVBeauR-hRlucΔ3ab highlighted in grey shows the misalignment of genome due to the complete TRS-B. Yellow highlighted text represents the different E TRSs observed. Red highlighted text denotes sequence of the mRNA TRS from an unknown source. Genome and mRNA positions are given on the right.

The E mRNA sequences showed the presence of a shortened CTTAA TRS as highlighted in yellow. This TRS is derived from the altered TRS-B for E as seen in Fig. 4.9B. However, an additional CTTA sequence has also been added upstream of this TRS, with the derivation of this sequence less clear. The TTA of the additional sequence does align with the virus genome upstream of the CTTAA TRS-B suggesting transcription has continued past the standard TRS-B position to include additional nucleotides, possibly to compensate for the deletion at the 3'-end of the TRS-B. This still does not account for the addition of the C nucleotide, which may be a result of an aberrant hybridization event between the TR-L and the altered TRS-B, leading to transcription of nucleotides from the TRS-L. This explanation can be better understood in light of findings in Chapters 6 and 7. Given the presence of this mRNA for R1, and the belief that E is essential to virus replication, it is likely that E is translated by both virus replicates, even though a corresponding mRNA was undetected in R2.

Each of the rIBVs, with the exception of BeauR-hRlucΔ3ab R2, were shown to correctly transcribe a mRNA for the reporter gene with no apparent changes to mRNA expression of any other virus gene. Analysis of the rIBVs was therefore continued to confirm that synthesis of the modified mRNAs resulted in the correct translation of either eGFP or hRluc.

4.4 Reporter gene protein expression

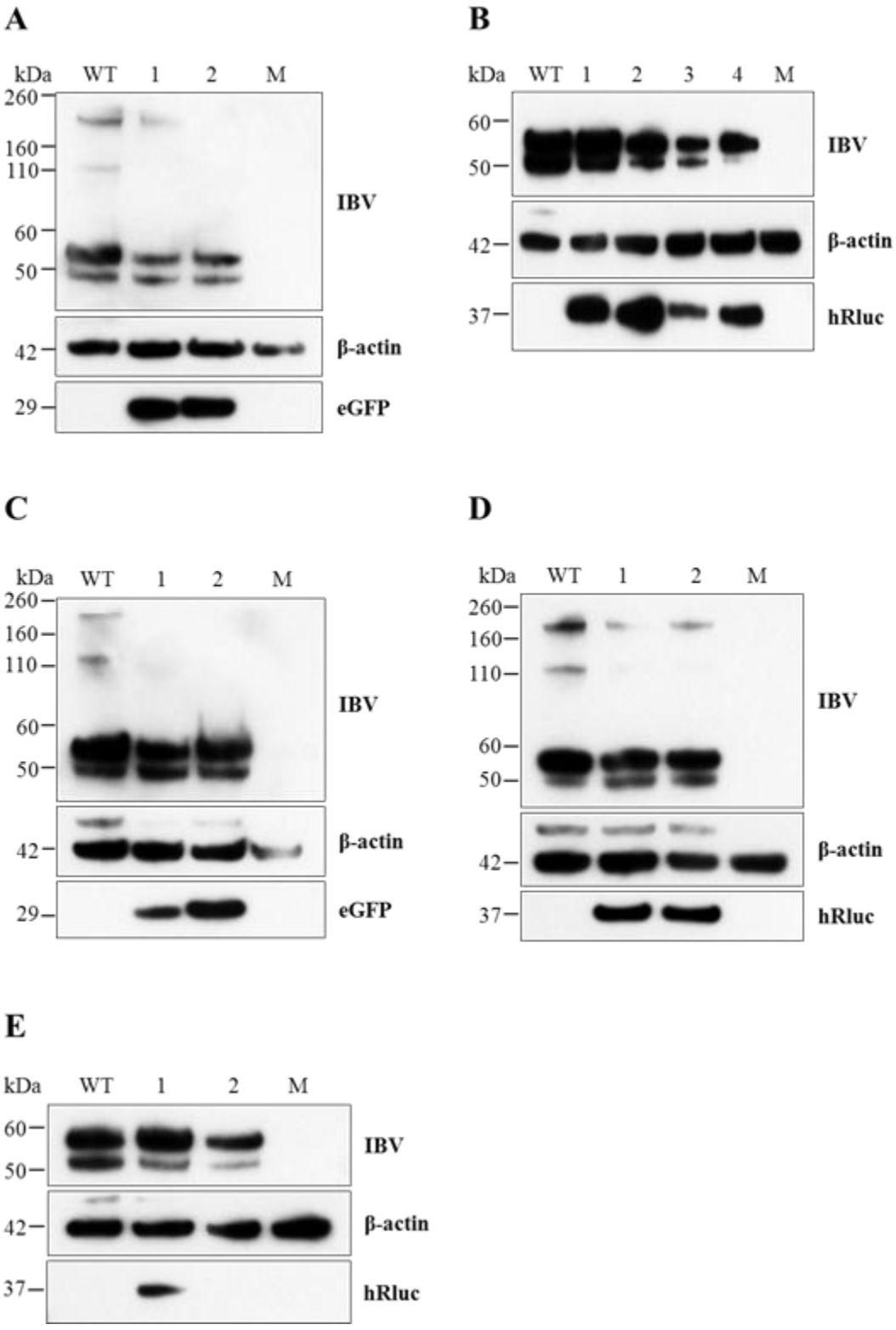
To confirm translation of a functional protein from the reporter gene mRNAs protein expression in CK cells was first analysed by western blot (Fig. 4.11). Correct protein

expression of eGFP and hRluc should yield proteins of 29kDa and 37kDa, respectively. BeauR-eGFPΔ5ab showed correct eGFP expression for both replicates with no expression from wild type or mock infected as expected (Fig. 4.11A). The same was true for hRluc expression from BeauR-hRlucΔ5ab and BeauR-IBVhRlucΔ5ab (Fig. 4.11B). Differences in the range of IBV proteins expressed (top panels) are due to the polyclonal nature of the anti-IBV primary antibody and varying sensitivities of blots from different infections.

Reporter protein expression from BeauR-eGFPΔIR and BeauR-hRlucΔIR was also shown to be correct for both replicates (Fig. 4.11C, D). For BeauR-hRlucΔ3ab luciferase protein expression could only be detected for R1 (Fig. 4.11E). This is consistent with the results seen with mRNA expression and sequence analysis of both replicates, suggesting that the problems seen at the 3'-end of the hRluc gene in R2 are affecting hRluc protein expression.

Protein expression was further analysed by confocal microscopy to confirm fluorescence from eGFP rIBVs, and to see if any differences could be observed in reporter protein distribution within the cell for any of the rIBVs. Confocal microscopy images are shown for infection of Vero cells as difficulties were encountered in obtaining clear images from primary CK cells due to the presence of high levels of cell debris. Confocal images confirmed expression of eGFP or hRluc in Vero cells following infection with R1 of each rIBV (Fig. 4.12). In all constructs proteins eGFP and hRluc were found distributed throughout the cytoplasm and nuclei of infected cells.

Fig. 4.11. Characterization of rIBV protein expression. Western blot analyses of rIBV infected CK cell lysates. (A) BeauR-eGFPΔ5ab, (B) BeauR-hRlucΔ5ab and BeauR-IBVhRlucΔ5ab (C) BeauR-eGFPΔIR (D) BeauR-hRlucΔIR (E) BeauR-hRlucΔ3ab. Each group of cell lysates were run on three separate gels and corresponding blots probed with either anti-IBV (top panels), anti-actin (middle panels) or, anti-eGFP (A, C) or anti-hRluc (B, D, E) (bottom panels). Lanes: (A, C, D, E) WT, Beau-R; 1, R1; 2, R2; M, mock; (B) WT, Beau-R; 1, BeauR-hRlucΔ5ab R1; 2, BeauR-hRlucΔ5ab R2; 3, BeauR-IBVhRlucΔ5ab R1; 4, BeauR-IBVhRlucΔ5ab R2; M, mock.


A
kDa WT 1 2 M
260
160
110
60
50
42
29
IBV
β-actin
eGFP
B
kDa WT 1 2 3 4 M
60
50
42
37
IBV
β-actin
hRluc
C
kDa WT 1 2 M
260
160
110
60
50
42
29
IBV
β-actin
eGFP
D
kDa WT 1 2 M
260
160
110
60
50
42
37
IBV
β-actin
hRluc
E
kDa WT 1 2 M
60
50
42
37
IBV
β-actin
hRluc

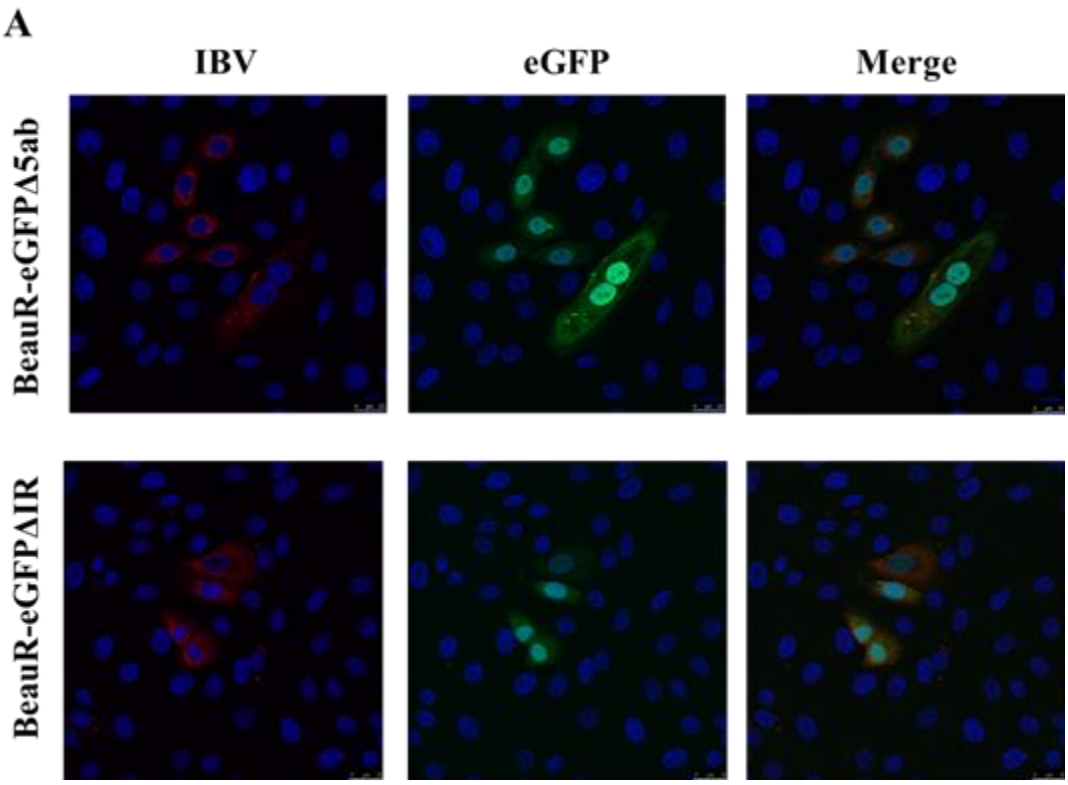
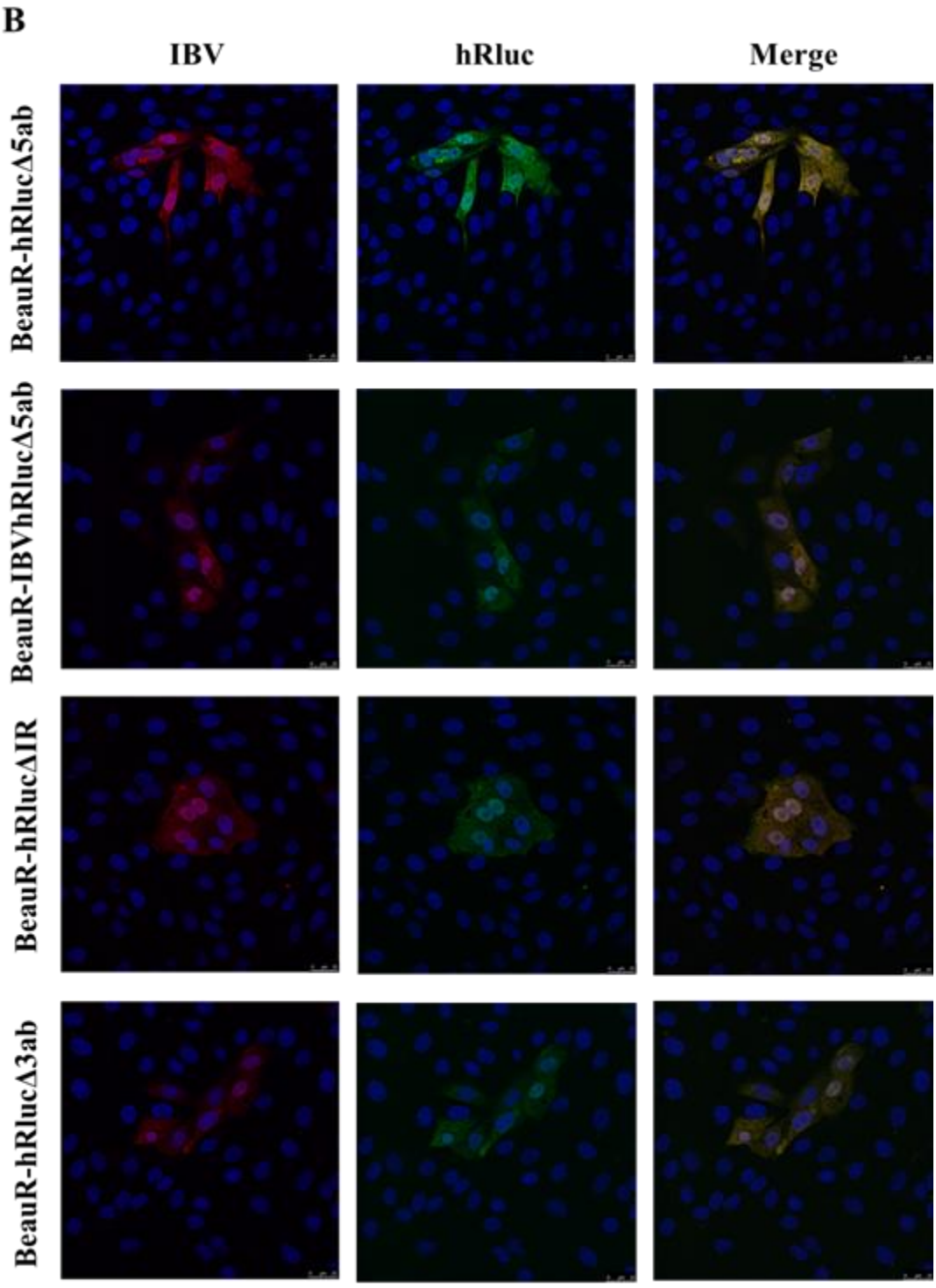


Fig. 4.12. Confocal microscopy of rIBVs. Monolayers of Vero cells at 80% confluency were infected with $1x10^5$ pfu of (A) eGFP rIBVs or (B) hRluc rIBVs. Cells were fixed 18 hpi with 4% paraformaldehyde. Cells were labelled with anti-IBV (red). For (A) eGFP was detected by fluorescence (green). For (B) cells were additionally labelled with anti-hRluc (green). Nuclei were stained blue with DAPI. Scale bars represent 25µm.


B
IBV
hRluc
Merge
BeauR-hRlucΔ5ab
BeauR-IBVhRlucΔ5ab
BeauR-hRlucΔIR
BeauR-hRlucΔ3ab

4.5 Luciferase expression

As luciferase expression can be quantified using a luminometer, the rIBVs expressing hRluc could be used as tools to investigate whether heterologous protein expression levels varied based on the location of the heterologous gene. Utilizing cell lysates harvested at the same time as extracellular virus for growth curves, protein expression was analysed from each of the rIBVs expressing hRluc (Fig. 4.13).

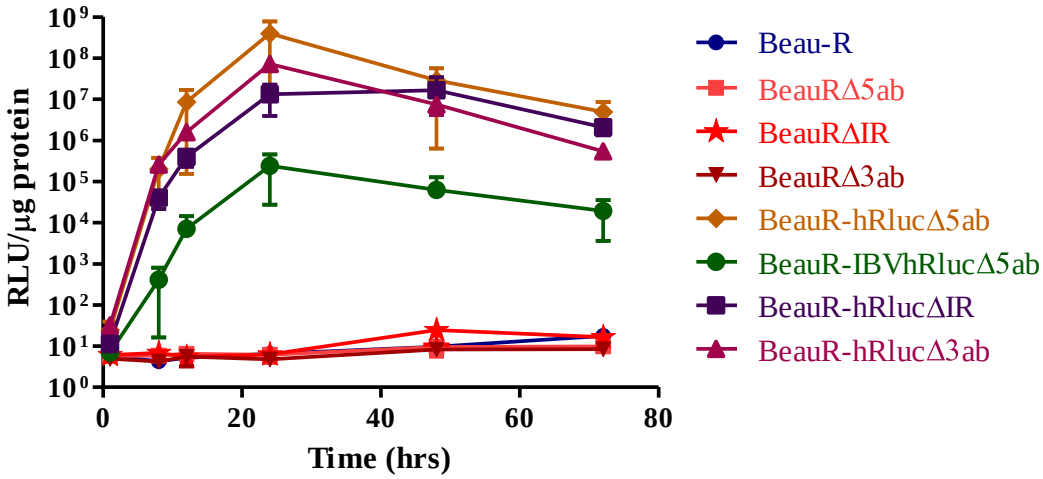


Fig. 4.13. Luciferase expression. Luciferase expression was measured in cell lysates taken at various time points during infection and expressed as relative light units (RLU) per µg of protein. Error bars represent SD of triplicate readings of three independent experiments.

Protein expression peaked for all rIBVs at 24 h except BeauR-hRlucΔIR, which peaked at 48 h. This coincides with peak virus titres obtained for rIBVs, except in the case of BeauR-hRlucΔIR. Comparing the 24 h time point data luciferase

expression was found to be highest in BeauR-hRlucΔ5ab followed by BeauR-hRlucΔ3ab and then BeauR-hRlucΔIR.

Surprisingly, BeauR-IBVhRlucΔ5ab was found to have the lowest levels of luciferase expression with peak levels approximately 3.5-$\log_{10}$ lower than for BeauR-hRlucΔ5ab. Codon optimization of the heterologous gene had been hypothesised to improve virus stability but was not expected to decrease protein expression from that gene. Due to the differences observed between BeauR-hRlucΔ5ab and BeauR-IBVhRlucΔ5ab sequence analysis of all Δ5ab rIBVs was carried out at this time to check that no errors had occurred within the modified regions of the rIBV genomes.

Sequence analysis showed that a mutation had occurred in the hRluc gene of BeauR-IBVhRlucΔ5ab R1 that was not present in either of the BeauR-hRlucΔ5ab replicates or R2 of BeauR-IBVhRlucΔ5ab. This mutation consisted of the insertion of two additional T nucleotides in a stretch of eight T nucleotides beginning at genome position 26267. The additional residues caused a frameshift and alteration to the hRluc amino acid sequence downstream of amino acid 262 (Fig. 4.14). This frameshift also resulted in the addition of eight amino acids to the 3'-end of the hRluc gene before a stop codon was reached. The modified hRluc gene generated by these changes was referred to as hRluc*.

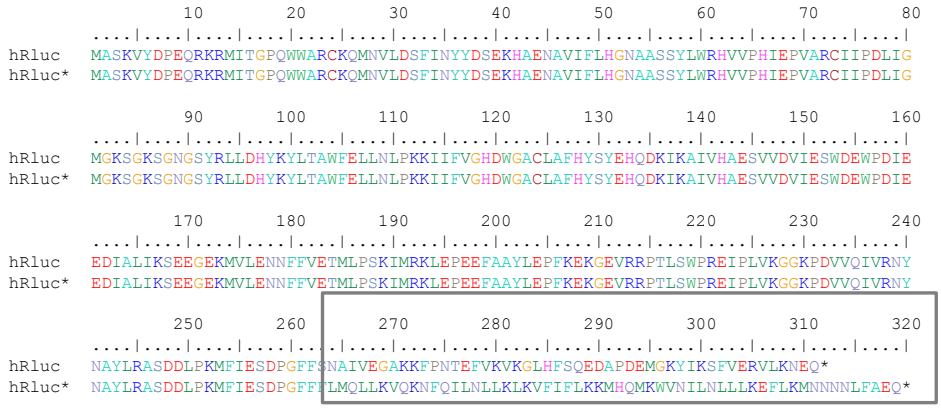


Fig. 4.14. Amino acid alignment of codon optimized hRluc. The altered hRluc gene (hRluc*) of BeauR-IBVhRlucΔ5ab R1 was sequenced and compared to wild type hRluc. Grey box highlights the amino acid positions affected by the mutations arising in hRluc*.

To investigate whether the changes in the hRluc gene of BeauR-IBVhRlucΔ5ab R1 were responsible for the lower than expected levels of luciferase expression the time course experiment was carried out for R2 (Fig. 4.15).

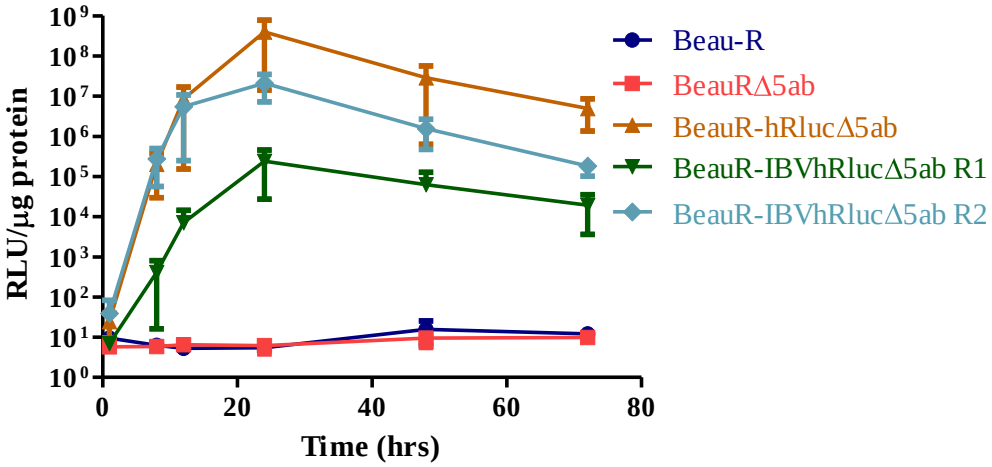


Fig. 4.15. Codon optimized luciferase expression. Luciferase expression was measured in cell lysates taken at various time points during infection and expressed as relative light units (RLU) per µg of protein. Error bars represent SD of triplicate readings of three independent experiments.

Luciferase expression for BeauR-IBVhRlucΔ5ab R2 was found to be greater than R1 suggesting that the changes within the hRluc gene were likely responsible for the lower levels of luciferase expression. However, expression levels were diminished slightly by 72 h compared to BeauR-hRlucΔ5ab; highlighting that in this case codon optimization may have a small negative impact on protein expression.

4.6 Utilizing eGFP rIBVs for live cell imaging studies

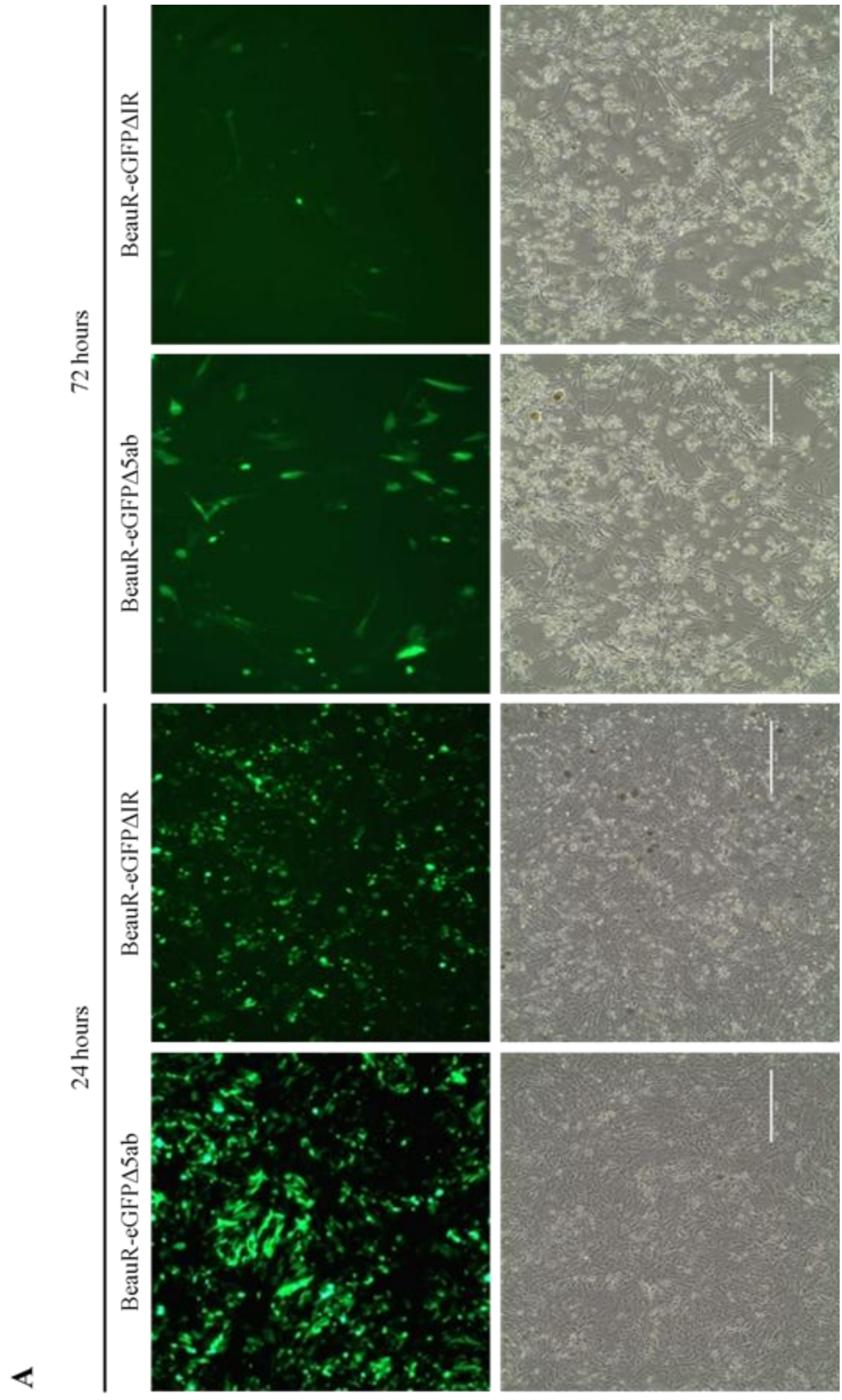
Having obtained two recombinant viruses expressing eGFP from different regions of the IBV genome it was decided to try and use these viruses as tools to study aspects of IBV infection by live cell imaging. As no previous publications or information could be found describing the use of CK cells for live cell imaging it was decided to first optimize infection of these viruses on Vero cells, as protocols were already in place for this cell culture. A trial infection was carried out, and monitored by standard fluorescence microscopy, to obtain timeframes for live cell imaging. However, it was observed that BeauR-eGFPΔ5ab did not propagate as well on Vero cells when compared to BeauR-eGFPΔIR with very few fluorescing cells visible.

To investigate this effect further, infections of BeauR-eGFPΔ5ab and BeauR-eGFPΔIR on both CK cells and Vero cells were carried out and monitored for 72 h (Fig. 4.16). At 24 h both viruses on CK cells showed wide spread virus infection as indicated by the high number of green fluorescent cells. By 72 h the number of fluorescent cells had decreased rapidly due to high levels of cell death, as expected for CK cells, and confirmed by bright field microscopy (Fig. 4.16A, lower panels). In contrast, on Vero cells at 24 h very few infected cells were observed for BeauR-

eGFPΔIR and only one or two infected cells observed for BeauR-eGFPΔ5ab (Fig. 4.16B).

At 72 h on Vero cells there was a distinct difference in the progression of virus infection between the two rIBVs. While BeauR-eGFPΔIR had spread throughout the cell monolayer, with large foci of infection and separate, individually infected cells visible, BeauR-eGFPΔ5ab was limited to very small, discrete foci of infection. These observations indicated that the loss of Gene 5 limits the ability of Beau-R to spread on Vero cells. Taken together with the plaque morphology data and the slight delay in replication kinetics observed for rIBVs lacking Gene 5 this suggested a role for one, or both, of the Gene 5 accessory proteins in cell-cell spread.

Following these observations it was decided not to pursue the use of BeauR-eGFPΔ5ab in live cell imaging as it could not be guaranteed that the effects of the loss of Gene 5 would not impact on the results obtained and yield inaccurate data. Additionally, whilst growth on Vero cells was not restricted for BeauR-eGFPΔIR the levels of fluorescence observed were considerably lower with long exposure times required to obtain images. Although slightly better fluorescence levels were observed in CK cells it was considered that the quality of images that could be obtained for BeauR-eGFPΔIR was not good enough to proceed with live cell imaging. For these reasons no live cell imaging experiments were carried out.


A
24 hours
72 hours
BeauR-eGFPΔ5ab
BeauR-eGFPΔIR
BeauR-eGFPΔ5ab
BeauR-eGFPΔIR

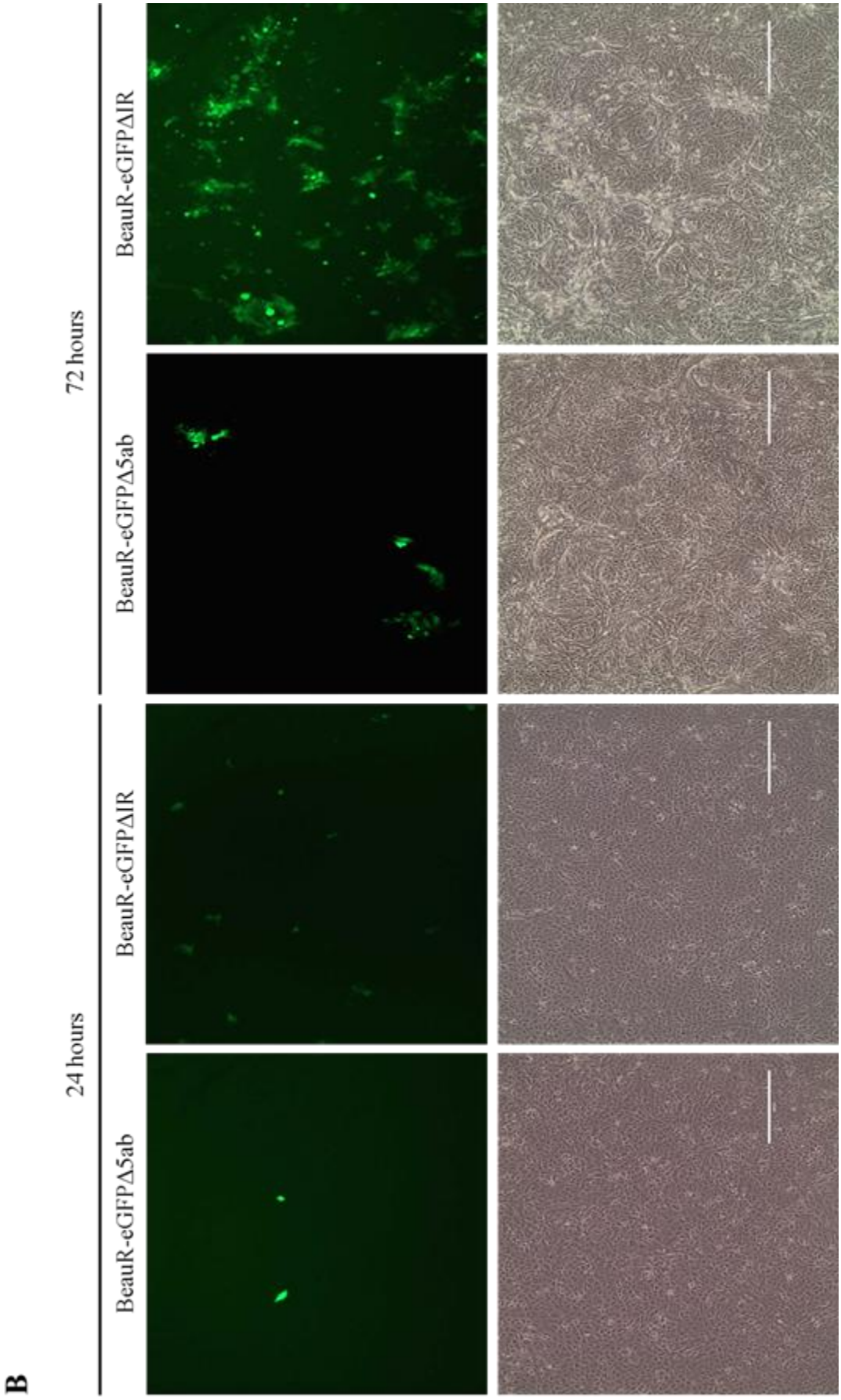

B
24 hours
72 hours
BeauR-eGFPΔ5ab
BeauR-eGFPΔIR
BeauR-eGFPΔ5ab
BeauR-eGFPΔIR

Fig. 4.16. Virus infection of Vero cells. Monolayers of CK (A) or Vero (B) cells at 80% confluency were infected with $1x10^5$ pfu of BeauR-eGFPΔ5ab or BeauR-eGFPΔIR. Images were taken at 24 h and 72 h by fluorescence microscopy (upper panels) at 10x magnification, or bright field microscopy (lower panels) where scale bars represent 400μm.

4.7 Summary

This chapter detailed the analysis of the rescued rIBVs in terms of growth characteristics and protein expression. The main findings from this chapter are summarized below:

- BeauR-eGFPΔ5ab and BeauR-hRlucΔ5ab replicated similarly to wild type but with 1-$\log_{10}$ lower peak titres and a delay in replication at 8 h. All virus replicates correctly transcribed and expressed reporter genes.
- BeauR-IBVhRlucΔ5ab replicated as other Δ5ab rIBVs but codon optimization resulted in diminished hRluc expression compared to BeauR-hRlucΔ5ab.
- BeauR-eGFPΔIR replicated with similar kinetics to wild type but with ~2-$\log_{10}$ lower titres from 24 h onwards. Replicate 1 was found to have a mutation leading to the additional loss of accessory protein 5b. Both virus replicates correctly transcribed and expressed eGFP.
- BeauR-hRlucΔIR replicated as wild type and both virus replicates correctly transcribed and expressed hRluc.
- BeauR-hRlucΔ3ab replicated to 1-$\log_{10}$ lower peak titres than wild type but both virus replicates contained sequence errors in, or around, the inserted hRluc gene. Replicate 2 did not express hRluc due to errors.
- The loss of Gene 5 resulted in an inability of Beau-R to undergo cell-cell spread on Vero cells.

As with Chapter 3, these results will be discussed in detail in an overall discussion at the end of Chapter 5.

Chapter 5: Analysis of rIBV stability in cell culture

5.1 Overview

This chapter details the experiments that were carried out to investigate the stability of the rescued rIBVs. By passaging each virus on CK cells the aim was to identify which rIBV constructs resulted in the expression of heterologous genes for the greatest number of passages. Initially both replicates of each rIBV were passaged, regardless of any issues identified in Chapter 4, in order to more accurately establish the stability of each construct. For the purposes of this chapter the rIBVs will be referred to as the wild type viruses with Beau-R referred to as the parental virus.

5.2 Recombinant IBV passaging

Initial experiments were carried out with rIBVs BeauR-eGFPΔ5ab, BeauR-hRlucΔ5ab, BeauR-eGFPΔIR and BeauR-hRlucΔIR as the first rIBVs to have been rescued. Recombinant viruses BeauR-IBVhRlucΔ5ab and BeauR-hRlucΔ3ab were rescued at a later date and therefore not included in initial experiments. The P3 stock of each virus was taken and blind passaged a further five times on CK cells. At each passage virus was diluted 1:100 prior to infection, to ensure the previous passage stock was not depleted, but no quantification of the virus inoculum was carried out.

All passaging was carried out in 6-well plates of confluent CK cells. Each rIBV was used to infect one well of CK cells (500μL/well). Cells were incubated for 1 h after which virus inoculum was removed and fresh medium added to cells. At 36 hpi

extracellular virus was harvested from cells and reverse transcribed for PCR analysis. Stability was measured by the presence or absence of the reporter gene, based on visualisation of PCR products as previously described in section 3.3, Table 3.2. The expected PCR products for the rIBVs specific to this chapter are listed again in Table 5.1.

TABLE 5.1: Expected PCR Products for Wild Type Virus

Virus	Primer Set	PCR Product (bp)
BeauR-eGFPΔ5ab	5abS + BG147	963
BeauR-hRlucΔ5ab/ BeauR-IBVhRlucΔ5ab	5abS + BG147	1179
BeauR-eGFPΔIR	BG52 + BG146	1087
BeauR-hRlucΔIR	BG52 + BG146	1303
BeauR-hRlucΔ3ab	BG49 + BG144	1916

All RT-PCR products from P3 to P8 for BeauR-eGFPΔ5ab, BeauR-hRlucΔ5ab, BeauR-eGFPΔIR and BeauR-hRlucΔIR were subsequently analysed by gel electrophoresis at the same time to observe any changes in the detection of reporter gene PCR products (Fig. 5.1). A single PCR product of the expected size at each passage would represent a stably replicating virus. The presence of smaller PCR products would indicate deletions were occurring somewhere over the region of the PCR. The results for these rIBVs showed that even at early passages (P3-P5) a number of smaller PCR products could be seen for each virus, indicating the loss of genomic sequence over the region of the PCR. This suggested that the reporter genes were being rapidly deleted from the genome and that the rIBVs were unstable.

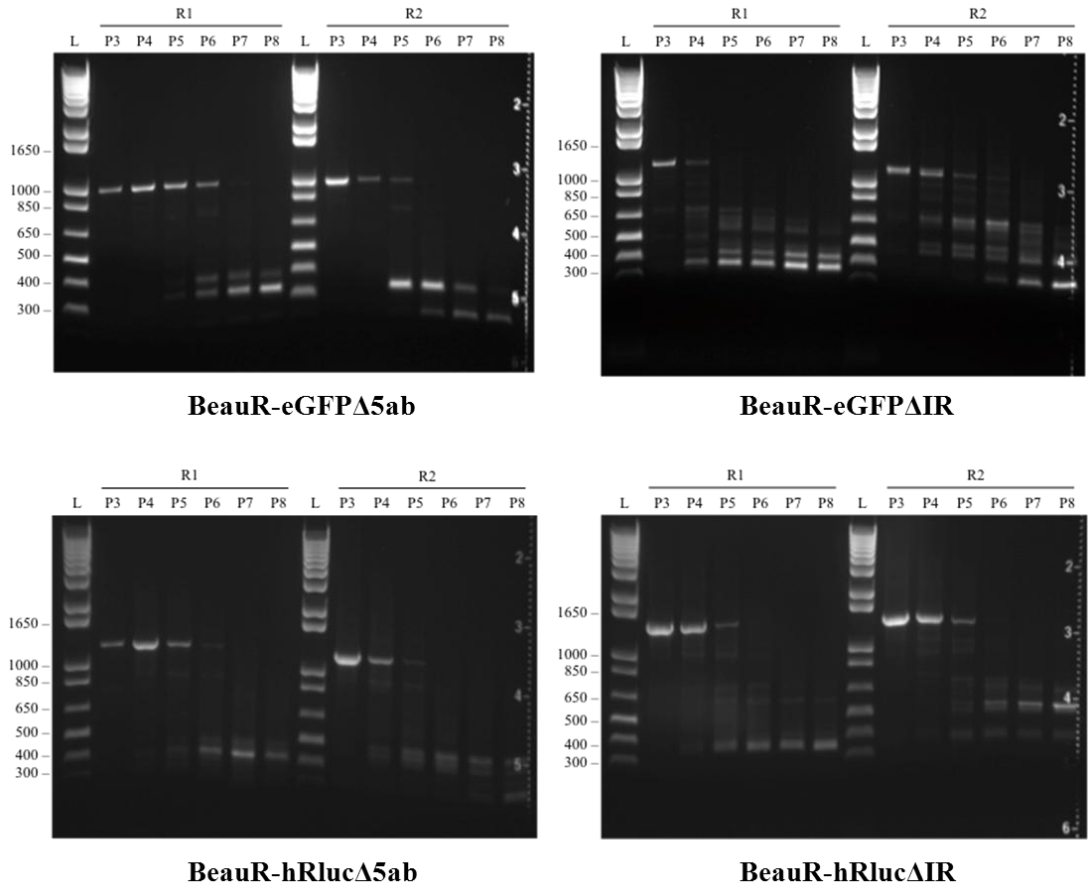


Fig. 5.1. Initial passaging of rIBVs. Recombinant viruses were blind passaged on CK cells. Confluent monolayers were infected with 500µl rIBV, extracellular virus was harvested 36 hpi and screened by RT-PCR. Gel electrophoresis shows PCR results for R1 and R2 of each virus from P3 to P8. DNA marker is shown as lane L with selected values identified on the left.

Complete loss of wild type PCR product was observed by P7 for BeauR-eGFPΔ5ab, P5 for BeauR-eGFPΔIR, P7 for BeauR-hRlucΔ5ab and P6 for BeauR-hRlucΔIR, although PCR products representing virus populations containing suspected deletions could be seen before said passages. Both R1 and R2 of each rIBV showed instability arising in the virus within one or two passages of each other, as would be expected for replicates of the same virus, and confirmed that both replicates are under the same replication pressures.

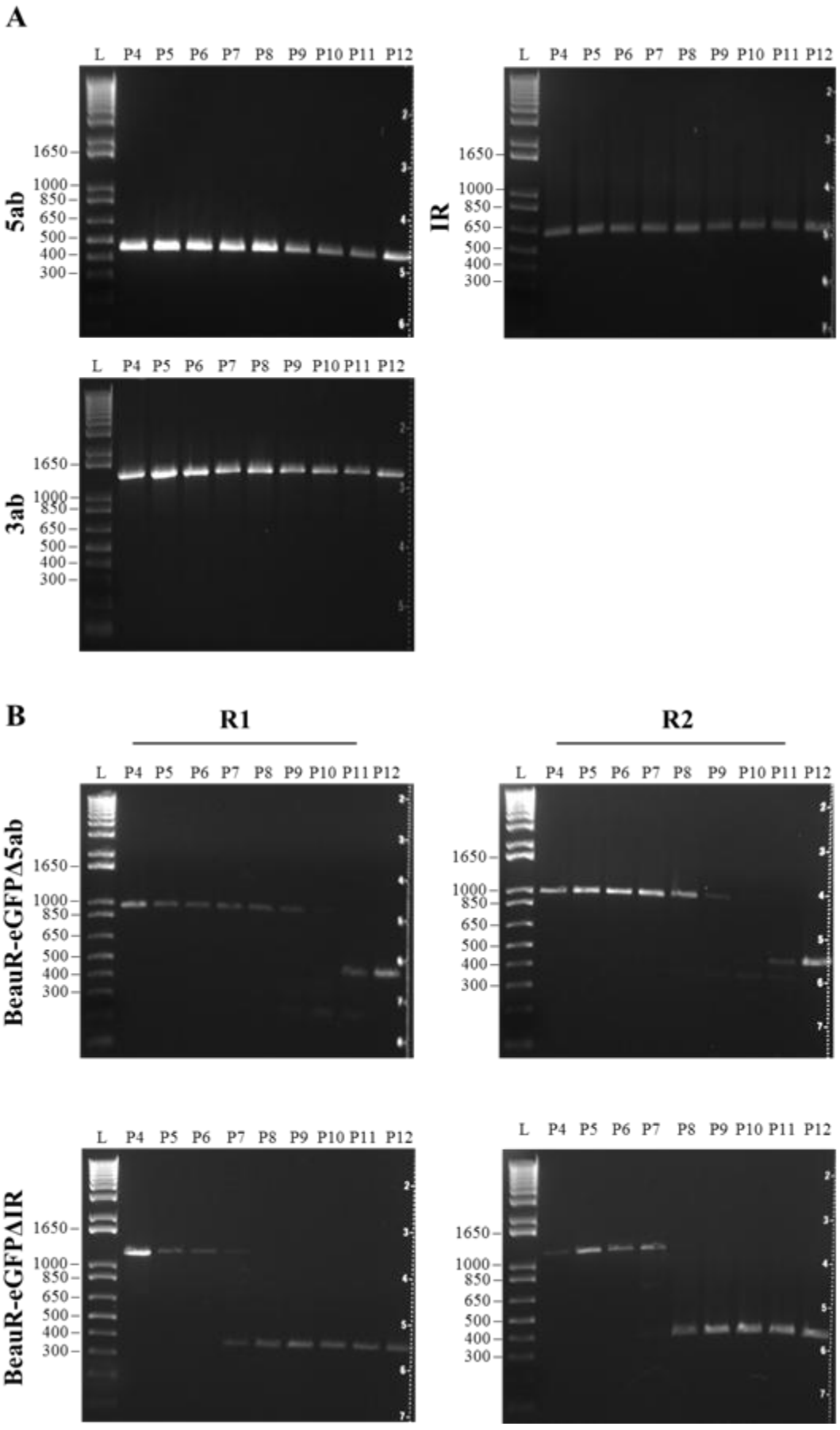
Due to the nature of the passaging, at potentially high virus titres, it was initially hypothesised that these suspected deletions in the reporter genes might be occurring in conjunction with other deletions across the virus genome, and thus leading to the formation of D-RNAs. Defective RNAs have previously been reported to occur for coronaviruses, including IBV, following serial passage in cell culture (Makino, *et al.* 1985, Penzes, *et al.* 1994). It is known that to control the generation of D-RNAs viruses must be passaged at a low MOI to prevent the smaller, faster replicating, defective genomes from being packaged and passaged on with the aid of helper, intact, virus genomes. Control of D-RNAs can also be achieved by simple dilution of virus stocks (without the need for calculation of an MOI) as this is known to reduce the co-infection of potential D-RNAs with helper viruses and therefore reduce, or abolish, subsequent passage of D-RNAs.

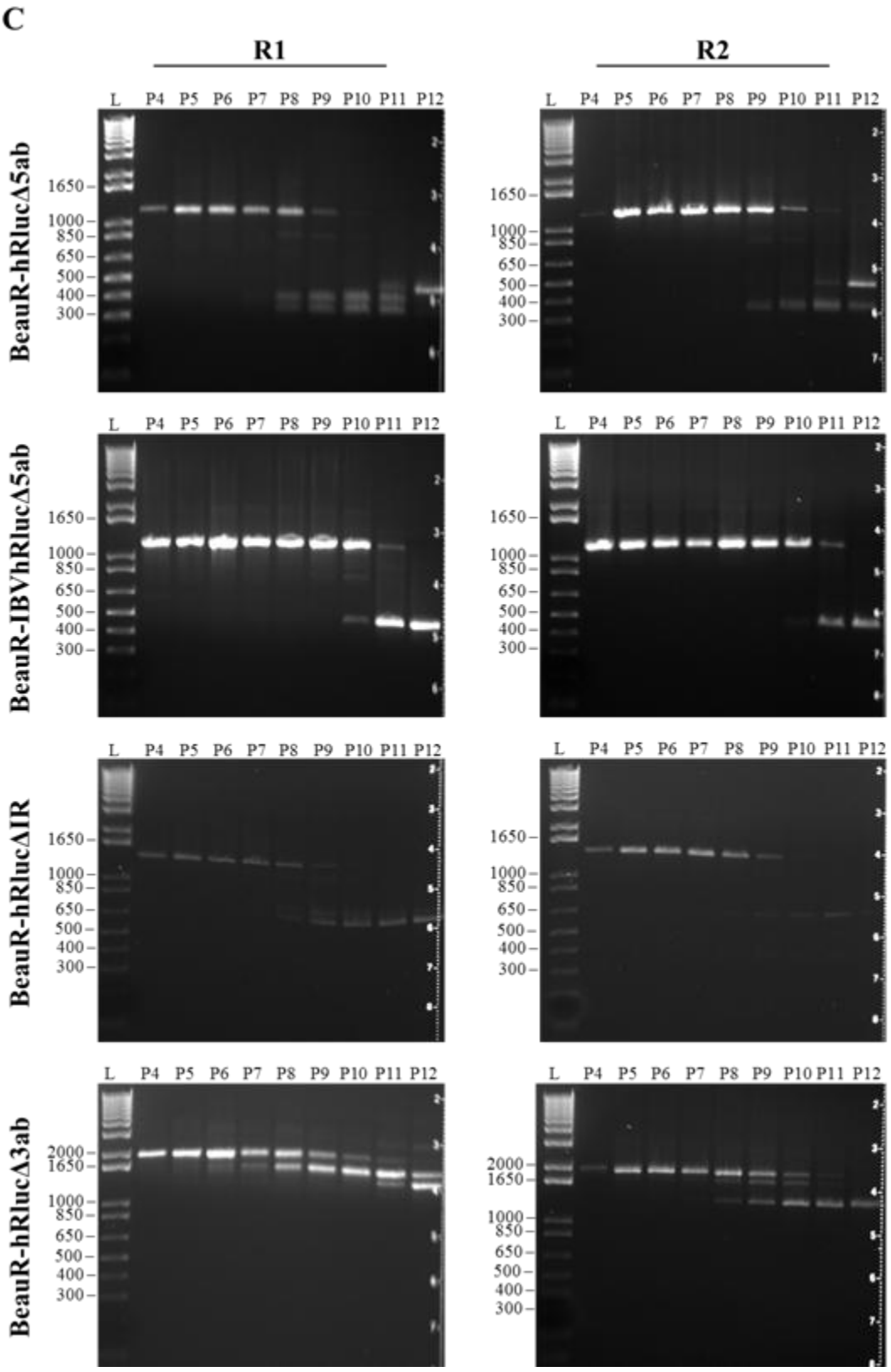
To investigate if D-RNAs were the cause of the observed rIBV instability the rescue process was repeated from P1 for each rIBV to establish new virus stocks at P4. This time each rIBV was subject to greater dilution prior to infection of CK cells at each passage. Due to potentially low virus titres at P1, P2 was carried out using a dilution

of 1:1000, followed by P3 and P4 stocks at a dilution of 1:10,000. The expectation was that these new P4 stocks would be free of any D-RNAs and would give a more reliable estimate of the stability of each rIBV when passaged. This process was also carried out with the additional rIBVs, BeauR-IBVhRlucΔ5ab and BeauR-hRlucΔ3ab, that had been subsequently rescued.

To examine the effect on stability by passaging from potentially D-RNA-free stocks of virus the passaging was repeated for BeauR-eGFPΔ5ab, BeauR-hRlucΔ5ab, BeauR-eGFPΔIR and BeauR-hRlucΔIR from P4 onwards. Those viruses not previously passaged, BeauR-IBVhRlucΔ5ab and BeauR-hRlucΔ3ab, were also included and passaging of all rIBVs was extended to P12. Parental virus Beau-R was also passaged and screened with each primer set to confirm whether deletions were arising due to the insertion of heterologous genes, or if deletions occur naturally over these regions when Beau-R is passaged (Fig. 5.2). As previously, viruses were diluted 1:100 from P4 onwards to ensure the previous passage stock was not depleted.

Passaging of Beau-R, under the conditions described, showed that deletions did not naturally occur within the unmodified genome regions, with the expected PCR products of 455bp, 623bp and 1324bp observed at all passages for the 5ab, IR and 3ab PCRs respectively (Fig. 5.2A). The detection of the non-wild type PCR products from the rIBVs, and the deletion of reporter genes, is therefore a result of the modifications made to the virus genomes.


A
5ab
IR
3ab
L P4 P5 P6 P7 P8 P9 P10 P11 P12
1650
1000
850
650
500
400
300
B
R1
R2
BeauR-eGFPΔ5ab
BeauR-eGFPΔIR


C
R1
R2
L P4 P5 P6 P7 P8 P9 P10 P11 P12
BeauR-hRlucΔ5ab
BeauR-IBVhRlucΔ5ab
BeauR-hRlucΔIR
BeauR-hRlucΔ3ab
2000
1650
1000
850
650
500
400
300

Repeat passaging of the rIBVs with potentially D-RNA-free stocks resulted in an overall improvement in stability for rIBVs BeauR-eGFPΔ5ab, BeauR-hRlucΔ5ab, BeauR-eGFPΔIR and BeauR-hRlucΔIR, as measured by the passage number at which the first non-wild type PCR products appeared, and at which the wild type product was no longer detected. Recombinant IBVs BeauR-IBVhRlucΔ5ab and BeauR-hRlucΔ3ab, which had not been analysed previously, also showed the accumulation of non-wild type PCR products over the course of passaging, indicating that all of the rIBVs are subject to deletion of the heterologous reporter gene sequences.

To more easily compare the data from the first two experiments the passage number at which instability was first detected is summarized for each rIBV in Table 5.2, along with the passage numbers at which the wild type product became undetectable.

TABLE 5.2: Recombinant IBV Stability

Virus	R1		R2	
	Instability Arises	**Reporter Gene Lost**	**Instability Arises**	**Reporter Gene Lost**
BeauR-eGFPΔ5ab				
Initial Passaging	*P5*	*P7*	*P5*	*P6*
Repeat Passaging	P9	P11	P8	P10
BeauR-hRlucΔ5ab				
Initial Passaging	*P5*	*P7*	*P4*	*P6*
Repeat Passaging	P8	P10	P9	P12
BeauR-eGFPΔIR				
Initial Passaging	*P4*	*P5*	*P4*	*P6*
Repeat Passaging	P7	P8	P7	P8
BeauR-hRlucΔIR				
Initial Passaging	*P4*	*P7*	*P4*	*P6*
Repeat Passaging	P8	P10	P8	P10
BeauR-IBVhRlucΔ5ab	P10	P12	P10	P12
BeauR-hRlucΔ3ab	P5	P12	P7	P12

If stability is defined by a single measurement, such as the passage number at which a non-wild type PCR product was first detected, then those rIBVs with reporter genes replacing Gene 5 are, on average, the most stable. Only one passage separates the instability observed for BeauR-eGFPΔ5ab from BeauR-hRlucΔ5ab suggesting that the heterologous gene itself has little impact on virus stability. Comparing the passaging data of BeauR-hRlucΔ5ab and BeauR-IBVhRlucΔ5ab (Table 5.2) it appears that codon optimization of the heterologous gene may have a small positive effect on stability with two additional passages achieved before non-wild type PCR products were observed.

The replacement of the IR resulted in rIBVs that were only one or two passages less stable than those containing a replacement of Gene 5. Again, only one passage separated the eGFP and hRluc recombinants confirming that once rescued the heterologous gene has a limited impact on virus stability. The least stable construct was found to be the replacement of ORFs 3a and 3b (BeauR-hRlucΔ3ab), with non-wild type PCR products detectable as early as P5. Interestingly however, this virus displayed the highest number of passages between the appearance of non-wild type PCR products and complete loss of the wild type PCR product.

It was apparent from these results that once instability had arisen within the virus population, detection of wild type virus expressing the reporter gene was lost after one or two further passages, with the exception of BeauR-hRlucΔ3ab. This suggested that once instability arises there is rapid selection of a virus population in which the reporter gene has been deleted and the virus is, presumably, able to replicate more efficiently. By beginning the passaging with a P4 stock generated by

dilution of the rIBV between P2 and P4, an additional 3-4 passages were achieved prior to the appearance of instability of the virus. This suggested that some control could be exerted on the rIBVs to maintain a stable virus population for longer, and reinforced the idea of D-RNAs arising during passaging.

Given the positive impact on stability observed by using a potentially D-RNA-free stock as the start point it was hypothesised that controlling the MOI at every passage would additionally benefit rIBV stability. To investigate this possibility a third passaging experiment was carried out with R1 of each of the hRluc rIBVs. This time, instead of blind passaging with a 1:100 dilution of virus, each rIBV was titrated on CK cells between passages and all infections were carried out at a MOI of 0.01. Recombinants were again passaged from P4 to P12 (Fig. 5.3).

The greatest effect of passaging at a set MOI was seen with BeauR-hRlucΔ5ab and BeauR-IBVhRlucΔ5ab where no non-wild type PCR products were observed up to, or including at, P12. This equated to an increase in stability of four rounds of passaging compared to the blind passaging experiment shown in Fig. 5.2. Although the image for BeauR-IBVhRlucΔ5ab appears to show a reduction in size of the PCR product this was in fact caused by a technical error during electrophoresis resulting in an uneven electrical flow through the gel.

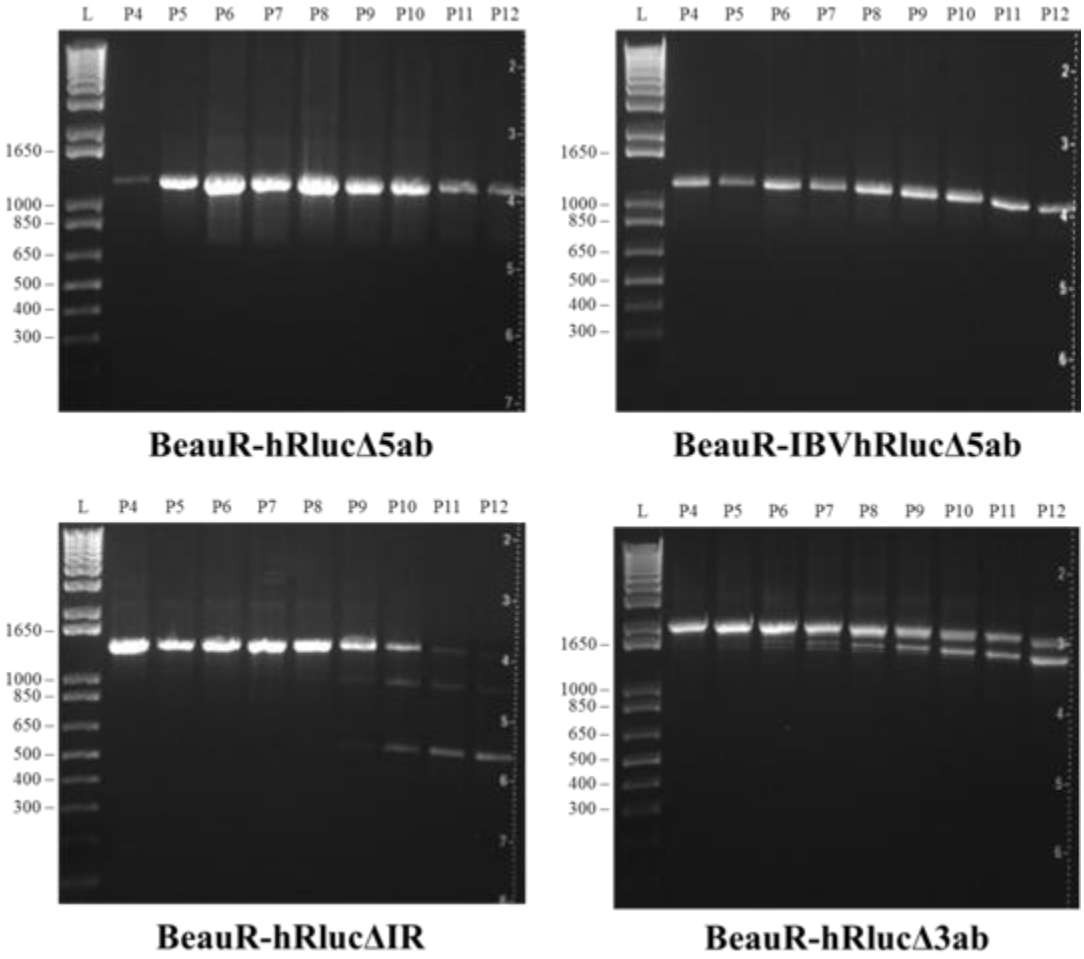


Fig. 5.3. MOI passaging of hRluc recombinants. Replicate 1 of each hRluc recombinant was re-passaged on CK cells at a MOI of 0.01. Virus was screened by RT-PCR from P4 to P12. DNA marker is shown as lane L with selected values identified on the left.

Passaging at a set MOI also had a small impact on BeauR-hRlucΔIR with non-wild type PCR products detected from P9 compared to P8 and wild type PCR products still detectable, although very faint, at P12. Recombinant BeauR-hRlucΔ3ab was again shown to be the least stable with non-wild type PCR products detected from P6. Similar to the previous experiment however, the wild type population was still clearly detected at P12 suggesting that although deletion of the reporter gene occurs early in this construct there is some selection pressure to also maintain the wild type population.

The MOI passaging data confirms what was suspected from the blind passaging experiments that the replacement of Gene 5 yields the most stable virus recombinants, followed by the replacement of the IR and then replacement of ORFs 3a and 3b. It is possible that codon optimization of the heterologous gene does improve rIBV stability but more data is required to fully support this hypothesis. The stability of the rIBVs was not found to be determined by the specific heterologous gene in these experiments, with similar results observed for both eGFP and hRluc recombinants.

5.3 Investigation of the non-wild type PCR products

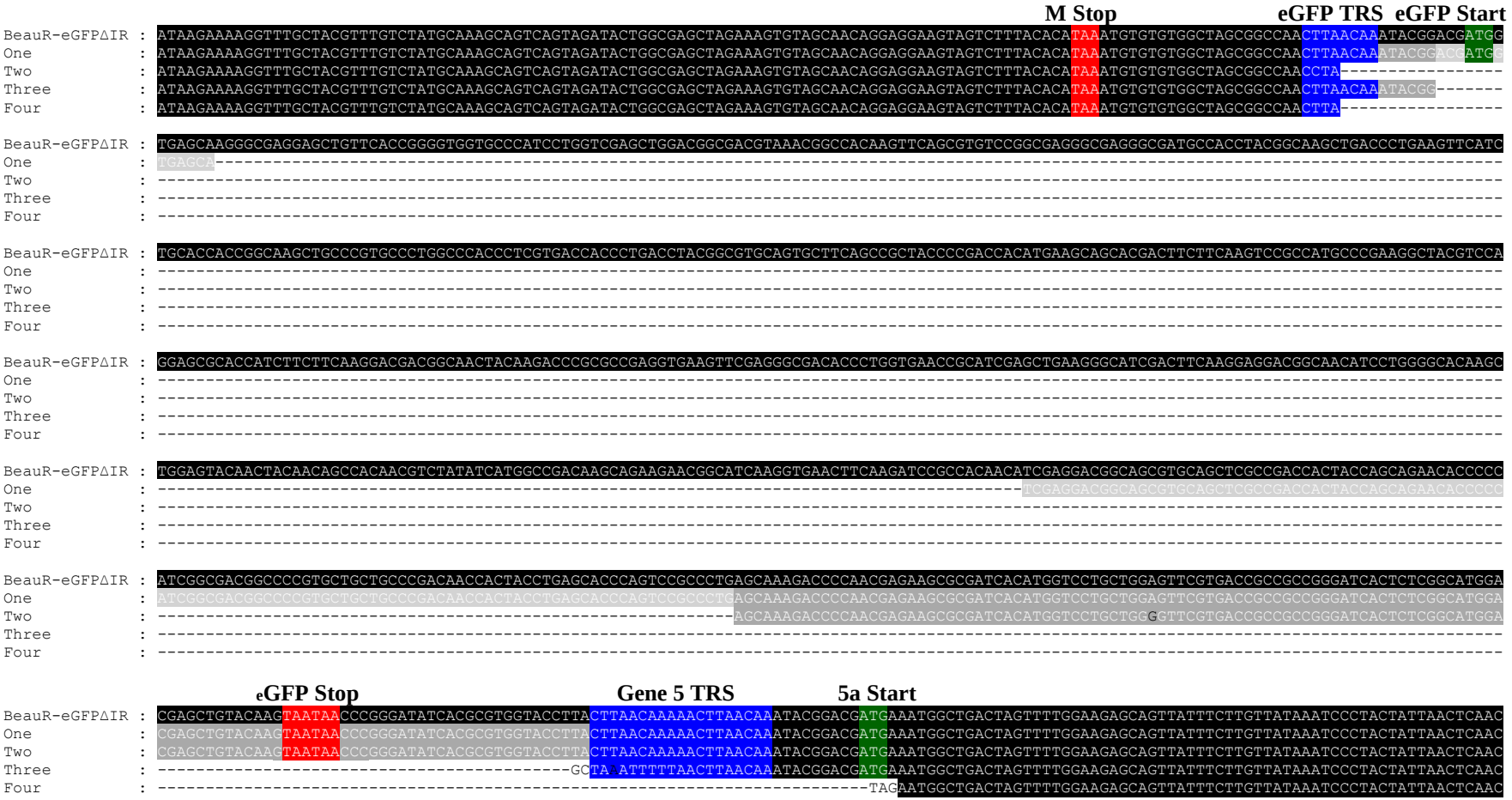
The passaging experiments had been repeated using a set MOI on the basis that the suspected deletions arising in the reporter genes were occurring in combination with other deletions across the genome, leading to the generation of D-RNAs. This theory was backed up by the fact that passaging at a set MOI did control the appearance of non-wild type PCR products, as would be expected with true D-RNAs. Previous IBV

D-RNAs were shown to have major deletions in the replicase ORFs, as well as at positions across the structural and accessory genes (Penzes, *et al.* 1994). Efforts were made to establish the nature of the non-wild type viruses and confirm the presence of D-RNAs. A number of experiments were carried out to investigate the rIBVs by northern blot analysis however; no evidence of true D-RNAs could be obtained suggesting that the rIBVs are only unstable across the region of the inserted reporter gene.

To investigate the nature of the non-wild type PCR products, and confirm the deletion of the reporter genes, the PCR products from two rIBVs, BeauR-eGFPΔIR and BeauR-hRlucΔ5ab, from P7 of the initial passaging experiment were cloned and sequenced. Alignments were then carried out against wild type sequences to show examples of which areas of the genome are deleted (Fig. 5.4).

The alignments showed that four different deletions were identified for BeauR-eGFPΔIR (Fig. 5.4A) while three different deletions were identified for BeauR-hRlucΔ5ab (Fig. 5.4B). For both rIBVs the deletions ranged in size across the reporter gene. Additionally, for BeauR-eGFPΔIR variant four also contained a deletion of the Gene 5 TRS and ORF 5a start codon thus resulting in the additional loss of both Gene 5 accessory proteins for this individual virus. This suggested that deletions were not limited to the reporter gene sequence and could spread into other non-essential genes resulting in smaller, but viable, genomes.

A


M Stop
eGFP TRS
eGFP Start
BeauR-eGFPΔIR :
One :
Two :
Three :
Four :
eGFP Stop
Gene 5 TRS
5a Start

B

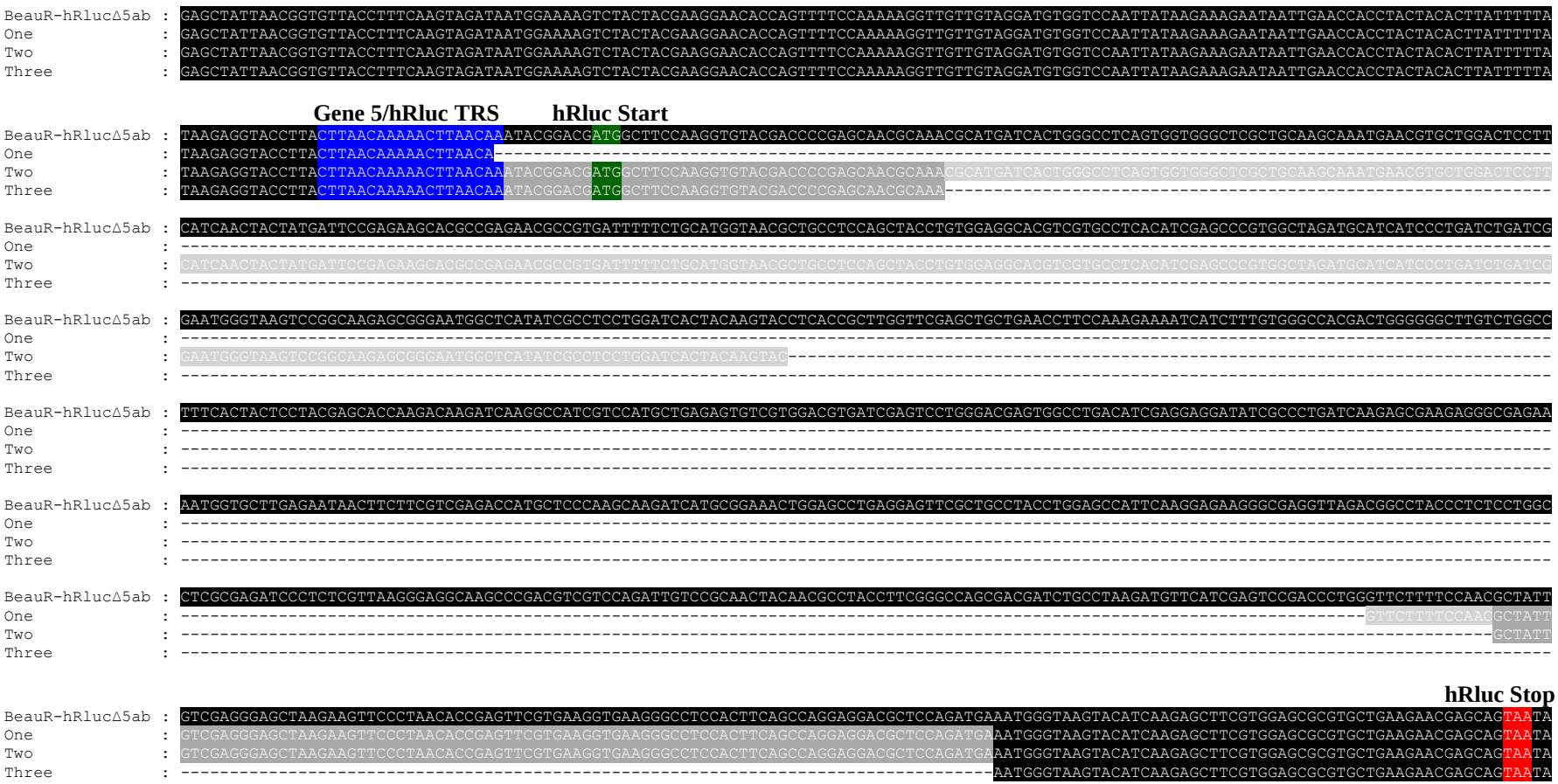

Gene 5/hRluc TRS
hRluc Start
hRluc Stop
BeauR-hRlucΔ5ab
One
Two
Three

Fig. 5.4. PCR cloning sequence analysis. PCR products from P7 of initial passaging experiment were cloned for (A) BeauR-eGFPΔIR and (B) BeauR-hRlucΔ5ab to demonstrate deletions arising in rIBVs. Wild type sequences are shown on top row with sequences from deletion variants shown below. Relevant genome positions are highlighted and labelled.

5.4 Summary

This chapter outlines the findings from investigations into the stability of rIBVs expressing reporter genes. The main findings from this chapter are summarized below:

- Initial attempts to passage rIBVs led to rapid deletion of reporter genes.
- Passaging at a set MOI improved the overall stability of each rIBV.
- Replacement of Gene 5 yields the most stable recombinants, followed by replacement of the IR and then replacement of ORFs 3a and 3b.
- Deletion variants of rIBVs are not true D-RNAs but can be controlled in a similar manner.
- Deletions are primarily located within the reporter genes but can also encompass non-essential flanking genes.

5.5 General discussion of Chapters 3-5

The aim of this project was to investigate the potential for using IBV as a vaccine vector and to identify factors that are important for generating stable virus recombinants that successfully express a heterologous protein. Chapters 3, 4 and 5 have detailed the results of this investigation and have highlighted some of the issues relevant to further pursuing the use of IBV as a vaccine vector. A series of nine rIBVs were designed to express reporter genes eGFP or hRluc, with subsequent analyses to determine how best to utilize IBV as a vaccine vector. A total of six of these viruses resulted in the successful rescue of a rIBV, as defined by the observation of IBV-induced CPE and the presence of the reporter gene by RT-PCR.

Three of the virus constructs did not result in rescued virus due to problems at different stages of the reverse genetics system.

Recombinant BeauR-IBVeGFPΔ5ab was unable to progress past the initial plasmid generation stage due to an inability to clone the genome fragment containing the codon optimized eGFP sequence in *E. coli*. Given the successful cloning of plasmid pGPT-eGFPΔ5ab, which utilizes the same genome fragment arrangement, it is not clear why pGPT-IBVeGFPΔ5ab could not be cloned, other than to suggest that the altered eGFP sequence rendered the plasmid toxic to bacterial cells. Although a number of different competent cells, and growth conditions, were tried none resulted in successful cloning of the genome fragment, therefore this potential virus could not be pursued further.

The inability to rescue rIBV BeauR-eGFPΔ3ab can be explained in context of the findings from analysis of BeauR-hRlucΔ3ab in Chapter 4. Although two replicates of BeauR-hRlucΔ3ab were rescued a number of problems were identified with these viruses, including the presence of deletions in the inserted TRS for E, and mixed populations arising around the hRluc gene in R2 (Fig. 4.9). This suggested a high degree of instability in this virus construct due to the sequence modifications made to accommodate transcription of E as a new sg mRNA from an artificially introduced TRS. Given these observations it is probable that lethal errors arose within this region of the BeauR-eGFPΔ3ab genome during initial rounds of replication, and thus a viable virus population could not be established. Without complete sequencing of the two vaccinia virus clones used for the rescue it is not possible to rule out the presence of sequence errors elsewhere in the modified IBV cDNA that could have

led to non-viable virus. However, given the stability of the vaccinia virus system, the use of two distinct plaque lineages of vaccinia virus and the successful sequencing of the modified region makes this a less likely explanation for the failure to rescue.

Recombinant virus BeauR-5a/eGFP also failed to rescue successfully. While rIBVs were in fact rescued, as identified from IBV-induced CPE, the eGFP gene had been deleted from all viruses screened by RT-PCR (Fig. 3.9). This virus had been designed so that eGFP was expressed as a C-terminal fusion of accessory protein 5a. The precise deletion of the eGFP sequence suggests that it was not tolerated in this construct. A possible explanation for this is the occurrence of recombination events that led to deletion of the eGFP sequence, as will be discussed later with regards to rIBV stability. It is possible that modifications to the sequence, i.e. an N-terminal fusion, may have yielded a successful rIBV but time constraints meant that these possibilities could not be followed up.

The analyses of the six rescued rIBVs revealed some important points for considering the use of IBV as a vaccine vector. The rescue of two replicates for each virus was intended to allow confirmation of the viability of each virus construct but instead highlighted the fact that each independent rescue can result in a different outcome in terms of errors introduced into the virus genome. Three of the rIBVs, BeauR-eGFPΔ5ab, BeauR-hRlucΔ5ab and BeauR-hRlucΔIR, had two replicates rescue without any known errors introduced into the genome. For BeauR-eGFPΔIR and BeauR-IBVhRlucΔ5ab only one replicate of each was rescued without errors, while neither replicate of BeauR-hRlucΔ3ab rescued without errors. Differences in

the successful rescue of different replicates of the same virus suggest that there is an unpredictable element in replication for some constructs.

Of the errors identified in the rescued rIBVs the one with the least impact on reporter gene expression or virus stability was that identified in BeauR-eGFPΔIR R1 (Fig. 4.4). Interestingly, this double T→A mutation, that results in a premature stop in ORF 5b, has been independently observed before in rIBVs generated in the IAH Coronavirus group (Armesto, *et al.* 2009). In this study by Armesto *et al.* the structural and accessory proteins of Beau-R were replaced with those of the pathogenic M41 strain of IBV. The resultant chimaeric virus contained several changes, one of which was the double T→A mutation in ORF 5b witnessed here. The repeated occurrence of these mutations in some rIBVs suggests that they may arise as a result of replication pressures brought about by the modifications made to the virus. As Gene 5 is known to be dispensable for replication in cell culture the loss of ORF 5b may be a trade-off for a benefit elsewhere in the virus. As suggested by Armesto *et al.* these mutations may lead to an improved recognition of the N gene TRS, and subsequent N gene expression. These mutations showed no effect on virus replication (Fig. 4.5) or virus stability (Fig. 5.2) following comparison of R1 and R2 of BeauR-eGFPΔIR, suggesting the additional loss of Gene 5 proteins does not affect this particular virus construct.

Expression of eGFP or hRluc in cell culture was confirmed by western blot for all rIBVs, with the exception of BeauR-hRlucΔ3ab R2 (Fig. 4.11). Protein expression was additionally confirmed for R1 of each rIBV by confocal microscopy (Fig. 4.12). This demonstrates that heterologous genes can be expressed from numerous

positions within the IBV genome and it is probable that genes from different avian viruses could therefore be expressed in a vaccine construct.

Analysis of heterologous protein expression from each of the Δ5ab hRluc rIBVs showed that codon optimization had a slight negative impact on protein expression (Fig 4.15). Replicate 2 of BeauR-IBVhRlucΔ5ab demonstrated lower levels of protein expression from 24 hpi compared to sister virus BeauR-hRlucΔ5ab. There are two possible explanations for the decrease in hRluc expression. First, it is possible that by maximizing the codon usage for IBV, pressure has been put on the availability of tRNAs and, as replication proceeds over time, pools of limited tRNAs are depleted and translation of the optimized gene is restricted. As peak virus titres at 24 h were identical for BeauR-hRlucΔ5ab and BeauR-IBVhRlucΔ5ab (Fig. 4.2), differences in protein expression levels cannot be explained by differences in virus replication. Additionally, at earlier time points expression levels are the same as the sister virus suggesting that codon optimization cannot increase translation of the heterologous gene above the wild type level for that gene.

A second explanation relates to the efficiency of expression of the hRluc protein in avian cells and whether this was affected by the changes introduced into the nucleotide sequence. The native *Renilla* luciferase gene was originally substantially altered at the nucleotide sequence level to generate the synthetic hRluc gene, which was optimized for expression and sensitivity in mammalian cells (Zhuang, *et al.* 2001). It is possible that sequence changes introduced during IBV codon optimization also introduced, or re-introduced, sequence elements that resulted in a reduction in the efficiency of expression of the luciferase gene in avian cells, and

counteracted the benefits of the original optimization process. However, it is also possible that these sequence changes resulted in a more stable viral RNA molecule thus explaining the small increase in virus stability observed for rIBV BeauR-IBVhRlucΔ5ab over BeauR-hRlucΔ5ab.

Interestingly R1 of BeauR-IBVhRlucΔ5ab had considerably reduced luciferase expression compared to R2 (Fig. 4.15). Alteration to the amino acid sequence of the codon optimized luciferase arose due to the insertion of two additional T nucleotides at position 26267 (Fig. 4.14). The active site for *Renilla* luciferase is proposed to consist of a triad of residues; D120, E144 and H285 (Woo, *et al.* 2008). The unintentional alteration to the hRluc sequence resulted in amino acid change H285F. This amino acid change is therefore likely to have altered the active site of hRluc with the result being a less efficient reaction of hRluc protein with the hRluc substrate and thus lower levels of luminescence detected.

Although reporter gene expression was observed for the majority of rIBVs, expression was only possible for a limited number of rounds of replication before the reporter gene was deleted from the virus. This highlights an issue of virus stability that has been seen previously with recombinant coronaviruses, and closely related arteriviruses, expressing heterologous genes (de Haan, *et al.* 2005, de Vries, *et al.* 2001, Shen, *et al.* 2009, Youn, *et al.* 2005).

Stability of the rIBVs was found to correlate with the location of the reporter gene, which therefore indirectly correlated with the level of modifications required in order to accommodate the reporter gene in the IBV genome. The most stable rIBVs, in

terms of the number of cell culture passages prior to reporter gene deletion, were those replacing Gene 5 (Table 5.2). The Δ5ab rIBVs required a direct replacement of virus sequence with the reporter gene, which could then be transcribed using the existing TRS for Gene 5. No additional sequences needed to be included to ensure continued transcription of virus genes and it is likely that this contributed to the increased stability of these rIBVs.

In contrast, the least stable construct, BeauR-hRlucΔ3ab, required not only alteration of the existing Gene 3 mRNA to incorporate the hRluc sequence in place of ORFs 3a and 3b, but also required the insertion of an additional TRS to ensure the transcription of the E gene as a new independent mRNA. Coronavirus transcription is a complex process that, although not completely understood, is likely to be tightly regulated and attempts to alter the number of mRNAs transcribed appears to be poorly tolerated. The ΔIR rIBVs were also found to be less stable than the Δ5ab rIBVs, and also had more substantial changes made to the virus sequence to accommodate transcription of the reporter gene as a new mRNA.

The finding that stability of the rIBVs could be improved by passaging at a set MOI raised the question of the mechanism of deletion of the reporter genes. Initially it was believed that reporter genes were being lost through the natural generation of D-RNAs but no evidence to support this could be obtained. Instead it appeared that only the reporter gene was deleted, as well as non-essential flanking regions in some cases (Fig. 5.4).

Sequence analysis of non-wild type PCR products from BeauR-eGFPΔIR, generated during passaging, showed that deletions were not necessarily limited to the reporter genes and could spread into flanking regions (Fig. 5.4). In the ΔIR constructs the reporter gene is flanked upstream by the essential M gene, and downstream by the non-essential Gene 5. It has been demonstrated that the loss of Gene 5 can be tolerated by the virus, raising the possibility that deletions may occur over a larger region, encompassing both the reporter gene and Gene 5, while still resulting in a viable virus population. This larger region over which deletions can occur increases the likelihood of a virus population arising that no longer contains the reporter gene and, with a smaller genome, is able to replicate more efficiently to rapidly become the dominant virus population.

In contrast, for rIBV BeauR-hRlucΔ3ab the reporter gene is flanked by two essential genes, Spike and E. Any deletions arising that encompass these genes would be lethal to the virus and consequently fewer deletion variants are likely to arise that result in a viable virus population. This may explain why in the passaging experiments detection of the wild type PCR product persists over a greater number of passages for BeauR-hRlucΔ3ab than any of the other rIBVs. As any deletions are limited to within the reporter gene, so that a viable virus population is maintained, it is likely that fewer deletion variants will arise that can replicate more efficiently than wild type. Deletion variant populations will therefore take longer to establish than for other rIBVs. As a result wild type genomes would encounter less competition for resources and packaging into new virions, and the wild type population will persist for longer.

As mentioned there have been previous studies of corona- and arteriviruses as vaccine vectors and many of these showed similar deletions of heterologous genes. One proposed explanation for this is the occurrence of homologous recombination events between similar sequences of virus genomes, such as TRSs, that result in the loss of the heterologous gene. It is probable that this is also the situation with the rIBVs analysed in this study. Recombination events occur as a result of template switching during negative-strand synthesis of virus genomes and sg mRNAs. The concurrent synthesis of negative-sense copies of both genomes (anti-genomes) and sg mRNAs (sg RNAs) means recombination events are possible between either anti-genome and anti-genome, or anti-genome and sg RNA. Although only changes introduced into the anti-genome will be hereditary, the compound effect of anti-genome recombinations with potentially defective sg RNAs will increase the chance of errors being introduced into the virus population.

This theory of recombination-based deletion could also explain why passaging at a low MOI resulted in viruses that maintained stability over a greater number of passages. By infecting at a low MOI it is less likely that a single cell is infected by more than one virus particle. This would initially limit the number of recombination events that can occur until replication has progressed sufficiently for an abundance of virus templates to be available, and so decreasing the likelihood of an early recombination event with a negative outcome.

Overall the data shows that all vaccine vector constructs are capable of expressing a heterologous gene. The stability of these viruses varies considerably depending on genome location and/or the level of modifications required to the genome. The

replacement of Gene 5 proved to be the most successful in terms of both stability and protein expression while the replacement of ORFs 3a and 3b was found to be the least stable. The type of heterologous gene inserted was found to have little effect on protein expression or stability, although eGFP resulted in poorer growth kinetics than hRluc following replacement of the IR (Fig. 4.3). Unfortunately as not all eGFP recombinants were rescued a comparison to hRluc is limited. Codon optimization may result in a slight improvement in virus stability (Table 5.2) although further passaging at the set MOI of 0.01 (past P12) would be required to observe any potential differences between BeauR-hRlucΔ5ab and BeauR-IBVhRlucΔ5ab. Codon optimized luciferase protein levels were slightly reduced compared to wild type hRluc but this may not be an unfavourable outcome depending on the heterologous gene expression levels required for an effective vaccine.

During the analysis of the rIBVs two unexpected observations were made. First, when attempting to establish protocols for the use of eGFP rIBVs in live cell imaging it was discovered that viruses lacking Gene 5 had a reduced capacity for cell-cell spread on Vero cells (Fig. 4.16). It is possible that this effect is also seen on CK cells although to a much lesser extent, as demonstrated by the plaque morphology data (Fig. 4.1) and growth kinetics (Fig. 4.2) of the Δ5ab recombinants. While the effects are much more pronounced in Vero cells than CK cells this may simply be representative of the inherently lower replication efficiency of IBV observed in Vero cells, as Beau-R has not been adapted for growth in this cell culture system. These results are the first indication of a role for the Gene 5 accessory proteins in IBV replication. Second, there was a suggestion from northern blot analysis that a mRNA may be being transcribed from the IR (Fig. 4.8). This

finding was investigated further and led to a second project for this thesis that will now be described in Chapters 6 and 7.

Chapter 6: Investigation of the IBV intergenic region

6.1 Introduction

The analysis of rIBVs in Chapter 4 led to interesting observations regarding the uncharacterized RNA species often visible on northern blots between the M gene and Gene 5 mRNAs of IBV (Fig. 4.8). The uncharacterized RNA species was first identified by Stern and Kennedy as one of a number of minor bands occasionally observed by northern blot analysis of the RNA species produced by the Beaudette strain of IBV (Stern and Kennedy 1980). It was not determined at that time whether this RNA was of cellular or viral origin and this question has remained ever since. The lack of consistent detection of this RNA is apparent throughout the literature with many publications making no reference to an additional RNA species at all, while others simply reference back to the Stern and Kennedy paper.

In Chapter 4 northern blot analysis of the Δ5ab rIBVs demonstrated that this uncharacterized RNA species shifted in size relative to the changes made to the Gene 5 mRNA (Fig. 4.8). Additionally, this RNA species was no longer detected during analysis of ΔIR rIBVs. Together, this data suggested that this RNA may in fact be a mRNA and furthermore may be linked to the region of the genome referred to as the intergenic region.

As described in more detail in Chapter 1 the IR of IBV, originally determined by sequence analysis of the Beaudette strain, is located between the M gene and Gene 5 and, in the majority of IBV strains, contains a putative ORF for a protein of 94

amino acids. The IR has also received little attention in terms of research due primarily to the lack of a TRS-B of CUUAACAA, or a similar sequence, upstream of the potential coding sequence. The presence of a TRS-B is one of the central principles underlying the mechanism of coronavirus sg mRNA transcription, and therefore despite the IR sequence being located between the M gene and Gene 5, as is also observed for the uncharacterized RNA species on northern blots, the two have never been linked. Furthermore, the theory of discontinuous transcription during negative strand synthesis is widely accepted, in part, due to the discovery that the TRS of each sg mRNA is derived from the associated TRS-B (Pasternak, *et al.* 2001, van Marle, *et al.* 1999, Zuniga, *et al.* 2004). This suggests an absolute requirement of a TRS-B for transcription and a lack thereof would mean no sg mRNA is transcribed from the IR.

As mentioned the IR contains a putative ORF, encoding a protein of 94 amino acids in the majority of IBV strains. Some, such as Beaudette and other lab-adapted strains, contain deletions within the IR ORF and therefore have lost the potential to code for this protein. The IR ORF is also known to be present in the closely related *gammacoronavirus*, turkey coronavirus (TCoV). The more common nomenclature for the TCoV IR in the literature is ORF 4b, presumably based on the assumption that without a known TRS-B any potential protein of the IR would be translated from a functionally bi-cistronic M mRNA in a manner similar to Gene 3 and Gene 5.

To date neither the IR and the putative ORF, nor the uncharacterized RNA have been studied in detail and the observation that the uncharacterized RNA may be linked to the IR provided the opportunity to fully investigate these two aspects of IBV

biology. The initial aim of this project was to determine the nature of the uncharacterized RNA species and what, if any, link exists between it and the IR. Further characterization of the RNA and the IR would be carried out in light of any findings.

6.2 Analysis of the IBV uncharacterized RNA species

The first stage of the investigation was to confirm whether a mRNA is transcribed from the IR, and if this mRNA corresponds to the previously uncharacterized RNA species seen on northern blots of IBV. To do this recombinant virus BeauRΔIR, as used in previous chapters, was subjected to northern blot analysis and the mRNA pattern compared to wild type Beau-R (Fig. 6.1).

The northern blot showed the typical mRNA profile for Beau-R in lane 1 while in lane 2 the profile for BeauRΔIR showed an altered pattern. The RNA species located between the M gene and Gene 5 (the uncharacterized RNA species) was observed to have been reduced in size following infection with BeauRΔIR. This size reduction is equivalent to the 268nt deletion introduced into the IR. The mRNAs for all genes upstream of the IR were also shown to be reduced in size as would be expected due to the nested nature of the mRNAs. This result provided strong evidence that the previously uncharacterized RNA species is in fact an IBV sg mRNA and is linked to the IR. An additional, previously unidentified RNA was also observed below the Spike gene mRNA (designated by an arrow) and is investigated further in section 6.6.

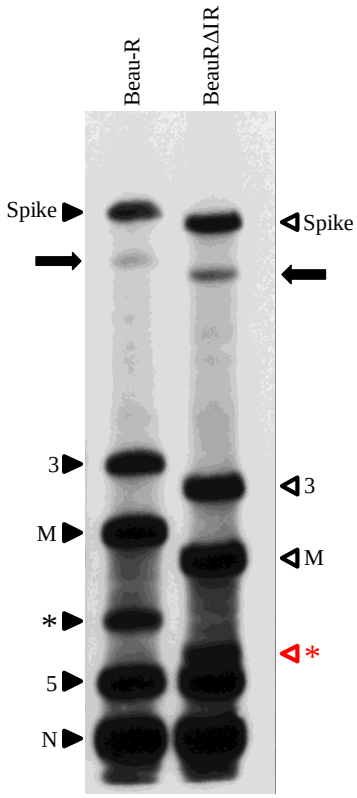


Fig. 6.1. Northern blot analysis of BeauRΔIR. Confluent monolayers of CK cells were infected with Beau-R or BeauRΔIR and 20 hpi intracellular RNA was harvested and subjected to northern blot analysis with a probe complementary to the 3'-end of the Beaudette genome. Closed arrow heads point to wild type mRNAs, open arrow heads point to altered mRNAs. The * indicates the uncharacterized RNA species. Arrows point to an additional unidentified RNA species.

As no TRS-B resembling the conserved IBV TRS, CUUAACAA, has ever been identified upstream of the IR the question of how this mRNA is transcribed was addressed. A PCR technique known as a leader-body junction PCR was used to investigate the IR further. This PCR, as detailed in section 2.9, utilizes the 5' and 3' co-terminal nature of the coronavirus sg mRNAs to amplify specific virus mRNAs as determined by the position of the reverse primer (Fig. 6.2).

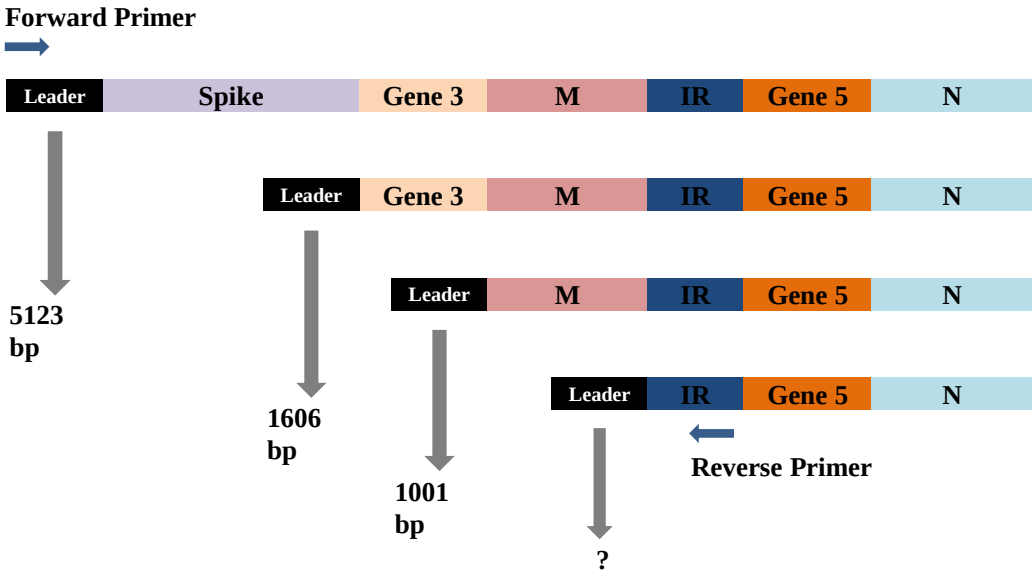


Fig. 6.2. Leader-body junction PCR schematic. This schematic outlines the amplification of coronavirus mRNAs by PCR. A forward primer located in the 5' Leader sequence is paired with a reverse primer located in one of the structural or accessory genes. Due to the nested nature of the sg mRNAs all mRNAs upstream of the reverse primer will be amplified. In the example a reverse primer located in the IR will yield PCR products of 5123bp, 1606bp and 1001bp representing Spike, Gene 3 and M mRNAs respectively. An additional PCR product would be amplified representing an IR mRNA although the size of this is unknown as the site of leader-body TRS junction is unknown.

Using this technique a PCR product suspected to correspond to the IR mRNA was amplified from Beau-R infected CK cells (Fig. 6.3).

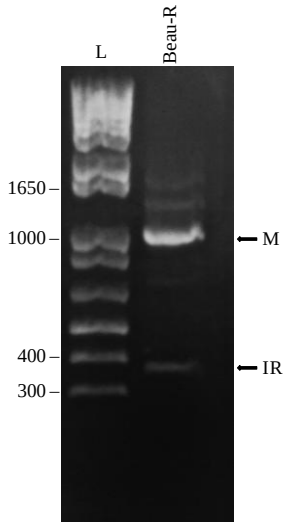
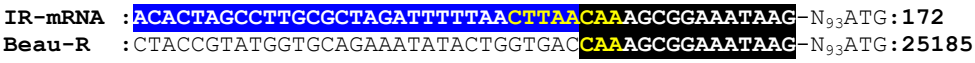


Fig. 6.3. Leader-body junction PCR of Beau-R. Intracellular RNA was extracted from Beau-R infected CK cells and reverse transcribed. Messenger RNAs were amplified by PCR with primers Leader1 and IR(R). PCR products were separated by gel electrophoresis. Arrows point to IR and M mRNAs. Lane L: DNA marker.

To facilitate further analysis of the suspected IR mRNA, PCR products were cloned and resulting bacterial colonies were screened by PCR to identify plasmids containing the original IR mRNA PCR product. Appropriate PCR products from the colony screening were then sequenced. Sequence analysis of the suspected IR mRNA yielded an interesting result (Fig. 6.4).

A



B

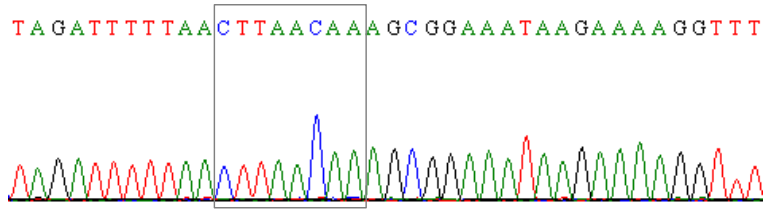


Fig. 6.4 Messenger RNA alignments. PCR products of suspected IR mRNAs were sequenced with primers M13(F) and M13(R). (A) The mRNA sequences were aligned to Beau-R genome sequence. Bottom line shows genomic sequence. Top line shows mRNA sequence with Leader sequence highlighted in blue and TRS highlighted in yellow. Relative mRNA and genome positions are given on the right along with the distance to the IR ATG. (B) Sequence trace of IR mRNA showing the CTTAACAA TRS.

The PCR product corresponding to the potential IR mRNA was found to be consistent with other IBV sg mRNAs in the sense that it contained the IBV Leader sequence at the 5'-end, followed by a full copy of the IBV TRS, CTTAACAA, followed by genomic sequence corresponding to IR. The analysis of the sequence showed that the IR mRNA aligned to a position 105nts upstream of the IR ATG, within the M gene. However, unlike for all other IBV sg mRNAs, a complete TRS-B

was not found in the genome sequence. Interestingly, the alignment of the mRNA to the genome showed that only positions 6-8 of the IBV TRS, CAA, were present as a potential TRS-B site. This suggested that the TRS-B for the IR is a non-canonical, shortened TRS-B consisting of CAA, and not CUUAACAA (or CUGAACAA) as observed for all other Beau-R structural and accessory genes.

Overall these results confirm that the RNA species present between the M gene and Gene 5 sg mRNAs on IBV northern blots fulfils the criteria for definition as an IBV sg mRNA, i.e. the 5' Leader sequence and TRS fused to genomic sequence corresponding to an IBV gene.

As the IR is known to be present in some form in all IBV strains it was necessary to confirm transcription of the IR mRNA from a variety of field and vaccine strains of IBV. These strains included Beau-CK, the parent virus of Beau-R used to generate the cDNA clone and M41, a pathogenic laboratory strain of IBV. Additionally, vaccine strains H120 and CR88, and field strains D1466, Italy-02 and QX were available in the lab and had recently been deep sequenced across the entire genome. A leader-body junction PCR was carried out on each of these strains to confirm that transcription of the IR mRNA is not unique to Beau-R. Beau-R was also repeated as a control. Gel electrophoresis showed that for each IBV strain a product was amplified relating to an IR mRNA providing strong evidence that all IBVs transcribe a sg mRNA corresponding to the IR. The size difference between Beau-R and Beau-CK, and all other strains is due to the known deletion present within Beaudette strains of IBV (Fig. 6.5).

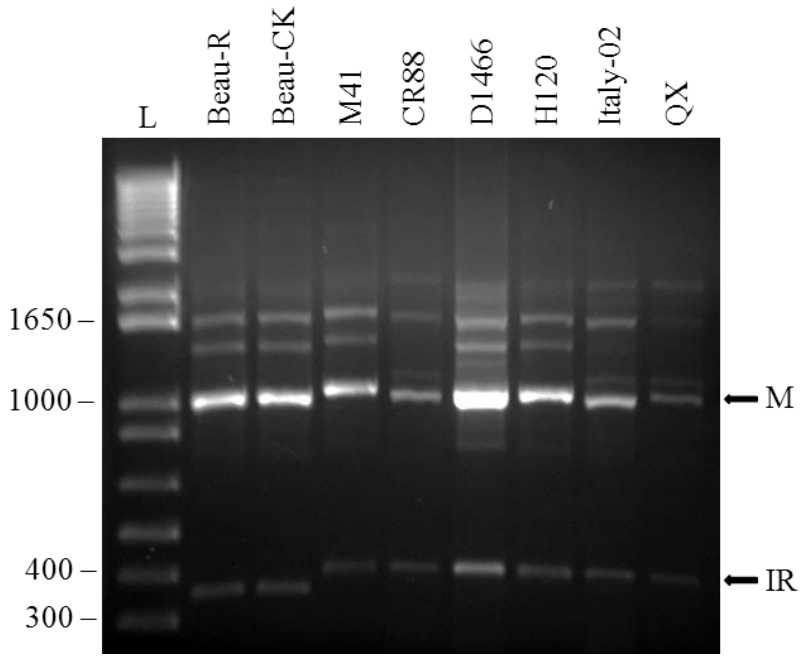


Fig. 6.5. Leader-body junction PCR of IBV field and vaccine strains. CK cells were infected with various strains of IBV and incubated for 20 to 48 h depending on strain. Intracellular RNA was extracted from cells and reverse transcribed. Messenger RNAs were amplified by PCR with primers Leader1 and IR(R). PCR products were separated by gel electrophoresis. Arrows point to IR and M mRNAs. Lane L: DNA marker.

The PCR product derived from the IR mRNA of each virus was cloned and sequenced as before. In all cases the IR mRNA sequence showed the presence of the complete CTTAACAA TRS followed by genomic sequence corresponding to the 3'-end of the M gene and the IR sequence. In all cases the alignments identified the same non-canonical TRS-B of CAA as for Beau-R (Fig. 6.6). These results confirmed that all of the IBV strains tested produce a sg mRNA corresponding to the IR sequence located between the M gene and Gene 5.

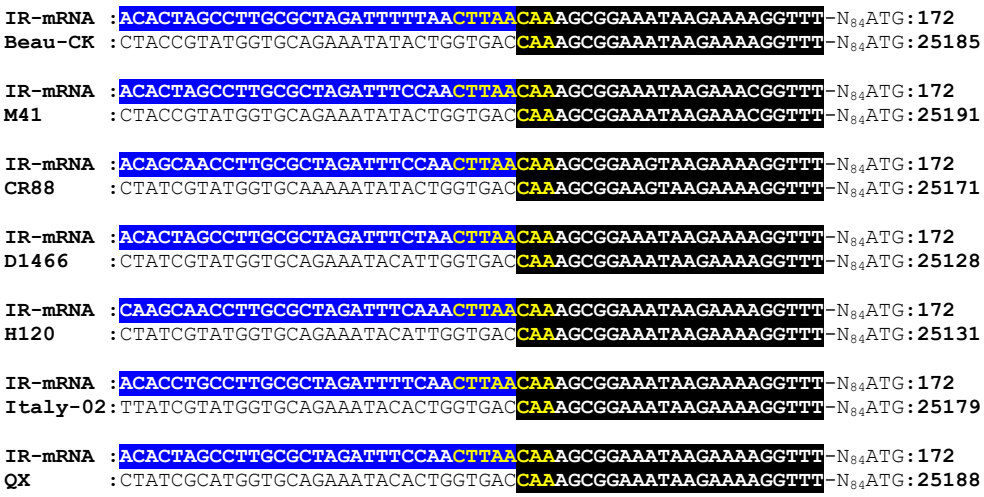
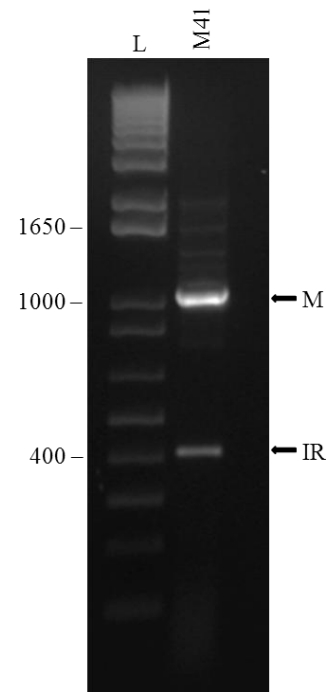


Fig. 6.6. Vaccine and field strains mRNA alignments. PCR products of each IR mRNAs were sequenced with primers M13(F) and M13(R). The mRNA sequences were aligned to the genome sequence of the relevant strain. Bottom line shows genomic sequence. Top line shows mRNA sequence with leader sequence highlighted in blue and TRS highlighted in yellow. Relative mRNA and genome positions are given on the right along with the distance to the IR ATG.

Finally, to confirm that transcription of the IR mRNA, using the suspected non-canonical TRS-B, is not an artefact of growth in CK cells, analysis was carried out on RNA samples isolated from tracheas of chickens experimentally infected with the M41 strain of IBV (Fig. 6.7).

A



B

```
IR-mRNA :ACACTAGCCTTGCGCTAGATTTCCAACTTAACAAAGCGGAAATAAGAAACGGTTT-N84ATG:172
M41     :CTACCGTATGGTGCAGAAATATACTGGTGACCAAAGCGGAAATAAGAAACGGTTT-N84ATG:25191
```

Fig. 6.7 Analysis of IR mRNA *in vivo*. Tracheal samples from IBV M41 infected birds were analysed for the presence of the IR mRNA. (A) Gel electrophoresis of leader-body junction PCR products. Lane L: DNA marker. (B) Sequence alignment of IR mRNA against M41 genome sequence. Bottom line shows genomic sequence. Top line shows mRNA sequence with leader sequence highlighted in blue and TRS highlighted in yellow. Relative mRNA and genome positions are given on the right along with the distance to the IR ATG.

The same PCR product corresponding to the IR was observed as for the M41 CK infection (Fig. 6.7A). Sequence analysis confirmed that the PCR product was a mRNA for the IR, as seen previously, and that the same non-canonical TRS-B was present (Fig. 6.7B).

The data confirms that a mRNA is transcribed from the IR of all IBV strains, both *in vitro* and *in vivo*. Additionally, transcription is suspected to be mediated by a previously unseen non-canonical TRS-B consisting of only the final three nucleotides of the known IBV TRS.

6.3 Characterization of the IR mRNA of turkey coronavirus

Given the knowledge obtained for IBV, and the known presence of the IR sequence in TCoV (ORF 4b), it was hypothesised that an independent mRNA was also being transcribed from the TCoV IR, as for IBV. There is currently no cell culture system for growing TCoV and the virus can only be grown in embryonated turkey eggs. No facilities were available at the IAH to obtain virus samples that could be screened for the presence of the IR mRNA. Instead a request was sent to Dr. Nicolas Eterradossi at ANSES in France who kindly provided a sample of total RNA extracted from homogenised turkey embryos infected with the FR080385d strain of TCoV.

The complete genome sequence was unavailable for TCoV FR080385d and so it was not possible to check that IBV primers Leader1 and IR(R), used for the leader-body junction PCR, were a sequence match to TCoV. Initial attempts at the PCR did not amplify any products suggesting a sequence mismatch in one, or both, of the

primers. To attempt to overcome this problem all of the available TCoV sequences in GenBank were analysed at the sites of the two primers to identify potential differences in sequence between TCoV and IBV. Based on the most frequent nucleotide differences two new primers were designed (Table 6.1).

TABLE 6.1: Leader-Body Junction PCR Primers

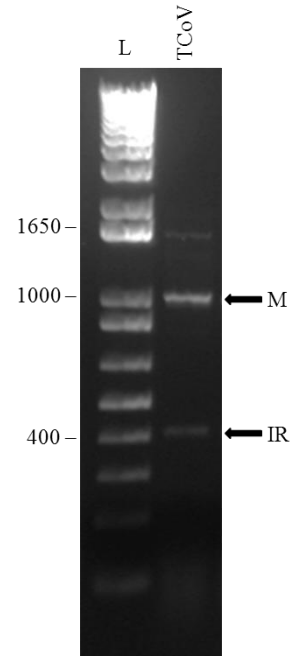
Primer	Sequence[a]
Leader1	CTATTACACTAGCCTTGCGC
TCoV Leader1	CTATCATACTAGCCTTGTGC
IR (R)	GGTTGTTGTAGGATGTGGTC
TCoV IR (R)	GGTTGTTGTAGAATGTGGTC

[a] nucleotides in red highlight changes from IBV to TCoV

A series of leader-body junction PCRs were carried out using the IBV and TCoV primers in each possible combination to find a pair that could amplify the viral mRNAs. The primer pair that was found to work was TCoV Leader1 with IBV IR(R) and mRNA products were amplified as previously (Fig. 6.8A).

A PCR product of the same size as previously obtained from IBV RNA, and representing the IR, was observed for TCoV. This PCR product was cloned and sequenced to confirm the presence of a TCoV IR mRNA. As the complete genome sequence for TCoV FR080385d was not available, in order to complete the alignments a section of the virus was sequenced using primers BG52 and IR(R). The TCoV IR mRNA sequence was then aligned to this, as well a complete TCoV genome sequence (GenBank NC_010800) for confirmation (Fig. 6.8B).

A



B

```
TCoV IR-mRNA:ATACTAGCCTTGTGCTAGATTTCTAACTTAACAAAGCGGAAATAAGAAAAGGTTT-N84ATG:
FR080385d    :CTATCGTATGGTGCAGAAATACACTGGTGACCAAAGCGGAAATAAGAAAAGGTTT-N84ATG:
NC_010800    :CTATCGTATGGTGCAGAAATACACTGGTGACCAAAGCGGAAATAAGAAAAGGTTT-N84ATG:25311
```

Fig. 6.8. Characterization of turkey coronavirus IR. Intracellular RNA was extracted from turkey embryos infected with FR080385d strain of TCoV. (A) Messenger RNAs were amplified by PCR with primers TCoV Leader1 and IR(R) and separated by gel electrophoresis. Lane L: DNA marker. (B) The TCoV IR mRNA was sequenced with primers M13(F) and M13(R). The mRNA was aligned to FR080385d genome sequence as well as GenBank sequence NC_010800. Top line shows mRNA sequence with Leader sequence highlighted in blue and TRS highlighted in yellow. Genome position was known only for GenBank strain NC_010800 and is given on the right along with the distance to the IR ATG.

The sequence alignment confirmed that a mRNA is also transcribed from within the 3'-end of the TCoV M gene and includes the potential TCoV ORF 4b protein sequence. Transcription of this mRNA also appears to be mediated by a suspected non-canonical TRS-B of CAA, as for the IR of IBV. Based on these findings it was speculated that other closely related members of the *gammacoronavirus* genus might also transcribe a mRNA from this region. Unfortunately it was not possible to test this using virus samples, however investigation of available complete genome sequences revealed that at least three further species also contain both the putative coding region and the conserved region of the non-canonical TRS-B; duck coronavirus (JF705860), a coronavirus isolated from partridge (AY646283) and pheasant coronavirus UK/602/95 (full sequence unpublished).

6.4 Investigation of the *gammacoronavirus* intergenic region protein

Sequence analyses have shown that the majority of gammacoronaviruses containing the IR contain a putative coding sequence that encodes a protein of 94 amino acids, but this has never been investigated due to the perceived lack of a mRNA for this protein. Having now identified a mRNA transcribed for the IR the question was raised as to whether it is possible that a 5th novel accessory protein is being translated by these viruses. Due to the unusual nature of the potential TRS-B for the mRNA, and the potentially altered transcription mechanism, it was first necessary to confirm that a coding region contained within the IR mRNA could be translated into a functional protein. At this time no tools were available to detect translation of the potential protein itself; therefore a rIBV expressing eGFP was generated for this purpose.

This new rIBV was similar to BeauR-eGFPΔIR used in the vaccine vector work (Chapters 3-5) and was generated from a pGPT selection plasmid synthesised by GeneArt® (see Appendix 2 and 3 for plasmid map and sequence). The construction of this rIBV involved the direct replacement of the IR/ORF 4b protein coding sequence with the sequence for eGFP, i.e. the ATG of the IR located directly downstream of the M gene stop codon became the ATG for eGFP. Unlike BeauR-eGFPΔIR, where a copy of the Gene 5 TRS-B was inserted to ensure transcription of a novel sg mRNA for eGFP, the new rIBV would utilize the suspected non-canonical TRS-B of the IR. To avoid confusion with rIBV BeauR-eGFPΔIR this new virus was called BeauR-eGFPΔ4b based on the alternative nomenclature used for the IR. BeauR-eGFPΔ4b was successfully rescued and fluorescence microscopy of CK cells infected with P3-derived virus showed high numbers of green fluorescent cells, confirming translation of eGFP (Fig. 6.9).

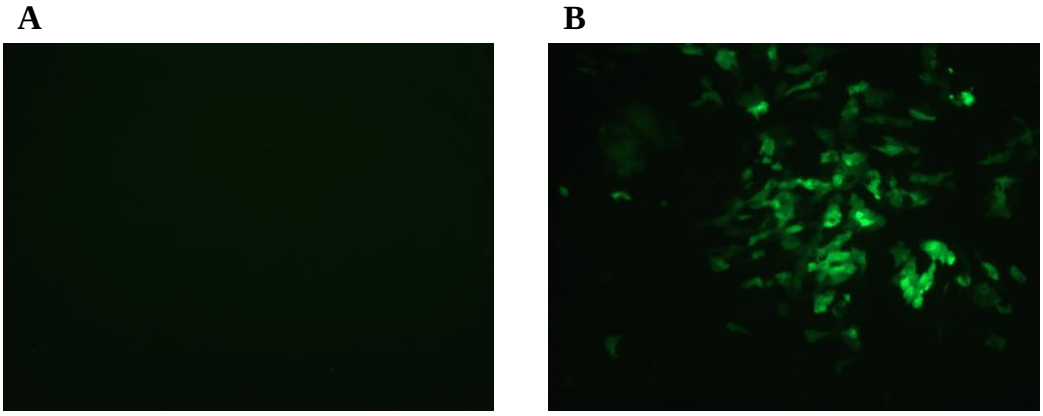


Fig. 6.9. Fluorescence microscopy of BeauR-eGFPΔ4b. Confluent monolayers of CK cells were (A) mock infected or (B) infected with BeauR-eGFPΔ4b P3 virus and examined by fluorescence microscopy 20 hpi. Images were taken at 10x magnification.

Using the leader-body junction PCR technique, the eGFP mRNA of BeauR-eGFPΔ4b was amplified with primers Leader1 and eGFP S(R) and sequenced to confirm transcription via the suspected non-canonical TRS-B (Fig. 6.10).

Fig. 6.10. Messenger RNA alignment of eGFP. A mRNA relating to eGFP from BeauR-eGFPΔ4b was sequenced and aligned to the virus genome. The M stop and eGFP start are labelled to indicate positions in the mRNA. Top line shows mRNA sequence with leader sequence highlighted in blue and TRS highlighted in yellow. Relative mRNA and genome positions are given on the right.

Sequence analysis confirmed that the eGFP mRNA was synthesised as an independent mRNA transcribed from the suspected non-canonical TRS-B located at the 3'-end of the M gene. This provided strong evidence that the potential protein coded for in the IR mRNA is translated during virus infection. To obtain more information on this potential protein some basic analysis was carried out using the coding sequence of the M41 IR. Using the Compute pI/Mw programme at www.expasy.org the theoretical isoelectric point (pI) and molecular weight were calculated at 9.06 and 11kDa respectively. Unfortunately additional analysis of the protein sequence did not reveal the presence of any specific domains or features that would provide information on the nature of the protein, and BLAST searches revealed no matches outside of the known IBV and related *gammacoronavirus*

sequences. It is therefore not possible to infer any possible functions or roles of the IR protein at this time.

With limited information available it was decided to try to detect the IR protein in virus infected CK cells. For this the generation of a peptide antibody was commissioned that could subsequently be used in different detection techniques. The antibody was generated against what was determined to be the most antigenic peptide within the M41 IR coding sequence. The peptide selected corresponded to amino acids 68-84 (5'-DNGKVYYEGKPIFQKGC-3') of the IR protein. The purified antibody was first tested by western blot of cell lysates from M41 infected CK cells. However, despite attempts to optimize the process, this did not result in visualisation of the expected 11kDa protein as was hoped. The antibody was further tested by confocal microscopy to try and visualise the protein within M41 infected CK cells however, this also proved to be unsuccessful and it remains to be confirmed that a protein is translated from the IR mRNA.

6.5 Investigation of reporter gene TRS-B usage in recombinant IBVs

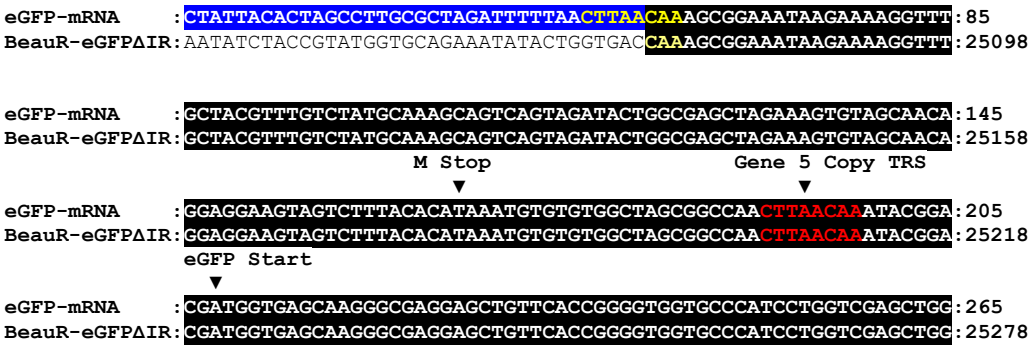
The initial observations in Chapter 4 that led to the investigation of the IR were made during the analysis of rIBVs in which the IR had been replaced with one of two reporter genes. In both rIBVs, BeauR-eGFPΔIR and BeauR-hRlucΔIR, an additional copy of the Gene 5 TRS-B was inserted upstream of the reporter gene to ensure transcription of a sg mRNA specific to the reporter gene. However, given the finding that the IR potentially has an associated TRS-B the question was raised as to how exactly the reporter genes in these rIBVs were being transcribed. Originally it

was believed that the reporter genes were transcribed via the inserted Gene 5-derived TRS-B but it was apparent from subsequent findings that this may not be the case.

Two possibilities were conceived for how the reporter genes are transcribed from BeauR-eGFPΔIR and BeauR-hRlucΔIR: (1) the polymerase recognises the inserted TRS-B and generates a mRNA with a leader-body junction at this position, (2) the polymerase ignores the inserted TRS-B and transcription continues until the natural TRS-B is reached. In this case the mRNA transcribed would be slightly larger than the one expected if the inserted TRS-B was utilized.

To confirm which of these alternative strategies is used for synthesis of the reporter gene mRNAs, the mRNAs from both BeauR-eGFPΔIR and BeauR-hRlucΔIR were amplified by PCR, cloned and sequenced using the techniques previously described (Fig. 6.11). For both reporter genes the alignments show that the inserted Gene 5-derived TRS-B is ignored by the IBV polymerase and mRNA synthesis is initiated from the natural TRS-B. This result confirms what was observed for the northern blots for these rIBVs in that the IR mRNA shifted in size according to the presence of the reporter gene sequence rather than a new mRNA being generated via the inserted TRS-B (see Chapter 4, Fig. 4.8A-D). This finding may help explain why the ΔIR rIBVs were more stable than BeaR-hRlucΔ3ab (see Chapter 5); although extensive sequence changes were made in both cases, replication of the ΔIR rIBVs ultimately did not involve synthesis of an additional mRNA that could disrupt the tightly regulated transcription process. This data showed that although consisting of only three nucleotides the proposed natural IR TRS-B is preferentially used over a complete TRS-B inserted nearby.

A



B

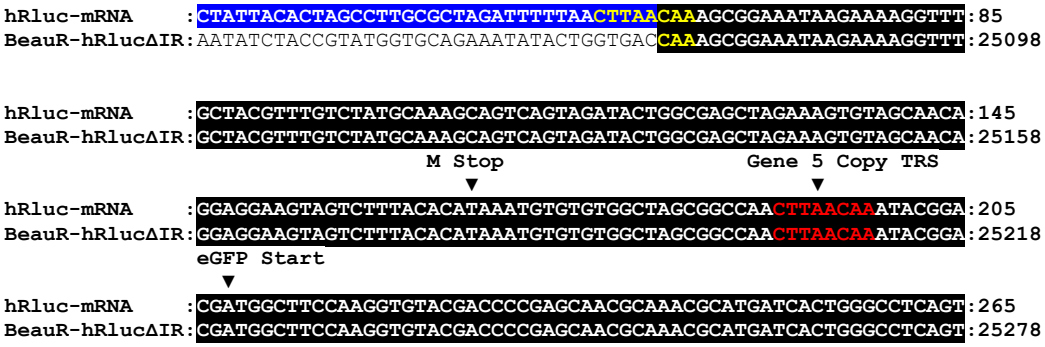


Fig. 6.11. Reporter gene mRNA alignments. Intracellular RNA was extracted from CK cells infected with BeauR-eGFPΔIR or BeauR-hRlucΔIR. Reporter gene mRNAs were amplified using primers Leader1 and (A) eGFP S(R) or (B) hRluc S(R). PCR products were cloned and sequenced with primers M13(F) and M13(R). Top line shows mRNA sequence with Leader sequence highlighted in blue and TRS highlighted in yellow. The inserted copy of a TRS-B is highlighted in red. Relative mRNA and genome positions are given on the right.

6.6 Identification of an additional mRNA also transcribed via a non-canonical TRS-B

Further analysis of the northern blot shown in Fig. 6.1 revealed the presence of an additional RNA species for Beau-R just below the Spike mRNA (indicated by an arrow). This RNA species was seen to decrease in size when produced from rIBV BeauRΔIR suggesting that this was also a potential IBV sg mRNA and, based on the

size, transcribed from within the Spike gene. Examination of a number of northern blot analyses of IBV showed that they also had a faint RNA species at this position in some but not all blots, mirroring the inconsistent detection originally seen with the IR mRNA. Using the approximate size of the RNA as a guide the genome sequence was screened upstream of the Spike S1/S2 cleavage site for a potential TRS-B but none could be identified.

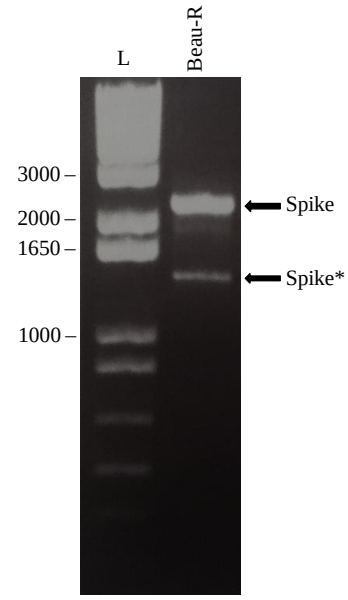
To investigate whether this was an additional mRNA for IBV transcribed via a non-canonical TRS-B the same analysis was applied as for the IR mRNA. To amplify this potential mRNA by PCR, a reverse primer (BG137) located downstream of the S1/S2 cleavage site was used along with the forward primer Leader1 (Fig. 6.12A). Two PCR products were visualised by gel electrophoresis corresponding to the Spike-derived RNAs identified on the northern blot: the Spike mRNA resulted in a 2334bp product as expected and a smaller product (Spike*) that resolved between 1000- and 1650bp. Sequence analysis of cloned PCR products confirmed the larger PCR product to be generated from the Spike mRNA, and indicated that the Spike* RNA was also a potential IBV mRNA (Fig. 6.12B, C).

The sequence alignment of the Spike* mRNA showed that this mRNA is potentially transcribed via a non-canonical TRS-B in a similar manner to the IR mRNA. In this case the potential TRS-B encompasses positions 4-8 of the canonical CTTAACAA TRS, compared to positions 6-8 for the IR mRNA. This non-canonical TRS-B is located at nucleotide position 21249-21253 of the Beau-R genome, 879nts downstream of the Spike ATG. To investigate whether this mRNA contains a protein coding sequence, the genome sequence downstream of the non-canonical TRS-B

was scanned for a potential ATG. The first ATG in the sequence is located at position 21328-21330, 74nts downstream of the non-canonical TRS-B. This distance is consistent with the gap between the TRS-B and ATG of all other IBV mRNAs and it is therefore possible that this ATG represents the start of an additional coding region. Using this ATG as a start codon it was observed that the proceeding sequence was in frame with the full Spike coding sequence and could therefore theoretically encode for a truncated version of the Spike protein, 320 amino acids shorter than the full-length protein. This truncated protein would encompass approximately the 3'–most third of the S1 subunit plus the S2 subunit.

To determine if this truncated Spike mRNA is a feature of all IBV strains attempts were made to amplify the Spike and truncated Spike mRNAs from the field and vaccine strains used to study the IR mRNA. No consistent PCR products could be obtained for any of the other strains that resembled the truncated Spike with the exception of parent virus Beau-CK, which showed the same result as Beau-R. The lack of identification of this mRNA in any other strain suggests that this may be unique to Beau-R, and the parent virus Beau-CK, and is possibly an artefact of the repeated passaging undergone by this virus during adaptation to growth on CK cells.

A



B

```
Spike-mRNA:ACACTAGCCTTGCGCTAGATTTTTAACTGAACAAAAGACAGACTTA-N40ATG:119
Beau-R    :AGCAACGCCAGTTGTTAATTTGAAAACTGAACAAAAGACAGACTTA-N40ATG:20370
```

C

```
Spike*-mRNA:ACACTAGCCTTGCGCTAGATTTTTAACTTAACAAAAACAGCTCAGA:76
Beau-R     :AGTGGTGTTCAGAATATTCAAACTTACCAAACAAAAACAGCTCAGA:21265
```

Fig. 6.12 Characterization of additional spike mRNA. Intracellular RNA was extracted from Beau-R infected CK cells and reverse transcribed. Messenger RNAs were amplified by PCR with primers Leader1 and BG137. (A) PCR products were separated by gel electrophoresis. Arrows indicate Spike and Spike* mRNAs. Lane L: DNA marker. (B, C) Clones of PCR products for Spike (B) and Spike* (C) were sequenced with primers M13(F) and M13(R) and aligned to Beau-R genome. Top line shows mRNA sequence with Leader sequence highlighted in blue and TRS highlighted in yellow. Relative mRNA and genome positions are given to the right along with (B) distance to the spike ATG.

6.7 Discussion

The work in this chapter describes the investigation of a previously uncharacterized RNA species observed during IBV infection, and the possibility that this RNA is linked to the region of the IBV genome known as the IR. The analysis of rIBV RNA species by northern blot analyses for the vaccine vector project had provided initial evidence that a previously described, but uncharacterized RNA species (Stern and Kennedy 1980) was in fact an IBV sg mRNA. This was based on the pattern of RNA species observed by northern blot analyses and the fact that the uncharacterized RNA had shifted in size relative to changes made in the recombinant viruses (Fig. 4.8).

The fact that the RNA has previously been inconsistently observed is likely due to the quality and clarity of northern blots combined with the potentially low levels of synthesis of this RNA. Many of the northern blots available for analysis used formaldehyde denaturing gels and/or radioactive isotope detection methods which can result in blots with less well defined bands and higher levels of background. The methods used during this project resulted in much clearer blots and therefore consistent detection of the RNA. Given the previous lack of consistent detection, compared to each of the other viral mRNAs, it is easy to understand why there were doubts over the nature and source of this RNA.

Utilizing rIBV BeauRΔIR it was possible to demonstrate that the RNA species is a viral mRNA and that it is associated with the IR of IBV (Fig. 6.1). The RNA species was reduced in size when produced from BeauRΔIR indicating that it was a potential

mRNA. Furthermore, this reduction correlated to the size of the deletion introduced into the IR sequence confirming that the mRNA was derived from the IR. To understand how this mRNA is transcribed, given the apparent lack of a TRS-B, a technique known as a leader-body junction PCR was applied to virus samples to amplify and analyse the IR mRNA.

Sequence analysis of the potential mRNA produced by Beau-R, and corresponding to the IR, identified the presence of the IBV 5' Leader sequence as would be expected for all IBV mRNAs thus confirming that this RNA is an IBV sg mRNA (Fig. 6.4). However, it was surprising to find that the mRNA contained a complete copy of the IBV TRS, CTTAACAA, as alignments showed that only nucleotide positions 6-8 of the TRS were present in the genomic sequence upstream of the IR. This suggested that a short, non-canonical TRS-B of CAA is utilized for synthesis of the IR sg mRNA and explains why there has previously been no identification of a TRS-B for the IR, as no genome search would be carried out for such a short sequence.

As mentioned, the presence of the full-length TRS in the IR mRNA was surprising given the non-canonical nature of the TRS-B, as there is substantial evidence demonstrating that the TRS of a mRNA is derived solely from the associated TRS-B (Pasternak, *et al.* 2001, van Marle, *et al.* 1999, Zuniga, *et al.* 2004). This observation was investigated further and forms part of the work described in Chapter 7.

The position of the proposed TRS-B in the genome, and the distance from the IR ATG, fits with what is known for the other viral mRNAs and corroborates the suggestion that this is the true TRS-B for the IR (Table 6.2).

TABLE 6.2: Distance of TRS-B from Associated Gene

Gene	Distance of TRS-B from ATG[a]
Spike	60
Gene 3	31
M	85
IR	105
Gene 5 TRS 1[b]	28
Gene 5 TRS 2[b]	17
N	101

[a]Nucleotides from position 1 of TRS-B (CTTAACAA)
[b]IBV Beaudette Gene 5 has a duplicated TRS-B

Having confirmed the existence of a sg mRNA for the IR of Beau-R it was necessary to rule out the possibility that this mRNA was an artefact of Beau-R transcription or a result of *in vitro* replication in CK cells (Fig. 6.6; Fig. 6.7). The data obtained from sequence analysis of the mRNAs from a variety of field and vaccine strains of IBV, as well as analysis of RNA samples from chickens experimentally infected with the M41 strain of IBV, confirmed that this mRNA is transcribed during IBV replication under all conditions. Furthermore, analysis of a second *gammacoronavirus*, TCoV, confirmed that this mRNA and the non-canonical TRS-B are not unique to IBV (Fig. 6.8). Analysis of genomic sequences revealed that all closely related members of the *gammacoronavirus* genus (Avian coronaviruses) contain the IR sequence between the M gene and Gene 5 and also contain the associated non-canonical TRS-B, suggesting a conserved mechanism for transcription of the IR gene.

Supporting the argument for the existence of the CAA TRS-B was the finding that rIBVs BeauR-eGFPΔIR and BeauR-hRlucΔIR both utilized the non-canonical TRS-B for transcription of the reporter genes (Fig. 6.11). In spite of the presence of a complete TRS-B, inserted upstream of the IR ATG, transcription of the reporter genes proceeded via the natural, non-canonical TRS-B. This demonstrates that despite its shortened form the natural TRS-B is preferentially used for transcription of mRNAs from this region.

The discovery of a second mRNA transcribed from the Beau-R genome via a shortened version of the IBV TRS suggests that non-canonical TRS-Bs may be an alternative method for the transcription of coronavirus mRNAs (Fig. 6.12). For the truncated Spike mRNA the TRS-B was slightly longer than for the IR, AACAA compared to CAA, but still sufficiently short that this too was not identified in genome searches, and amplification and sequence analysis of the mRNA was required to establish the nature of the TRS-B. Unlike for the IR, this mRNA could not be isolated from any other IBV strains and for this reason no attempts were made to isolate the mRNA from *in vivo* or TCoV samples. While it appears unlikely that a functional protein is translated from the truncated Spike mRNA it demonstrates the possibility that some coronaviruses may utilize similar shortened versions of a TRS-B to transcribe additional mRNAs that have yet to be identified.

Following confirmation of the presence of an IR mRNA it was speculated that the 11kDa protein coded for within this mRNA would be translated. While there is no evidence to suggest that there would be an impact on the ability of proteins to be translated from a mRNA derived via a non-canonical TRS-B, the lack of information

regarding the IR meant that this needed to be investigated. Recombinant IBV BeauR-eGFPΔ4b was generated to show that functional protein, in the form of eGFP, could be translated from a mRNA transcribed via the non-canonical TRS-B (Fig. 6.9).

This finding supported the idea that the 11kDa IR protein is translated during virus replication. In an attempt to identify this protein a peptide antibody was generated against the IR protein sequence. As Beau-R contains a 50nt deletion within the IR, disrupting the protein coding sequence, laboratory strain M41 was used for this work. All experiments attempting to visualise the IR protein were unsuccessful, but this was believed to be a failure of the antibody rather than indicating a lack of the protein.

Given the data outlined in this chapter it is clear that a mRNA is transcribed from the IR of IBV and that this is achieved via a previously unseen non-canonical TRS-B. In addition, although the evidence of its existence is yet to be obtained it is proposed that this mRNA codes for an 11kDa protein that constitutes a 5th novel accessory protein of IBV and closely related *gammacoronaviruses*.

Chapter 7: Investigation of the non-canonical IR TRS-B and the role of the TRS in regulating subgenomic mRNA synthesis

7.1 Introduction

In Chapter 6 a series of experiments were described that led to the identification of a novel sg mRNA of IBV. Work presented in Chapter 6 showed that this mRNA is transcribed from within the 3'-end of the M gene of all strains of IBV, and closely related gammacoronaviruses, via a suspected non-canonical TRS-B consisting of only the final three nucleotides of the consensus IBV TRS of CUUAACAA. Sequence analysis of the IR mRNA showed that it contained a complete copy of the consensus IBV TRS. However, the origin of this TRS was unknown as the lack of a consensus TRS-B meant that the current model of coronavirus sg mRNA transcription, in which the TRS of a mRNA is derived from the TRS-B of the associated gene, was not appropriate.

This chapter details the experiments carried out, utilizing IBV reverse genetics technology, to investigate the IR mRNA further and determine the origin of the TRS within the IR mRNA. This investigation also aimed to understand the role that a non-canonical TRS-B may play in transcription of the IR mRNA and how that fits with the current model of coronavirus transcription.

7.2 Generation of recombinant IBVs to study the IR mRNA

Sequence analysis of each of the IR mRNAs for IBV and TCoV in Chapter 6 was highly suggestive of a TRS-B for the IBV IR consisting of only the final three nucleotides of the consensus IBV/TCoV TRS of CUUAACAA. However, the IR mRNA did contain a complete copy of the CUUAACAA TRS suggesting that for the IR mRNA TRS positions 1-5 are derived from an alternative source to the TRS-B. The current model of coronavirus transcription, in which the discontinuous step occurs during negative-strand synthesis, implies that the TRS of each mRNA is derived from the TRS-B. However, Pasternak *et al.* had previously shown in a study of the related arterivirus EAV that mutations in the TRS could result in TRS-L-derived nucleotides being present in the mRNA TRS (Pasternak, *et al.* 2001) and thus raising the possibility of mRNA TRSs derived partly from the TRS-L.

In recent years a greater understanding of the process of coronavirus sg mRNA transcription has been sought with evidence strongly suggesting that, along with recognition of the TRS-B, elements such as RNA structures and/or protein interactions are also involved (Dufour, *et al.* 2011, Ozdarendeli, *et al.* 2001, Yang, *et al.* 2011). With this in mind it was hypothesised that in the case of the IR mRNA these additional factors may play a greater role in initiating the template switch of the RTC than recognition of/signalling from the TRS-B, and transcription of a sg mRNA for the IR is possible despite the lack of a consensus TRS-B. This hypothesis dictates that the non-canonical TRS-B acts in a similar manner to all other TRS-Bs by mediating hybridisation of the nascent negative-sense RNA to the 5' Leader sequence at positions 6-8 only of the TRS-L. The majority of the IR mRNA TRS

would subsequently be derived from transcription of the TRS-L, contrary to the current model of coronavirus transcription, but in keeping with the findings of Pasternak *et al.*(Pasternak, *et al*. 2001).

The availability of the IBV reverse genetics system opened the possibility of investigating this theory and for identifying the origin of the IR mRNA TRS. For example, the introduction of point mutations into the TRS-L of Beau-R, and subsequent analysis of the IR mRNA, would allow the source of the IR mRNA TRS to be established. To investigate this it was necessary to construct two rIBVs: (1) to identify the source of positions 1-5 of the mRNA TRS and (2) to confirm positions 6-8 of the mRNA TRS are derived from the non-canonical TRS-B.

In order to generate the two rIBVs with modified TRS-Ls a genome fragment had to be produced that contained the TRS-L with approximately 400bp of virus sequence either side to allow for recombination events. Due to the proximity of the TRS-L to the 5'-end of the IBV genome this meant that the genome fragment would also need to encompass vaccinia virus sequence. A suitable region of vNotI/IBV$_{FL}$ was amplified by PCR using a forward primer within the vaccinia virus sequence: (5'-TTAATTAACACACCGATTGATACATATC-3') and a reverse primer within the IBV cDNA virus sequence: (5'-AAGCTTTCACTAAACACCACCAGAAC-3'). To facilitate cloning of the genome fragment into plasmid pGPTNEB193 restrictions sites *Pac*I and *Hin*dIII were added to the 5'- and 3'-end respectively during PCR. The genome fragment was ligated into *Pac*I/*Hin*dIII digested pGPTNEB193 to give plasmid pGPT-Vaccinia/IBV (Fig. 7.1).

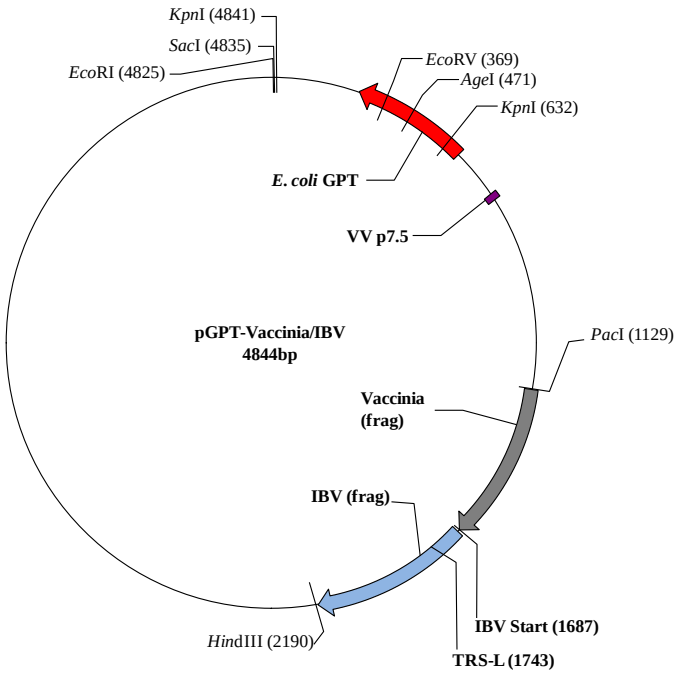


Fig. 7.1. Plasmid map of pGPT-Vaccinia/IBV. The location of the genome fragment of vaccinia virus and IBV sequence is shown in the pGPTNEB193 plasmid. The IBV TRS-L is shown approximately central to the virus cDNA fragment.

Due to the importance of the TRS-L in transcription of all IBV genes it was necessary to ensure that any mutations introduced into the TRS-L would not be detrimental to the virus. For this reason in the first rIBV a single T→G point mutation was introduced at position 3 of the TRS-L to mutate the TRS-L from CTTAACAA to CTGAACAA, and yield rIBV BeauR-CTGAACAA. This mutation was chosen because the TRS-B for both Spike and Gene 3 of Beau-R is CTGAACAA and so it was already known that a T/G mismatch between the TRS-B and TRS-L at this position could be tolerated during transcription.

The second of the TRS-L rIBVs was required to confirm that the proposed TRS-B is the source of positions 6-8 of the IR mRNA. A mutation at TRS-L position 8 was hypothesised to be less disruptive to transcription than a mutation at position 6 or 7

and so recombinant virus BeauR-L-CTTAACAT was generated with an A→T mutation at position 8 of the TRS-L. As there was no evidence to suggest which alternative base at position 8 would be least disruptive to virus replication this mutation was chosen to maintain the AU-rich nature of the TRS whilst also maintaining the free energy of the wild type A-U base pairing with that of a U-U base pairing.

To create rIBVs BeauR-L-CTGAACAA and BeauR-L-CTTAACAT the T→G and A→T point mutations were individually introduced into the IBV sequence of pGPT-Vaccinia/IBV by site directed PCR mutagenesis using primers given in Table A1.3, and reaction parameters as described in section 2.9. PCR products were *Dpn*I digested for 3 h at 37°C and concentrated by processing through a nucleotide removal column. The resulting plasmid DNA was used to transform TOP10 cells and colonies were screened by PCR and sequence analysis to identify plasmids containing the required mutations. Plasmid DNA stocks of pGPT-BeauR-L-CTGAACAA and pGPT-BeauR-L-CTTAACAT were obtained by maxi-prep of appropriate cultures and subsequently used for reverse genetics as described in section 2.10. Two replicates each of BeauR-L-CTGAACAA and BeauR-L-CTTAACAT were successfully rescued as determined by sequence analysis of the TRS-L region of the genome (Fig. 7.2).

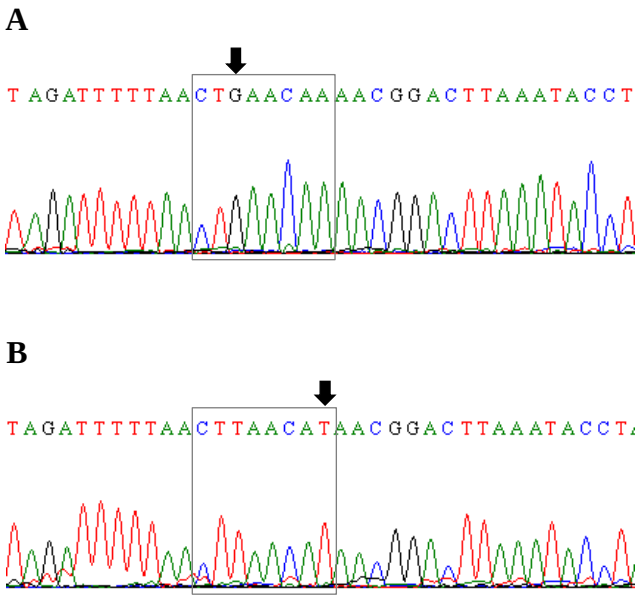
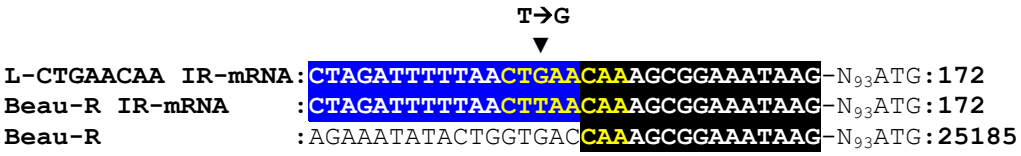


Fig. 7.2. Genome sequence analysis of TRS-L mutants. CK cells were infected with passage 4-derived (A) BeauR-L-CTGAACAA or (B) BeauR-L-CTTAACAT. Extracellular virus was harvested at 20 hpi. TRS-L region was amplified by RT-PCR with primers Leader1 and BG2 and sequenced with primer BG2. Sequence traces show the modified TRS-L outlined with grey boxes and mutations indicated by black arrows.

7.3 Analysis of IR mRNAs of BeauR-L-CTGAACAA and BeauR-L-CTTAACAT

Following the successful rescue of rIBVs BeauR-L-CTGAACAA and BeauR-L-CTTAACAT the IR mRNAs of these viruses could be analysed for presence or absence of the mutations. Working with the hypothesis that positions 1-5 of the IR mRNA TRS are transcribed from the TRS-L the IR mRNA of rIBV BeauR-L-CTGAACAA was analysed to determine whether this mutation was incorporated into the mRNA sequence. The IR mRNA was obtained by leader-junction PCR, as described in section 2.9 and Fig. 6.2, and subjected to sequence analysis. The IR mRNA sequence was aligned to the virus genome as well as an IR mRNA from a control infection with Beau-R (Fig. 7.3).

A



B

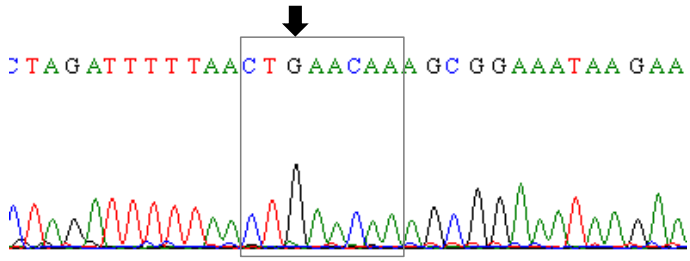


Fig. 7.3. Characterization of BeauR-L-CTGAACAA IR mRNA. CK cells were infected with either Beau-R or rIBV BeauR-L-CTGAACAA and 20 hpi intracellular RNA was harvested. IR mRNA was amplified with primers Leader1 and IR(R). (A) IR mRNA alignments. Leader sequence highlighted in blue and TRS highlighted in yellow. (▼) indicates position of the point mutation. Relative mRNA and genome positions are given on the right along with distance to the IR ATG. (B) Sequence trace of BeauR-L-CTGAACAA IR mRNA. Grey box outlines the TRS with position of TRS-L point mutation indicated by a black arrow.

The IR mRNA of BeauR-L-CTGAACAA was shown to contain the T→G point mutation that had been introduced into the TRS-L thus confirming the hypothesis that the IR mRNA TRS is partly transcribed from the TRS-L.

To confirm the source of positions 6-8 of the mRNA TRS as the non-canonical TRS-B the second rIBV, BeauR-L-CTTAACAT was analysed. The IR mRNA was subjected to sequence analysis to look for incorporation of the A→T mutation in the mRNA TRS (Fig. 7.4).

A

B

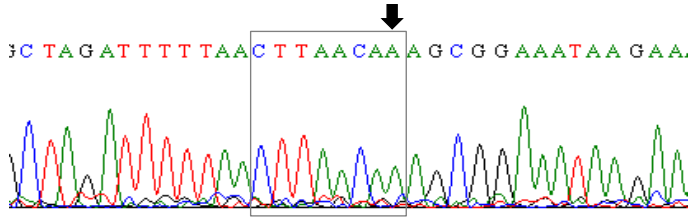


Fig. 7.4. Characterization of BeauR-L-CTTAACAT IR mRNA. CK cells were infected with either Beau-R or rIBV BeauR-L-CTTAACAT and 20 hpi intracellular RNA was harvested. IR mRNA was amplified with primers Leader1 and IR(R). (A) IR mRNA alignments. Leader sequence highlighted in blue and TRS highlighted in yellow. (▼) indicates position of the point mutation. Relative mRNA and genome positions are given on the right along with the distance to the IR ATG. (B) Sequence trace of BeauR-L-CTTAACAT IR mRNA. Grey box outlines the TRS with position of TRS-L point mutation indicated by a black arrow.

Sequence analysis showed that the A→T mutation was not present in the IR mRNA of BeauR-L-CTTAACAT thus confirming that the 3'-end of the IR mRNA TRS is derived from the non-canonical TRS-B. The data obtained from analysis of these two rIBVs demonstrates that although part of the IR mRNA TRS is derived from a non-canonical TRS-B, in line with the current model of coronavirus transcription, the IR is unique amongst currently known, naturally occurring, coronavirus mRNAs in that the majority of the mRNA TRS is derived from the TRS-L.

7.4 Generation of a rIBV to investigate the role of the IR TRS-B

The function of the TRS-B in coronavirus transcription is believed to be as a signal for the pausing/stopping of the replication complex during negative-sense RNA

transcription, and facilitating hybridisation of the nascent RNA to the TRS-L for addition of a negative-sense copy of the genomic 5' Leader sequence to the 3'-end of the nascent RNA. The finding that the TRS-B for the IR mRNA consists only of the sequence CAA brings into question the role of this TRS-B in acting as a pause/stop signal for the replication complex as this triplicate nucleotide sequence occurs >250 times throughout the genome without resulting in production of a mRNA. This suggests that the primary role of the TRS-B is to facilitate hybridisation of the nascent negative-sense RNA to the TRS-L to ensure the addition of a negative-sense copy of the 5' Leader sequence to the nascent RNA. As sg mRNAs can only be synthesised by copying of the completed negative-sense RNAs, the number of negative-sense RNAs synthesised based on the TRS-B/TRS-L interaction will subsequently determine the level of sg mRNAs observed.

Previous work has shown that the introduction of mismatches between the TRS-B and TRS-L can alter the mRNA level of the associated gene (Yount, *et al.* 2006, Zuniga, *et al.* 2004). Historically, northern blot analysis of IBV sg mRNAs has demonstrated that the IR mRNA is not always detected, or if it is, is detected in much lower amounts than the mRNAs representing the flanking genes at the same time point of infection. Having established that the TRS-B for the IR consists of only 3 nucleotides it was hypothesised that this shortened sequence may affect the efficiency of the hybridisation event between the TRS-B and TRS-L therefore resulting in lower levels of this mRNA. Consequently, it was further hypothesised that if a complete IBV TRS-B for the IR mRNA were introduced into the Beau-R genome the levels of the IR mRNA would increase due to a 100% match between the TRS-L and TRS-B.

To investigate this hypothesis a third rIBV was generated in which the TRS-B of the IR was extended. Initially the genome sequence was analysed to identify what changes could be made to accommodate a consensus IBV TRS-B. The non-canonical TRS-B of the Beau-R IR is located at nucleotides 25075-25077 within the M gene coding region. To generate a complete TRS-B at nucleotides 25070-25077 it was not possible to use the consensus CTTAACAA TRS as this would have introduced a premature stop codon into the M gene at amino acid 190. Instead a TRS-B of CTGAACAA, as found for the Beau-R Spike gene and Gene 3, was adopted. This required two modifications to the genome sequence, a G→C mutation at 25070 and a C→A mutation at 25074, resulting in amino acid changes G189A and D190E within the M gene (Fig. 7.5).

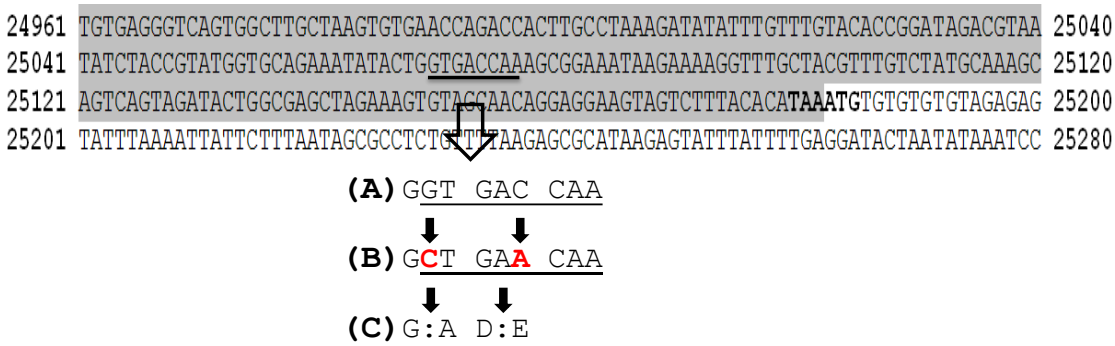


Fig. 7.5. Sequence modifications for replacement of IR TRS-B. Section of Beau-R genome sequence showing M highlighted in grey and IR in white, with the M stop and IR ATG shown in bold. (A) The sequence of the proposed TRS-B site is enlarged and underlined. (B) Point mutations required for a complete TRS-B of CTGAACAA are highlighted in red. (C) Resulting amino acid changes within M.

Initial attempts at site directed mutagenesis to generate a modified version of pGPTNEB193 containing a genome fragment with both point mutations proved unsuccessful, with large deletions occurring within the IBV sequence following introduction of the second mutation. The modified genome fragment sequence was

therefore sent to GeneArt® for synthesis and cloning into pGPTNEB193 to yield plasmid pGPT-BeauR-IR-CTGAACAA (Fig. 7.6).

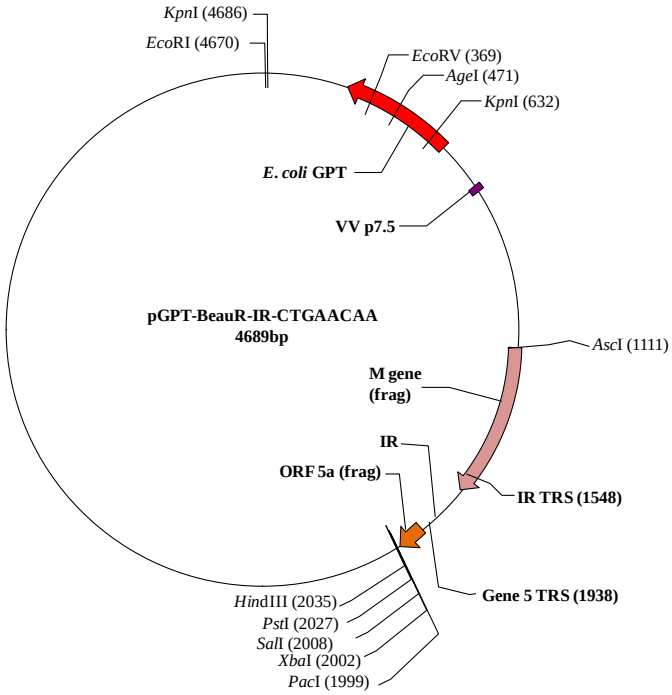


Fig. 7.6. Plasmid map of pGPT-BeauR-IR-CTGAACAAA. The location of the genome fragment containing the IR TRS-B is shown in the pGPTNEB193 plasmid. The IR TRS-B is shown approximately central to the virus cDNA fragment.

Two replicates of rIBV BeauR-IR-CTGAACAA were subsequently rescued using the reverse genetics system for IBV with the presence of the two mutations confirmed in both replicates by sequence analysis (Fig. 7.7). However, the sequence data also identified additional mutations arising independently upstream of the altered sequence. These mutations were not present in the IBV cDNA sequence in either vaccinia virus suggesting they had arisen during passage of the rescued virus on CK cells.

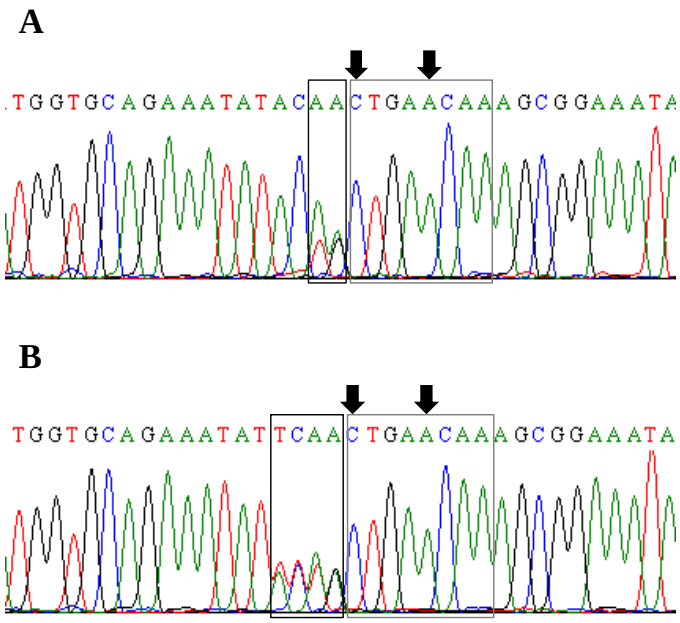
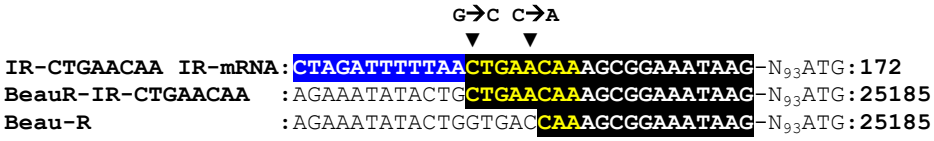


Fig. 7.7. Genome sequence analysis of BeauR-IR-CTGAACAA. CK cells were infected with passage 4 derived BeauR-IR-CTGAACAA. Extracellular virus was harvest 20 hpi. The IR TRS-B region was amplified by RT-PCR, and sequenced with primers BG51 and IR(R). Sequence traces show the modified TRS-B of (A) R1 and (B) R2 outlined with grey boxes and engineered mutations indicated by black arrows. Black boxes outline regions of additional mutations.

Given the amino acid changes introduced into the M gene sequence it was suspected that the mutations that arose upstream of the IR TRS-B were compensatory mutations. However, as the engineered mutations had been maintained it was decided to proceed with analysis of this virus despite the additional changes developing in the virus sequence. It was first necessary to analyse the IR mRNA of BeauR-IR-CTGAACAA to determine the origin of the IR mRNA TRS as two possibilities were conceived of: (1) the IR mRNA TRS is derived entirely from the engineered complete TRS-B as predicted by the model of coronavirus transcription or (2) the IR mRNA TRS is still partly derived from the TRS-L via a modified version of the transcription model.

Sequence analysis of the IR mRNA for BeauR-IR-CTGAACAA showed the presence of a TRS of CTGAACAA, matching the sequence of the extended TRS-B and differing from the TRS of CTTAACAA observed for Beau-R (Fig. 7.8). This suggested that the IR mRNA TRS was derived from the entire TRS-B and no longer involved transcription of the TRS-L. However, while the G at position 3, and therefore the downstream C→A mutation at position 5 must now be derived from the TRS-B, and not the TRS-L, it cannot be ruled out that positions 1 and 2 are still derived from the TRS-L, as the now perfect match of these two nucleotides between the TRS-B and TRS-L means it is not possible to determine their source from sequence analysis. This finding suggested that the wild type IR mRNA TRS is partly derived from the TRS-L only in the absence of a complete TRS-B, and not because of any major differences in the mechanism of transcription for this mRNA.

A



B

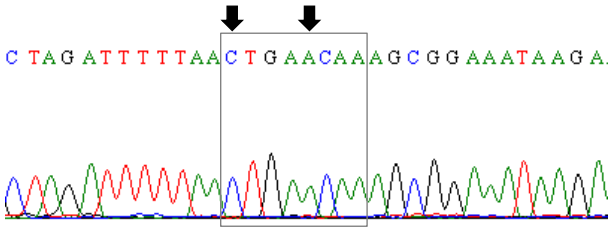


Fig. 7.8. Characterization of BeauR-IR-CTGAACAA IR mRNA. CK cells were infected with either Beau-R or rIBV BeauR-IR-CTGAACAA and 20 hpi intracellular RNA was harvested. (A) IR mRNA alignment. Leader sequence highlighted in blue and TRS highlighted in yellow. (▼) indicates positions of the point mutations. Relative mRNA and genome positions are given on the right. (B) Sequence trace of BeauR-IR-CTGAACAA IR mRNA. Grey box outlines the TRS-B and point mutations are indicated by black arrows.

Prior to using each of the TRS mutants to investigate of the role of the TRS in transcription it was necessary to confirm that the mutations introduced in the three rIBVs had not otherwise altered the replication kinetics of the virus. Growth curves were carried out for R1 of each rIBV to determine any effects of the introduced mutations on virus growth (Fig. 7.9). The growth curves demonstrated that replication kinetics were not altered for either of the TRS-L mutants, BeauR-L-CTGAACAA or BeauR-L-CTTAACAT. However, BeauR-IR-CTGAACAA showed peak titres at 24 h approximately 1-$\log_{10}$ lower than Beau-R, or the TRS-L mutants. This was suspected to be due to the amino acid changes introduced into the M gene resulting in a mildly impaired M protein.

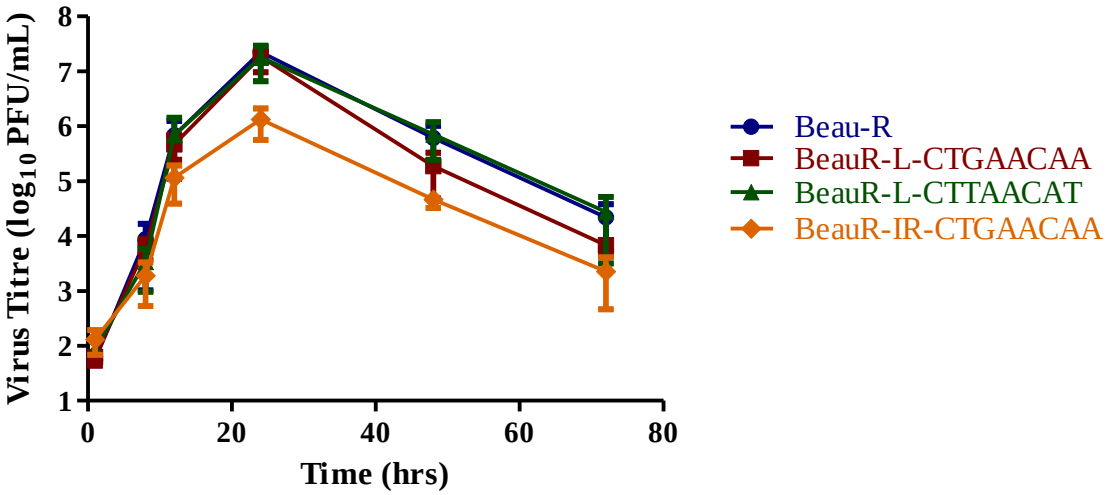


Fig. 7.9. Replication kinetics of TRS mutants. Confluent monolayers of CK cells were infected with 1x10^5 pfu virus and extracellular virus harvested at various time points for plaque assay. Error bars represent SD of results from three independent experiments each assayed in triplicate.

7.5 Investigating the role of the TRS in regulating mRNA levels

As a consequence of the non-canonical nature of the IR TRS-B the IR mRNA provided a tool for studying the effects of altering the IBV TRS-B/TRS-L interaction on mRNA levels. Following the generation of the three rIBVs in this project, BeauR-L-CTGAACAA, BeauR-L-CTTAACAT and BeauR-IR-CTGAACAA, it was hypothesised that the point mutations may alter the IR mRNA levels, when compared to the expression levels of the IR mRNA of parental virus Beau-R, due to changes in the number of nucleotide matches between the TRS-L and the IR TRS-B.

The expression levels of the IR mRNA for each of the TRS mutant viruses were compared to Beau-R using northern blot analysis to visualise any changes resulting from the TRS mutations (Fig. 7.10). As can be seen in Figure 7.10 it was clear from this analysis that the nucleotides comprising the TRS have an important role to play in regulating mRNA levels. No change in expression level was observed in any mRNA between Beau-R and Beau-R-L-CTGAACAA (Fig. 7.10A). This was expected as no changes in replication kinetics were observed for this virus and the mutation introduced into the TRS-L is not involved in the non-canonical IR TRS-B/TRS-L interaction. Analysis of BeauR-L-CTTAACAT however revealed that while no other mRNA levels were altered the IR mRNA was now almost undetectable (Fig. 7.10B). In contrast, the introduction of the complete IR TRS-B in BeauR-IR-CTGAACAA resulted in an increase in IR mRNA levels comparable to the levels of the flanking genes, M and Gene 5 (Fig. 7.10C). The northern blot analysis data was strongly suggestive that the TRS-B/TRS-L interaction is important for regulating the amount of mRNA transcribed from a particular gene.

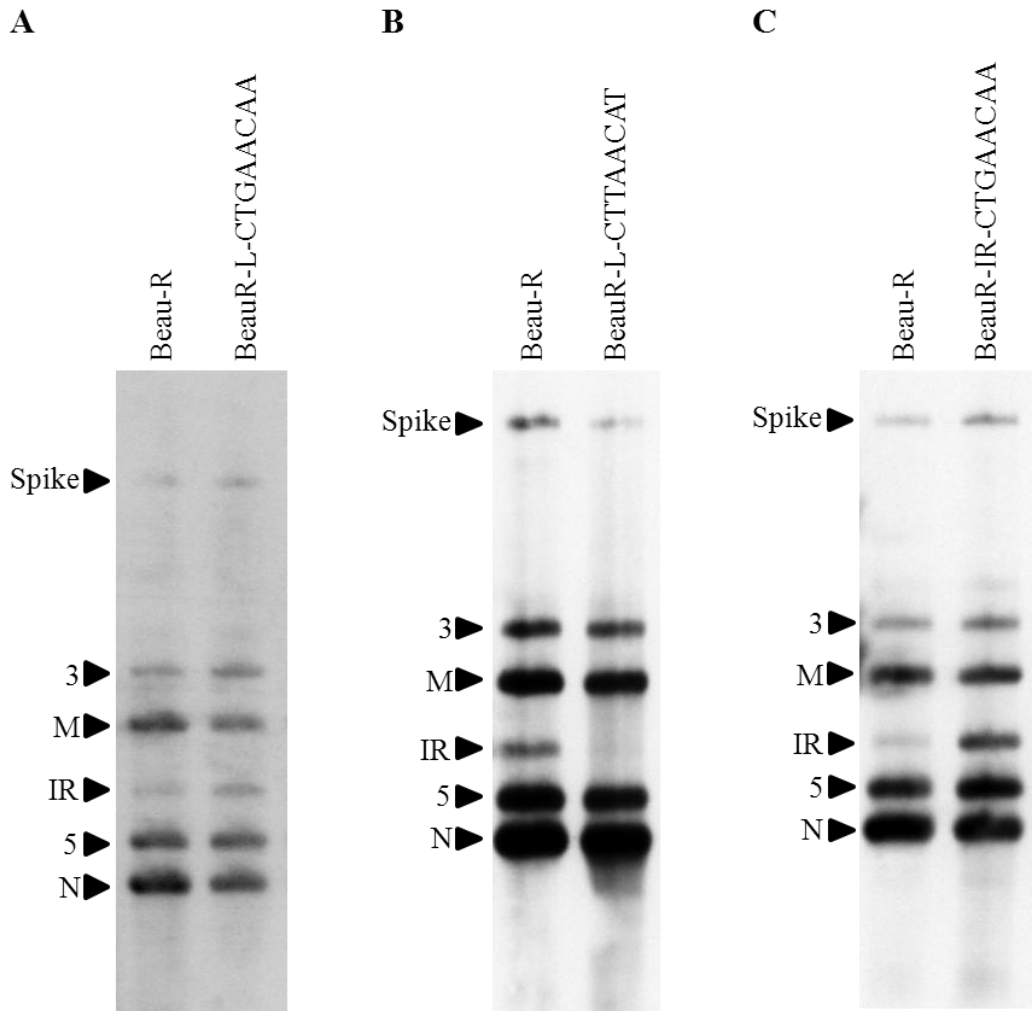


Fig. 7.10, Northern blot analysis of TRS mutants. CK cells were infected with Beau-R and rIBVs, intracellular RNA harvested 20 hpi and subjected to northern blot analysis. Each recombinant was compared to Beau-R. Data is shown for one replicate only of rIBV (A) BeauR-L-CTGAACAA, (B) BeauR-L-CTTAACAT and (C) BeauR-IR-CTGAACAA.

Although the changes in IR mRNA levels were visible by northern blot the techniques used did not allow for direct quantification of each mRNA. In order to obtain quantitative data that could be used to confirm the changes in mRNA levels the IR mRNA levels were additionally analysed by 2-step qRT-PCR. The IR mRNA levels were quantified for each of the rIBVs and expressed as a fold-change in amount compared to Beau-R (Fig. 7.11) (see Appendix 5 for sample standard curves and data calculations).

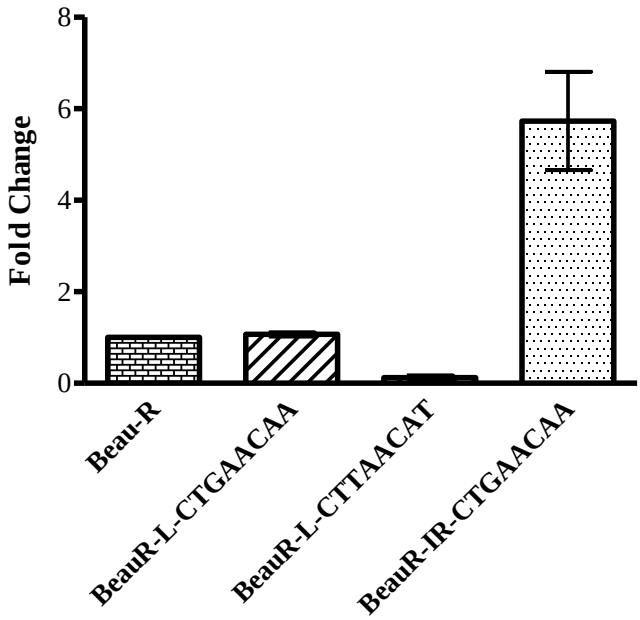


Fig. 7.11. Quantitative PCR analysis of IR mRNA levels. CK cells were infected with $1x10^5$ pfu of Beau-R, BeauR-L-CTGAACAA, BeauR-L-CTTAACAT or BeauR-IR-CTGAACAA. Intracellular RNA was harvested 18 hpi and IR mRNA levels quantified by qRT-PCR using 28S expression as an internal standard. Values were additionally normalised to Gene 3 mRNA levels to account for variations in infection rates. Error bars represent the SD of the mean of two independent experiments assayed in triplicate.

The qRT-PCR data confirmed the northern blot analysis data, showing a 6-fold increase in IR mRNA for BeauR-IR-CTGAACAA. A 0.1-fold change was observed for BeauR-L-CTTAACAT demonstrating that although barely detectable by northern blot analysis the IR mRNA is still transcribed during virus replication. Given that for this rIBV there is only a two-nucleotide match between the IR TRS-B and the TRS-L (at positions 6 and 7) this reduction is consistent with the knowledge that mismatches between the TRS-L and TRS-B can reduce mRNA levels. Again, no change is observed for the IR of BeauR-L-CTGAACAA as the mutation introduced into this virus does not alter the IR TRS-B/TRS-L interaction.

The experiments with the TRS mutants confirmed that the TRS can be responsible for regulating sg mRNA levels, with the number of nucleotide matches between the TRS-B and TRS-L the key factor in this regulation.

7.6 Flexibility of TRS usage

During analysis of the BeauR-L-CTGAACAA IR mRNA the mRNA corresponding to the M gene was also sequenced as a control. This was to confirm that the M mRNA TRS was derived from the complete M TRS-B as per the current understanding of sg mRNA transcription, and that the T→G mutation was not incorporated into the mRNA TRS in a similar mechanism to the IR mRNA. However, this was not found to be the case with the first M mRNA clone sequenced from BeauR-L-CTGAACAA unexpectedly showing the presence of the T→G mutation at position 3 of the mRNA TRS (Fig 7.12A). A second clone of the M mRNA was subsequently sequenced and while the consensus sequence gave the

mRNA TRS as CTTAACAA analysis of the sequence traces showed this was an error and that the dominant nucleotide at position 3 was a G with a minor T population (Fig. 7.12B).

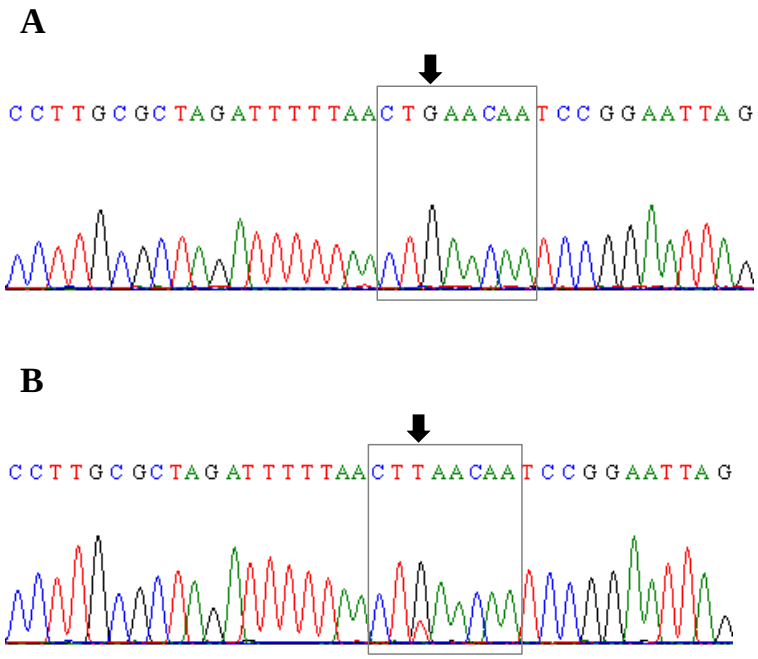


Fig. 7.12. Sequence traces of M gene mRNAs. CK cells were infected with BeauR-IR-CTGAACAA and 20 hpi intracellular RNA was harvested. M gene mRNAs were amplified with primers Leader1 and IR(R) and PCR cloned for sequence analysis with primers M13(F) and M13(R). (A, B) Sequence traces for two clones of M mRNA. Grey boxes outline the TRS-B with position of the TRS-L mutation indicated by black arrow.

Given this finding other sg mRNAs from a BeauR-L-CTGAACAA infection were cloned and sequenced to establish whether the T→G mutation is consistently incorporated into mRNA TRSs and therefore suggestive of a greater degree of flexibility in TRS usage than previously supposed (Fig. 7.13).

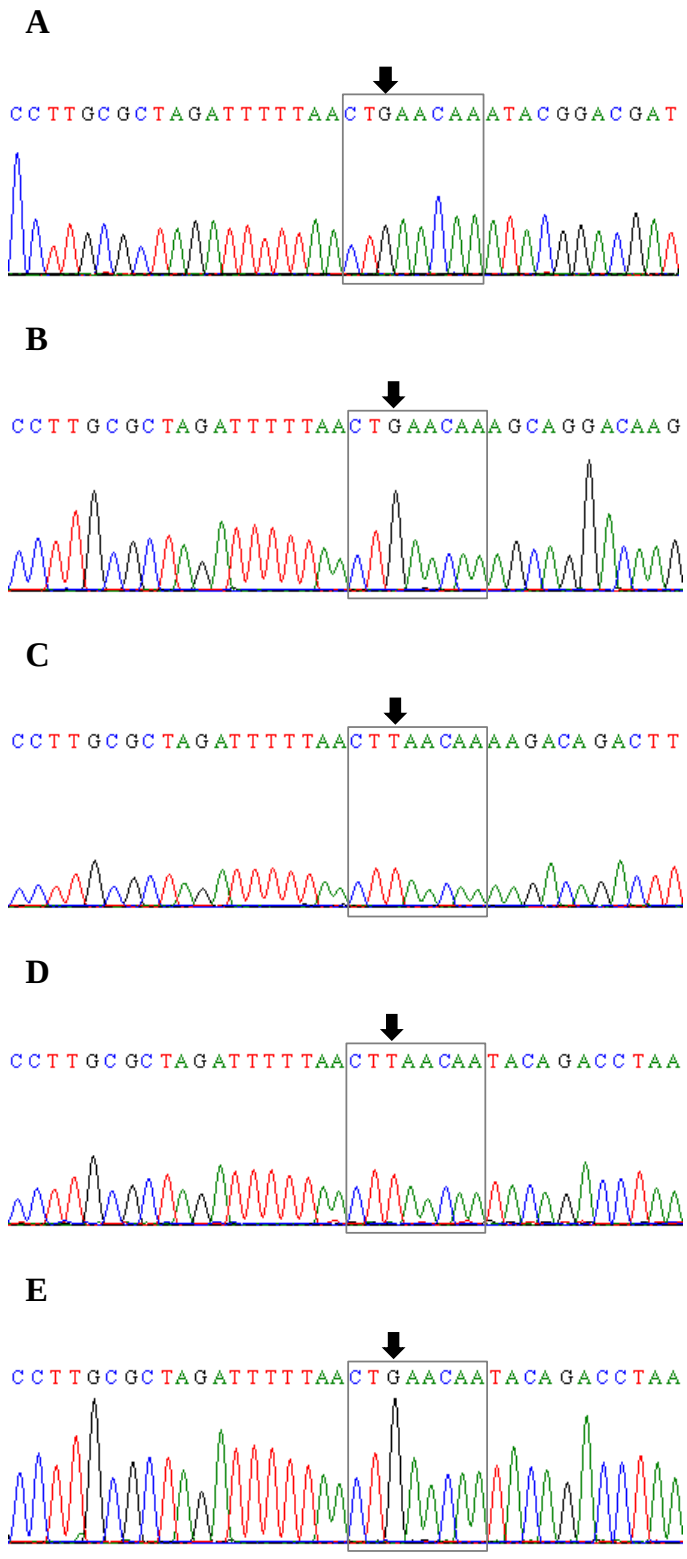


Fig. 7.13. Sequence traces of IBV mRNAs. CK cells were infected with (A, B) BeauR-L-CTGAACAA or (C, D, E) Beau-R and 20 hpi intracellular RNA was harvested. Messenger RNAs were amplified with primers Leader1 and (A) BG147, (B) BG149, (C) BG133, (D, E) BG142, and PCR cloned for sequence analysis with primers M13(F) and M13(R). Sequence traces are shown for (A) Gene 5 (B) N (C) Spike (D, E) Two clones of Gene 3. Grey boxes outline the TRS-B with TRS-L point mutation indicated by black arrow.

Two clones each of Gene 5 (Fig. 7.13A) and N gene (Fig. 7.13B) mRNAs were subject to sequence analysis and in all cases showed the incorporation of the T→G mutation. Given that for BeauR-L-CTGAACAA both the TRS-L and the TRS-Bs for Spike and Gene 3 are CTGAACAA it was not possible to determine the source of the G at position 3 by sequence analysis. However, it was possible to analyse the Spike and Gene 3 mRNAs from a Beau-R infection to investigate whether the mRNA TRS position 3 is a G, as derived from the TRS-B, or a T, as derived from the TRS-L. Due to limited availability of clones only one clone of the Spike mRNA was sequenced and found to contain a T at position 3 (Fig. 7.13C). Two clones of the Gene 3 mRNA were available and sequenced, with one containing a T and one containing a G at position 3 of the TRS (Fig. 7.13D, E).

This result demonstrated that even in a wild type virus infection the TRS of a mRNA can be partly derived from the TRS-L. Although limited numbers of clones were analysed the data suggests that for IBV mRNAs the complete TRS-B is not necessary to permit hybridisation to the TRS-L, and there may be some flexibility in the TRS-B requirements for generation of mRNAs.

These findings could be investigated further via a more in depth analysis of the IR mRNAs transcribed during infection with rIBV BeauR-IR-CTGAACAA. The original IR mRNA clone sequenced from this virus showed the incorporation of the entire TRS-B in the mRNA TRS (see Fig. 7.8). By carrying out sequence analysis of a further four clones of IR mRNA from each replicate of BeauR-IR-CTGAACAA it was possible to determine whether the mRNA TRS is consistently derived from the TRS-B alone, as demonstrated by the first clone, or if the IR mRNA TRS is often

derived from both the TRS-B and TRS-L as seen with the wild type IR mRNA. The sequence alignments of the IR mRNA showed that 4/4 clones for R1 and 2/4 clones for R2 contained a T nucleotide at position 3 of the TRS confirming that in many cases the mRNA TRS is still partly derived from the TRS-L (Fig. 7.14)

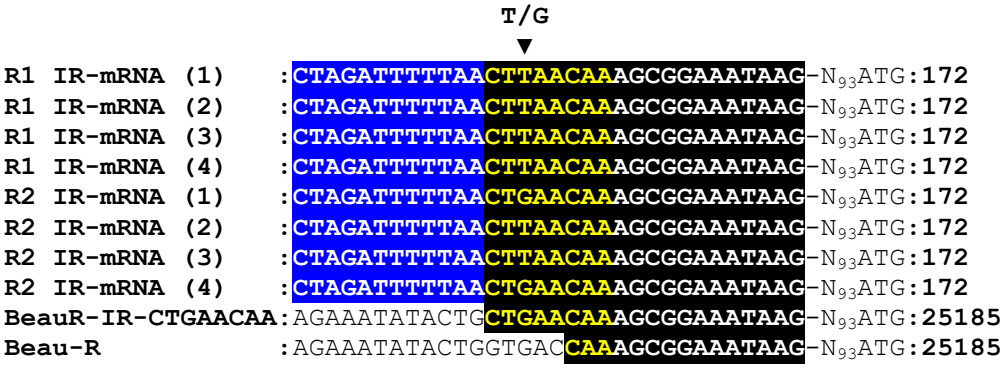


Fig. 7.14. Sequence alignments of BeauR-IR-CTGAACAA IR mRNA. Confluent monolayers of CK cells were infected with replicate 1 (R1) or replicate 2 (R2) of BeauR-IR-CTGAACAA and intracellular RNA extracted 20 hpi. IR mRNAs were amplified by PCR with primers Leader1 and IR(R) and PCR cloned for sequence analysis. Sequence alignments are shown for four clones of each replicate against BeauR-IR-CTGAACAA and Beau-R genomic sequence. Leader sequence highlighted in blue and TRS highlighted in yellow. (▼) indicates position of the variable nucleotide. Relative positions of the mRNA and genome are given on the right along with the distance to the IR ATG.

Overall the results in this chapter have demonstrated that the IBV IR mRNA is unique among the known IBV sg mRNAs in that the associated TRS-B is a non-canonical shortened TRS-B of CAA, not CTTAACAA, and that the IR mRNA TRS is mostly derived from the TRS-L and not the TRS-B as would be expected. Furthermore, by introducing point mutations into the IR TRS-B to extend it from CAA to CTGAACAA, thus better resembling the consensus TRS, it was shown that the IR mRNA expression level could be increased 6-fold over wild type and

demonstrated the importance of the TRS-L/TRS-B interaction in regulating mRNA expression levels. Sequence analysis of all IBV sg mRNAs showed that the TRS of individual mRNAs can be derived, in part, from the TRS-L, suggesting a greater degree of flexibility during transcription of IBV sg mRNAs with regards to the origin of individual mRNA TRSs.

7.7 Discussion

In this chapter the IR mRNA was investigated to determine the transcription mechanisms underlying the use of the non-canonical TRS-B. Until now the evidence has been strongly suggestive that the TRS of each sg mRNA is derived from the TRS-B of the associated gene, and this has led to the general acceptance of the model of transcription in which the discontinuous step occurs during negative-strand synthesis (Sawicki and Sawicki 1995, Schaad and Baric 1994, van Marle, *et al.* 1999, Zuniga, *et al.* 2004). Transcription of nascent negative-sense RNAs is understood to proceed 3'-5' until a TRS-B has been transcribed. The nascent negative-sense RNA then dissociates from the body sequence, is translocated to the 5'-end of the genome and the negative-sense TRS-B complementarily binds to the TRS-L. This initiates a template switch and transcription continues to the very 5'-end of the genome resulting in the addition of a negative copy of the leader sequence to the nascent RNA. The discovery that the IBV IR mRNA contains a complete copy of the IBV CTTAACAA TRS while the associated TRS-B consists of only the final 3 nucleotides of this sequence suggested an altered mechanism of transcription for the IR mRNA.

It was clear that at least part of the IR mRNA TRS originated from virus sequence other than the TRS-B, as would be expected. The IBV reverse genetics system was therefore used to produce a series of rIBVs that could be used to establish the source of the IR mRNA TRS. A hypothesis was established which centred on the idea that the switch in transcription from the genome body to the 5' Leader sequence could occur without the need for a complete TRS-B. This was founded on the understanding that, as well as recognition of the TRS-B, additional elements are involved in controlling the template switch of the RTC during transcription (Dufour, *et al.* 2011, Moreno, *et al.* 2008, Ozdarendeli, *et al.* 2001, Pasternak, *et al.* 2001, Yang, *et al.* 2011). Following hybridisation of the shortened TRS-B to positions 6-8 of the TRS-L transcription would then proceed 3'-5' through positions 5 to 1 of the TRS-L and continue to the 5'-end of the genome. This hypothesis explained the presence of a complete TRS within the IR mRNA and suggested that it is partly derived from the TRS-L.

The first rIBV generated to test this hypothesis, BeauR-L-CTGAACAA, included a T→G point mutation at position 3 of the TRS-L. Analysis of the IR mRNA from this virus showed that this mutation was incorporated into the mRNA TRS and confirmed that the TRS-L was involved in generation of the IR mRNA TRS (Fig. 7.3). While the presence of this mutation confirms that positions 1-3 of the mRNA TRS are derived from the TRS-L it does not necessarily confirm the source of the nucleotides at positions 4 and 5 (AA) as these lie downstream of the point mutation. The A at position 4 of the TRS does align with an A at what would be the corresponding position of a complete IR TRS-B. However, the A at position 5 of the TRS does not correlate with the sequence upstream of the shortened TRS-B and is

therefore most likely derived from the TRS-L, as for positions 1-3. Given the direction of RNA synthesis it is difficult to conceive of a mechanism in which positions 3 and 5 of the mRNA TRS, but not position 4, are derived from the TRS-L. It is therefore proposed that nucleotides 1-5 of the IR mRNA TRS are all derived from transcription of the TRS-L.

A previous model of coronavirus transcription, termed leader-primed transcription, proposed that the discontinuous step occurs during positive-strand RNA synthesis with transcription of the leader sequence occurring prior to hybridisation to a TRS-B (Baric, *et al.* 1983). This theory would suggest that each mRNA TRS is derived from the TRS-L and not the TRS-B. Given the proposed extent of involvement of the TRS-L in the IR mRNA TRS it was necessary to confirm that this alternative model was not the basis for generation of the IR mRNA and that the current model of transcription is valid, albeit in a modified form.

A second rIBV was generated, BeauR-L-CTTAACAT, in which an A→T point mutation was introduced at position 8 in the TRS-L of Beau-R. The IR mRNA TRS from this rIBV did not contain the A→T mutation, confirming that position 8 of the TRS is derived from the TRS-B, with positions 6 and 7 also likely to be derived from the non-canonical TRS-B, as suggested by sequence alignments (Fig. 7.4). Although it is possible that under the leader-primed model the nascent RNA could be transferred to a TRS-B prior to complete transcription of the TRS-L it is unlikely that two different transcription mechanisms are utilized by the virus. Therefore, it is proposed that the IBV IR mRNA is transcribed using an altered form of the model of

discontinuous transcription during negative-strand synthesis, in which part of the mRNA TRS is derived from the TRS-L.

During the identification of the IBV IR mRNA in the work described in Chapter 6 it was clear that one of the main reasons for the previous lack of identification of this mRNA is the low detection level observed by northern blot analysis. One possible explanation for this low level of detection is that the IR mRNA is only transcribed at late stages of replication and is therefore transcribed in lower amounts than all other mRNAs. However, the discovery of the non-canonical TRS-B led to the hypothesis that the lower transcription levels were in fact a direct result of an impaired interaction between the non-canonical TRS-B and the TRS-L and may be a strategy used by the virus to control sg mRNA synthesis and translation of the associated 11kDa protein. It was further hypothesised that the low expression level could be overcome by extending the IR TRS-B to the complete consensus sequence, CTTAACAA.

To test this hypothesis a rIBV, BeauR-IR-CTGAACAA, was generated in which two point mutations were introduced into the IBV genome to extend the non-canonical TRS-B from CAA to CTGAACAA. As mentioned in the results the CTGAACAA sequence was chosen over CTTAACAA to avoid the introduction of a premature stop codon within the M gene sequence, and maximise the chances of a successful rescue. Levels of IR mRNA from BeauR-IR-CTGAACAA were investigated by northern blot analysis and qRT-PCR.

The results showed that the IR mRNA level from BeauR-IR-CTGAACAA was increased 6-fold over the IR mRNA level observed for Beau-R (Fig. 7.10C; Fig. 7.11). This confirmed that the amount of mRNA transcribed from a gene correlates with the number of nucleotide matches between the TRS-L and TRS-B. This hypothesis was further confirmed by the IR mRNA level observed from BeauR-L-CTTAACAT (Fig. 7.10B; Fig. 7.11). For BeauR-L-CTTAACAT the presence of only two nucleotide matches between the TRS-L and TRS-B (positions 6 and 7) led to a further reduction in the IR mRNA level compared to wild type. This was an interesting result as it indicated that an IBV mRNA could be produced from only two common nucleotides within the TRS-L and TRS-B, suggesting that other factors may play a role in generation of a coronavirus sg mRNA, not simply complementation between the two TRSs.

Investigation of the replication kinetics of the TRS mutants demonstrated that the individual mutations introduced into the TRS-L had no effect on virus replication (Fig. 7.9). This suggests that IBV can tolerate minor changes to the TRS-L. In contrast, the peak titre for BeauR-IR-CTGAACAA was approximately 1-$\log_{10}$ lower than Beau-R. One possible explanation for this reduction in peak titres was that the increased efficiency of the IR TRS-L/TRS-B interaction for BeauR-IR-CTGAACAA led to a decrease in the levels of mRNA synthesis for the genes upstream of the IR, i.e. M, Gene 3 and Spike, which would have a negative impact on virus replication. If the RTC is halted more frequently at the IR TRS-B due to the increased complementarity between the TRSs, it is possible that initiation of mRNA synthesis from upstream TRS-Bs is reduced. However, northern blot analysis did not indicate any changes to mRNA levels for the upstream genes (Fig. 7.10C).

The reduction in peak virus titres was therefore most likely due to the generation of a slightly impaired M protein brought about by the amino acid changes introduced into the M gene sequence to accommodate the extension of the IR TRS-B. The amino acid changes G189A and D190E are located in the cytoplasmic tail of the M protein, a region that has been demonstrated to be important for interaction with the virus E protein and subsequently virus particle assembly (Corse and Machamer 2003). It is possible that the changes introduced into M have affected this interaction with the result being an impairment of virus particle assembly and therefore a reduction in virus titre.

It is interesting to note that in both replicates of BeauR-IR-CTGAACAA mutations independently arose directly upstream of the modified TRS-B. At passage 4 these mutations were not stable in the virus population with several interpretations of the sequence data possible, presumably owing to a mixed virus population. However, subsequent work by colleague Sarah Keep has shown that by passage 8 these mutations had settled into dominant populations resulting in additional amino acid changes A189T for R1, and T188L and A189T for R2. It is proposed that these mutations are potential compensatory mutations arising to counteract the changes made to the M gene.

During analysis of the IR mRNAs of each of the TRS-L rIBVs mRNAs corresponding to the M gene were also sequenced as controls. It was expected that the M mRNAs would fit the current understanding of coronavirus transcription, i.e. the mRNA TRS is derived wholly from the TRS-B, and neither TRS-L mutation would be incorporated into the M mRNA TRS. However, it was found that for the

M, and subsequently for the Gene 5 and N, mRNAs from a BeauR-L-CTGAACAA infection all showed incorporation of the T→G mutation (Fig. 7. 12; Fig. 7.13A, B). Similarly, the Spike and Gene 3 mRNAs from a Beau-R infection showed the incorporation of the T nucleotide from position 3 of the TRS-L and not the G nucleotide from the TRS-B as was expected (Fig. 7.13C, D, E). Although these observations are based on sequence analysis of only a very small number of mRNA clones this finding, due to the presence of variant mRNA species for a few clones, raises the possibility that the derivation of the mRNA TRS is more flexible than previously thought.

Attempts were made to investigate this finding in more depth by examining the whole virus population during cellular infection with either Beau-R or BeauR-L-CTGAACAA. The aim was to use alternative techniques to PCR cloning and sequence analysis that would allow the sequence of every mRNA transcribed during infection to be analysed. Initially the possibility of developing a qRT-PCR assay based on existing allelic discrimination assays was investigated. However, due to the limited sequence over which to design primers no primer-probe sets could be identified for several of the IBV mRNAs and it was not possible to pursue this method further.

An alternative approach was also investigated using locked nucleic acid (LNA^{TM}) primer technology (Exiqon) and PCR to specifically amplify mRNAs containing either a T or G at position 3 of the mRNA TRS, and thus demonstrate the presence of one, or both, mRNA species in a virus population. This technology uses a class of nucleic acid analogues in which the ribose ring is "locked" by a methylene bridge

connecting the 2'-O atom with the 4'-C atom. The LNA nucleosides conform to normal Watson-Crick base pairing rules; however the thermal stability of an LNA/DNA duplex is increased over standard DNA/DNA duplexes. Primers containing LNA nucleosides therefore have an increased Tm and can increase the specificity of a PCR.

By designing primers containing either a T or a G LNA nucleoside at the variable position of the TRS it was hoped that the mRNA populations present during virus replication could be established. For example, a PCR to amplify the IR mRNA from a Beau-R infection should yield a product when the primer contains a T LNA nucleoside at position 3 of the TRS, but no product should be amplified if the primer contains a G LNA nucleoside at this position. However, despite several attempts to develop a PCR assay the control PCRs did not show the expected results with products being amplified by primers that should not have been able to base pair to the template DNA. It was suspected that the difference in T_M between a perfect match and a 1-nucleotide mismatch at the T/G position was not great enough to generate the required specificity in the PCR and therefore these experiments were not pursued further.

Two theories are proposed for the incorporation of nucleotides from the TRS-L into all mRNA TRSs. First, it is possible that for IBV there is a natural wobble during transcription that may, or may not lead to complete transcription of the TRS-B prior to the switch to the 5'-end of the genome. This would suggest that the true core TRS for IBV might be the shorter sequence AACAA with the CUU not essential for hybridisation to the TRS-L. Evidence for this theory came from analysis of the

BeauR-IR-CTGAACAA IR mRNA. Sequence analysis of four IR mRNA clones for each of the two replicates of virus showed that in 4/4 clones for R1 and 2/4 clones for R2 the IR mRNA contained the position 3 nucleotide T, as derived from the TRS-L (Fig. 7.14). However, the IR mRNA level of R1 was greatly increased over Beau-R despite the apparent switch to the TRS-L at position 3 in a large proportion of those mRNAs sequenced. This suggests that the increase in mRNA level is largely, if not entirely, due to the extension of the TRS-L/TRS-B interaction to include positions 4 and 5 of the TRS-B.

The second possibility is that the presence of a mismatch between the two nucleotides at position 3, either naturally occurring or introduced, forces the early translocation and hybridisation of the nascent RNA to the TRS-L. Testing of this theory would require the generation of a large series of rIBVs, in which each nucleotide of the TRS-L and TRS-B of a gene are individually mutated, to examine the effects on the origin of the mRNA TRS.

Overall the experiments detailed in this chapter have highlighted some important issues for coronavirus transcription. It has been confirmed that the IBV IR mRNA is transcribed via a non-canonical, shortened TRS-B with the sequence CAA. In a departure from the current model of coronavirus transcription this non-canonical TRS-B leads to a mRNA in which the TRS is derived from transcription of the TRS-L as well as the TRS-B. Analysis of other mRNAs demonstrated that the presence of nucleotides in the mRNA TRS that are derived from the TRS-L might be a more common occurrence than previously suspected. It has additionally been shown for the first time with IBV that the TRS-L/TRS-B interaction directly mediates mRNA

levels during infection via the number of nucleotide matches existing between the two sequences. This work also raises the possibility that additional virus sg mRNAs may be produced by coronaviruses for the expression of as yet unidentified proteins.

Chapter 8: Discussion and Further Work

8.1 Discussion

The avian coronavirus IBV is a positive-sense, single stranded RNA virus belonging to the family *Coronaviridae*. Coronaviruses possess the largest RNA genomes ranging in size from 26-31kb and this, combined with the development of reverse genetics systems that allow manipulation of the genome, has resulted in the suggestion that coronaviruses may be suitable to use as viral vectors for the expression of heterologous genes. The initial aim of this PhD was to investigate the potential for utilizing IBV as a vaccine vector that could be used to vaccinate chickens against both IBV and a second avian viral disease based on the heterologous gene being expressed.

Utilizing the reverse genetics systems developed in the Coronavirus group at the IAH a series of rIBVs, expressing either eGFP or hRluc, were generated to establish what factors are important for the stable expression of foreign genes from IBV, and thereby demonstrating the potential to use IBV as a vaccine vector. Recombinant viruses were designed with the aim of investigating the effect on virus stability of four different factors: (1) genome position of the foreign gene, (2) the nature of the foreign gene, (3) codon optimization of the foreign gene and (4) fusion of the foreign gene to an existing IBV gene. The successful rescue of six different rIBVs out of a planned nine constructs allowed for investigation of factors 1-3 using a variety of analysis techniques as detailed in Chapters 3-5.

Two replicates each of rIBVs BeauR-eGFPΔ5ab, BeauR-hRlucΔ5ab, BeauR-eGFPΔIR, BeauR-hRlucΔIR, BeauR-IBVhRlucΔ5ab and BeauR-hRlucΔ3ab were rescued as defined by the observation of IBV-induced CPE, as well as detection of a PCR product corresponding to the appropriate reporter gene. However, as was discovered during analysis, in some cases nucleotide changes had occurred during the rescue process that differed between the two replicates of a particular rIBV. This was an important observation as it demonstrated an element of unpredictability in the rescue of rIBVs that could lead to inconsistent production of viruses for vaccine purposes. It is not always clear in the literature, when looking at similar studies of recombinant coronaviruses, whether the data presented represents one or more virus replicates, and whether any nucleotide changes outside of the heterologous genes have been presented. It is difficult therefore to establish whether the inconsistencies observed here are due to the specific rescue process, the basic design of the rIBV constructs, or whether this is a feature of all coronavirus recombinants expressing foreign genes.

The analyses of the rIBVs revealed that, in the absence of sequence changes within the heterologous genes, all rIBV constructs were capable of expressing the relevant protein. However, problems were encountered with the stability of the rIBVs following repeated passage on CK cells. Previous studies investigating the expression of heterologous genes from coronaviruses have also reported varying degrees of instability for recombinant viruses following passaging, suggesting that this is a common problem, potentially with the same general underlying cause. A summary of the stability of successfully rescued recombinant viruses is given in Table 8.1.

TABLE 8.1: Stability of Recombinant Coronaviruses

Virus Backbone	Recombinant	Reporter Gene	Stability (Passage at which expression no longer detected)	Reference
IBV (Vero cell-adapted strain P65)	**IBV-EGFP25180** (insertion into IR) **IBV-EGFP27105** (insertion between N and 3'UTR) **IBV-S-EGFP** (fusion to carboxy terminus of S) **IBV-LucΔ3ab** (replacement of ORFs 3a and 3ab) **IBV-Luc25180** (insertion into IR) **IBV-Luc27105** (insertion between N and 3'UTR	eGFP eGFP eGFP fLuc fLuc fLuc	P5 (Vero cells) P5 (Vero cells) P5 (Vero cells) P15 (Vero cells) P2 (Vero cells) P2 (Vero cells)	Shen *et al*. 2006
IBV (Vero cell-adapted strain)	**GIBV** (replacement of ORF 5a)	GFP	P5 (Vero cells)	Youn *et al*. 2005
TGEV (PUR46-MAD strain)	**rTGEV-Δ3-TRS$_{3a}$-GFP** (replacement of ORFs 3a and 3b)	GFP	≥ P20 (Swine testis cells)	Sola *et al*. 2003
FCoV (Black strain)	**recFCoV-GFP** (replacement of ORFs 3a, 3b and 3c) **recFCoV-RL** (replacement of ORFs 3a, 3b and 3c)	GFP *Renilla*	≥ P6 (FCWF-4 cells) ≥ P6 (FCWF-4 cells)	Tekes *et al*. 2008
HCoV-NL63	**icNL63gfp** (replacement of ORF 3)	GFP	≥ 2 months in human ciliated airway epithelium cultures[a]	Donaldson *et al*. 2008
MHV-A59	**MHV-SGFP** (Fusion to carboxy terminus of Spike)	GFP	P6 – but mixed population as early as P2 (LR7 cells)	Bosch *et al*. 2004
MHV-A59	**MHV-ERLM** (insertion between E and M) **MHV-EFLM** (insertion between E and M)	*Renilla* fLuc	≥ P8 (LR7 cells) ≥ P5 – but decreased expression from P5 (LR7 cells)	de Haan *et al*. 2005
FIPV[b]	**FIPV-Δ3abcRL** (replacement of ORFs 3a, 3b and 3c) **FIPV-Δ3abcFL** (replacement of ORFs 3a, 3b and 3c)	*Renilla* fLuc	≥ P8 (FCWF-4 cells) ≥ P8 (FCWF-4 cells)	

[a]No passage number indicated
[b]Feline infectious peritonitis virus

As indicated in Table 8.1 only two recombinants remained stable for over ten passages in cell culture, with the majority losing heterologous gene expression between P2 and P8; a similar result to that observed in this study where instability arose between P4 and P10 depending on the rIBV and conditions of passaging (Table 5.2). Several additional studies have also demonstrated the ability of coronaviruses to express heterologous genes, thus adding to the evidence in support of coronaviruses as viral vectors (de Haan, *et al.* 2005, Fischer, *et al.* 1997, Raaben 2009, Ribes, *et al.* 2011, van den Worm, *et al.* 2012). However, no passaging experiments were conducted in these reports and therefore the stability of the recombinant viruses is unclear.

In this study the differences observed in stability between those rIBVs expressing eGFP and those expressing hRluc was minimal, with only one or two passages between the related constructs, suggesting that the type of heterologous gene had little effect on virus stability. This was in contrast to the observations of Shen *et al*. (Shen, *et al.* 2009) where recombinants IBV-Luc25180 and IBV-Luc27105, expressing fLuc, were less stable than similar recombinants expressing eGFP, and expression of fLuc was not detected past P1 of the virus. This apparent discrepancy may be explained by the differences between fLuc and hRluc, such as protein size, and correlates with previous findings that fLuc could not be stably expressed from IBV D-RNAs (Stirrups, *et al.* 2000). Similarly, the work of de Haan *et al*. (de Haan, *et al.* 2005) demonstrated that fLuc was less stable than *Renilla* luciferase when expressed from between the E and M genes of MHV, although an investigation into the problems with fLuc determined that the reduced stability was not a consequence of the larger size of fLuc.

It has been suggested previously that the replacement of accessory genes results in recombinants with increased stability over those in which the heterologous gene has been inserted as an additional expression cassette. This was clearly demonstrated by Shen *et al*. (Shen, *et al.* 2009) where the authors found that the replacement of ORFs 3a and 3b by fLuc was considerably more stable than any other construct in the study, with expression of fLuc still detected at P15 compared to all others where expression was lost by P5 at the latest.

This finding can be compared to the results obtained in this study for BeauR-hRlucΔ3ab as, although the exact sequence of IBV-LucΔ3ab (Shen, *et al.* 2009) has been difficult to determine from the publication, the two recombinant viruses consist of the replacement of the same genome region, albeit in different virus strains. Analysis of BeauR-hRlucΔ3ab showed that although instability in the form of deletion variants of the virus arose in early passages, the wild type virus was also maintained up to P12 when passaged at an MOI of 0.01 (Fig. 5.3). Had this recombinant been assessed purely on the basis of detection of luciferase activity it would have been classed as a highly stable rIBV. With no analysis given in the Shen paper as to the virus populations present, it is interesting to speculate that at P15 IBV-LucΔ3ab was in fact present as a mixed population containing deletion variants, and therefore not as stable as suggested.

If no mixed populations are assumed for IBV-LucΔ3ab at P15 (Shen, *et al.* 2009) then comparison to rIBV BeauR-hRlucΔ3ab may highlight the importance of the sequence alterations made to viruses to accommodate the expression of heterologous genes. Sequence analysis of BeauR-hRlucΔ3ab showed that several unintended

changes had arisen in the two replicates of this rIBV in the region of the inserted TRS-B for expression the E gene, and along with the failure to rescue BeauR-eGFPΔ3ab, suggested that the sequence alterations made were not favourable to successful rescue. As mentioned, the precise sequence of IBV-LucΔ3ab was difficult to determine and so it is not clear how the expression of E was achieved for this virus. It is possible therefore that the overall construction of IBV-LucΔ3ab was more conducive to successful rescue and stable passaging than BeauR-hRlucΔ3ab. This would suggest that the specific sequence construction of a recombinant may be more important than either the heterologous gene being expressed, or the position it is located, and may help explain the differences observed in stability between seemingly analogous viruses.

The data from this, and similar, studies suggest that overall, the stability of coronaviruses expressing heterologous genes is low, and raises the question of how feasible it is to utilize coronaviruses as vectors. A second family of positive-strand RNA viruses, the *Togaviridae,* also utilize sg RNAs for the expression of virus structural proteins, in this case a single sg mRNA known as 26S, which is transcribed from an internal promoter present in the minus-strand RNA. As a result viruses from this family, particularly those in the *alphavirus* genus, have also been extensively studied as potential viral vectors. Alphaviruses have been utilized as expression vectors in several formats including, as replication-competent vectors, in which heterologous genes are expressed from additional sg mRNAs (Hahn, *et al.* 1992, Pugachev, *et al.* 1995) or as virus-like particles, in which the heterologous gene is expressed from the 26S sg mRNA through replacement of the structural genes (reviewed in (Rayner, *et al.* 2002)).

As with coronaviruses however, initial attempts to generate stable replication-competent vectors, in which a heterologous gene was expressed as an additional sg mRNA, showed little promise. Heterologous genes were found to be deleted following only a few passages in cell culture; a result presumed to be due to various aspects including homologous recombination between duplicated promoter sequences, the location of the additional promoter sequence and the nature of the heterologous gene (Hahn, *et al.* 1992, Pugachev, *et al.* 1995). A solution to this problem was found with the insertion of a picornavirus, encephalomyocarditis virus (EMCV), IRES element between ORFs instead of a second sg promoter, thus allowing expression of the heterologous gene from the existing sg mRNA (Pugachev, *et al.* 2000). A similar strategy was attempted for IBV using D-RNAs for the expression of CAT via an EMCV IRES (Dove, *et al.* 2004). However, during IBV helper virus rescue no IRES-mediated translation of the CAT protein could be detected, suggesting that this mechanism may not be suitable to improve the stability of recombinant coronaviruses, although D-RNAs may not be representative of the situation observed with full-length genomic RNA.

Understanding how instability arises in recombinant coronaviruses, and overcoming this, may be necessary before the full potential of coronavirus-based viral vectors can be realised. Homologous recombination, a result of template-switching during RNA replication, has been demonstrated to occur in several coronaviruses including MHV (Makino, *et al.* 1986, van der Most, *et al.* 1992), HCoV-OC43 (Lau, *et al.* 2011) and IBV (Hughes 2011, Kottier, *et al.* 1995, Lee and Jackwood 2000) and is the most likely explanation for the deletion of heterologous genes from recombinant coronaviruses. Recombination may occur due to the presence of similar sequences

between the foreign gene and the virus genome, the presence of similar sequences in the genome itself, such as TRSs, or the presence of secondary RNA structures within the foreign gene or virus genome.

In this study stability was found to increase upon passaging in cell culture at a low MOI of 0.01, a result that may be explained by the indirect control of the occurrence of recombination events. By infecting cells at a low MOI the number of genomes initially present in a single cell will be low and therefore limit the potential for recombination events to occur between genomes until further rounds of replication have taken place. However, even with passaging at an MOI of 0.05 the stability of an MHV Spike-GFP fusion recombinant did not extend beyond P2 before the occurrence of deletions within the GFP sequence and the appearance of a mixed virus population (Bosch, *et al.* 2004). Taken together with the data from this study, which showed that passaging at a low MOI could only improve stability by a small number of passages, this suggests that passaging at a low MOI is not sufficient to prevent the deletion of heterologous genes from recombinant coronaviruses, but may be beneficial in some circumstances.

Overall this project has demonstrated that IBV does have potential to be used as a vaccine vector although further work is required. The data suggests that a successful vector would be based on constructs in which Gene 5 has been replaced by the heterologous gene, as these were found to be the most stable, achieving ≥12 passages in cell culture whilst still expressing the heterologous protein (Table 5.2). The benefits of codon optimization of the heterologous gene were not fully established, although there was some evidence from one passaging experiment to suggest that

this may result in a modest increase in the stability of the virus (Table 5.2). The stability of recombinant coronaviruses expressing heterologous genes remains a problem, with no obvious solution as to how to overcome the natural process of homologous recombination that appears to lead to the deletion of heterologous genes. It remains to be determined how the stability of rIBVs in cell culture translates to stability *in vivo* and this will be the key factor in establishing whether IBV can be utilized as a vaccine vector.

The IR and a novel sg mRNA

The analysis of the rIBVs generated for investigation of the use of IBV as a vaccine vector led to the identification of a novel sg mRNA for IBV and the closely related TCoV, and the suggestion of the existence of a 5th accessory protein for the gammacoronaviruses. The novel sg mRNA was associated with a region of the IBV and TCoV genome located between the M gene and Gene 5, and referred to variably as the IR, ORF 4b or ORF X (Armesto, *et al.* 2009, Cao, *et al.* 2008, Gomaa, *et al.* 2008, Hewson, *et al.* 2011). Transcription of the sg mRNA was discovered to be initiated from a non-canonical TRS-B, of CAA, located approximately 100nts upstream of the AUG of the IR, within the M gene. It was demonstrated that the TRS of the IR mRNA was derived in part from this non-canonical TRS-B at positions 6-8, while positions 1-5 were derived from the TRS-L. This is unique amongst the currently known sg mRNAs of coronaviruses and contradicts the current understanding that a mRNA TRS is derived from the TRS-B (van Marle, *et al.* 1999, Zuniga, *et al.* 2004).

The current model of coronavirus sg mRNA transcription can be thought of as a three-step process consisting of: 1) pause/stop of the RTC following recognition and copying of a TRS-B, (2) base-pairing of the nascent negative-sense RNA to the TRS-L and (3) a template switch to allow transcription of the nascent RNA to continue to the very 5' end of the genome. The exact mechanisms of coronavirus transcription are not yet fully understood but there is increasing evidence to suggest that additional factors are involved such as specific sequence elements and RNA-RNA interactions, the formation of secondary RNA structures such as stem-loops, or RNA-protein interactions, (Baric, *et al.* 1988, Dufour, *et al.* 2011, Keane, *et al.* 2012, Mateos-Gomez, *et al.* 2011, Moreno, *et al.* 2008, Ozdarendeli, *et al.* 2001, Sola, *et al.* 2011, Sola, *et al.* 2011, Yang, *et al.* 2011, Zhang and Lai 1995, Zuniga, *et al.* 2010).

The TRS-B is believed to play an important role in the initiation of sg mRNA synthesis by acting as a pause/stop signal for the RTC. However, this belief is difficult to reconcile with the finding that the IR sg mRNA is transcribed via a TRS-B of CAA. This 3nt sequence occurs approximately 280 times throughout the Beau-R genome without generating sg mRNAs and therefore it is proposed that sequence recognition of a TRS-B may only have a limited role in the initial steps of sg mRNA transcription, i.e. pausing of the RTC. Particularly in the case of the IR sg mRNA, if not all sg mRNAs, the suggestion would be that the additional elements such as secondary RNA structures, are the dominant factor for initiation of sg mRNA synthesis.

Further evidence in favour of this possibility was gained from analysis of rIBVs BeauR-eGFPΔIR and BeauR-hRlucΔIR. In the case of BeauR-eGFPΔIR and BeauR-hRlucΔIR it was discovered that transcription of the reporter gene sg mRNAs was mediated by the naturally occurring IR TRS-B, and not the consensus TRS-B inserted upstream of the gene sequence, suggesting that simply the presence of a TRS-B is not sufficient for initiating transcription. Investigation of the BCoV Gene 5 sg mRNA also demonstrated the preferential use of a non-canonical TRS-B of GGUAGAC despite the presence of a consensus, canonical UCCAAAC sequence located 14nts downstream, again suggesting that sequence recognition of a canonical TRS-B is not the main factor in initiating the synthesis of sg mRNAs (Ozdarendeli, *et al.* 2001).

Additionally, analysis of the sg mRNAs produced by rIBV BeauR-L-CTTAACAT did not demonstrate the presence of any additional, unexpected, sg mRNAs despite the presence of a now perfect match between the TRS-L and a CUUAACAU sequence beginning at nucleotide 24685 within the M gene of the Beau-R genome. If sequence recognition of a TRS was sufficient to initiate transcription then a sg mRNA originating from this position would be expected, but this was not the case. However, this finding was in contrast to reports concerning SARS-CoV and arterivirus EAV, in which mutations in the TRS-L led to the activation of non-canonical TRS-Bs and the synthesis of novel sg mRNAs (Pasternak, *et al.* 2003, Yount, *et al.* 2006). These reports suggested that provided a stable TRS-L/TRS-B duplex could be formed new sg mRNAs could be generated from sequences that had previously not been recognised as a TRS by the RTC. It is possible that the mutations introduced in these cases resulted in changes, such as introduction of

secondary RNA structures or creation of protein binding sites, that were not previously present, and it is these factors that led to the synthesis of sg mRNAs.

Given the suggestion that the TRS-B is not as important for initiation of sg mRNA synthesis as previously thought, the non-canonical IR TRS-B provided a means to investigate its role in regulating the levels of sg mRNA transcription of the structural and accessory genes. Previous studies have demonstrated that altering the TRS-L/TRS-B interaction can lead to changes in sg mRNA levels for affected genes (Alonso, *et al.* 2002, Sola, *et al.* 2005, van Marle, *et al.* 1999, Yount, *et al.* 2006, Zhang and Lai 1995, Zuniga, *et al.* 2004). By extending the IR TRS-B to CUGAACAA, to match that of Gene 3 and the S gene, it was demonstrated that the levels of IR sg mRNA could be increased 6-fold over the natural level (Fig. 7.11), thus confirming the importance of the TRS-L/TRS-B interaction in sg mRNA synthesis.

The IR sg mRNA contains a putative ORF for expression of an 11kDa protein, with the evidence suggesting that this protein is translated during IBV infection (Fig. 6.9), therefore resulting in a 5^{th} accessory protein of IBV and related gammacoronaviruses. The non-canonical transcription mechanism utilized by the IR, and involving incorporation of TRS-L nucleotides into the mRNA TRS, suggests that there is potential for other coronaviruses to synthesise sg mRNAs in a similar way, via as yet unidentified shortened TRS-Bs. It is possible therefore, that there may be other additional proteins produced by coronaviruses that are yet to be discovered.

Studies of arteriviruses EAV and porcine reproductive and respiratory virus led to the discovery of a novel gene that overlaps the arterivirus GP5 structural gene, and encodes a potential additional structural protein referred to as ORF5a (Firth, *et al.* 2011, Johnson, *et al.* 2011). Expression of the ORF5a protein is predicted to be from the same sg mRNA as GP5, with translation proposed to involve a mechanism of leaky ribosomal scanning. These findings, along with the discovery of the IR sg mRNA, suggests that there are several mechanisms by which coronaviruses, and the closely related arteriviruses, may be able to expand their repertoire of proteins, not all of which are fully appreciated at this time.

The discovery of the IR sg mRNA, and the possibility that there may be other unidentified sg mRNAs/proteins, also has important implications for the use of IBV, or other coronaviruses, as vaccine vectors. It has been shown in several studies that replacement of accessory genes results in the most stable recombinant viruses. However, if the full complement of proteins is yet to be discovered for a virus, these gene replacements may indirectly affect more genes than those intended. While this may not alter growth characteristics in cell culture, as demonstrated by the limited effects of the deletion of coronavirus accessory proteins (Casais, *et al.* 2005, de Haan, *et al.* 2002, Hodgson, *et al.* 2006), effects on *in vivo* replication may be enough to alter the effectiveness of any potential vaccine.

In summary, the data presented in this thesis has demonstrated that IBV has the potential to be utilized as a vaccine vector based on the replacement of accessory Gene 5 with a heterologous gene. The stability of the Δ5ab rIBVs in cell culture was comparable to, if not better than, similar coronavirus constructs that have been

proposed as viral vectors, although *in vivo* stability was not determined. Important discoveries were made with the identification of a novel sg mRNA corresponding to the IR of the gammacoronaviruses, and the presence of a potential 11kDa protein that would constitute a 5th novel accessory protein of closely related gammacoronaviruses. The transcription mechanism involved in synthesis of the IR sg mRNA demonstrated the importance of TRS-L/TRS-B interactions in regulating the levels of sg mRNAs produced and suggests that current model of coronavirus sg mRNA synthesis may not account for the complexity of the interactions occurring during virus replication.

8.2 Further Work

IBV as a potential vaccine vector

The data presented in this thesis demonstrated that the replacement of Gene 5, encoding accessory proteins 5a and 5b, resulted in the most stable rIBVs and therefore offers the best target for further development of an IBV-based vector vaccine. Several follow-up experiments are suggested to explore this possibility further. First, BeauR-hRlucΔ5ab and BeauR-IBVhRlucΔ5ab should be passaged further, in cell culture, to determine if codon optimization does result in a more stable rIBV and whether such a process would be beneficial for vaccine design. Second, animal experiments are required to demonstrate the stability of the rIBVs *in vivo*, as well as the generation of antibodies against the heterologous protein and an effective immune response against IBV. Finally, it will be necessary to demonstrate that rIBVs can be generated expressing genes from other avian viruses in place of

Gene 5. These rIBVs should be at least as stable as the Δ5ab rIBVs tested in this study to warrant consideration as a viable vaccine vector candidate.

The IR

Despite the advances made in this thesis regarding the knowledge of the IR, and the associated ORF, much remains unknown. Further work is required not only with regards to the IR protein, but also to understand the mechanisms involved in coronavirus sg mRNA transcription. Ideally, an antibody against the IR protein is required to definitively prove that it is translated during virus infection. Due to the failure of the peptide antibody generated in this study an alternative approach, such as the generation of a recombinant protein, will be required. A functional antibody can subsequently be used to determine the cellular location of the protein during infection and possibly any viral, or host, proteins it interacts with.

The unique TRS-B of the IR can be used to further study the mechanisms involved in coronavirus sg mRNA transcription. Having demonstrated the role of the TRS in determining the levels of sg mRNA transcription, further rIBVs can be generated to fully establish the core sequence of the IBV TRS required for achieving the maximum level of transcription possible from a gene. It would also be interesting to determine whether such a short TRS-B could be introduced for one of the essential structural genes, thereby reducing the level of transcription of that gene, and offering a mechanism for the molecular attenuation of coronaviruses for vaccine purposes. Analysis of the sequence at the site of the IR TRS-B may help in identifying particular RNA structures or RNA, or protein, binding sites that are important for RTC recognition of a TRS-B.

APPENDICES

Appendix 1: Primer Lists

TABLE A1.1: Overlapping PCR Primers

Primer (F/R)	Sequence[a, b]
A(F)	TTT***GGCGCGCC***CGCTAAGTGTGAACCAGACCACTTGCC
C(R)	CCC***TTAATTAA***GTCGATCTAGATTTAGTATCAGCACCC
A-eGFP(R)	GCTCCTCGCCCTTGCTCACCATCGTCCGTATTTGTTAAG TTTTTG
B-eGFP(F)	CAAAAACTTAACAAATACGGACGATGGTGAGCAAGGG CGAGGAGC
B-eGFP(R)	GCTATTGCTCCGCGAAAAGGTTATTACTTGTACAGCTC GTCC
C-eGFP(F)	GGACGAGCTGTACAAGTAATAACCTTTTCGCGGAGCAA TAGC
A-hRluc(R)	GTCGTACACCTTGGAAGCCATCGTCCGTATTTGTTAAGT TTTTG
B-hRluc(F)	CAAAAACTTAACAAATACGGACGATGGCTTCCAAGGTG TACGAC
B-hRluc(R)	GCTATTGCTCCGCGAAAAGGTTATTACTGCTCGTTCTTC AGC
C-hRluc(F)	GCTGAAGAACGAGCAGTAATAACCTTTTCGCGGAGCAA TAGC

[a] Bold italics denotes added restriction sites
[b] Underlined sequence denotes sequence derived from reporter genes

TABLE A1.2: Reporter Gene Cloning Primers

Primer (F/R)	Sequence[a, b]
eGFP-IR(F)	TTT***GCTAGC***GGCCAACTTAACAAATACGGACGATGGTG AGCAAGGGCGAGG
eGFP-IR(R)	AAA***CCCGGG***TTATTACTTGTACAGCTCGTCCATGCC
hRluc-IR(F)	TTT***GCTAGC***GGCCAACTTAACAAATACGGACGATGGCT TCCAAGGTGTACG
hRluc-IR(R)	AAA***CCCGGG***TTATTACTGCTCGTTCTTCAGCACGCGC

[a] Bold italics denotes added restriction sites
[b] Underlined sequence denotes sequence derived from reporter genes

TABLE A1.3: Site Directed Mutagenesis Primers

Primer (F/R)	Sequence[a]
TRS-L T→G (F)	TAGATTTTTAACT**G**AACAAAACGGAC
TRS-L T→G (R)	GTCCGTTTTGTT**C**AGTTAAAAATCTA
TRS-L A→T (F)	TAGATTTTTAACTTAACA**T**AACGGAC
TRS-L A→T (R)	GTCCGTT**A**TGTTAAGTTAAAAATCTA

[a] Nucleotides in bold represent introduced mutations

TABLE A1.4: Confirmation PCR Primers

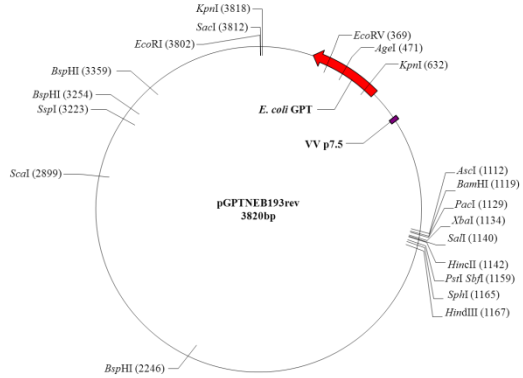
Primer (F/R)	Sequence
5abS (F)	GAGCTATTAACGGTGTTACC
BG147 (R)	AATCTAATCCTTCTCTCAGA
BG52 (F)	GAATGGTGTTCTTTATTG
BG146 (R)	TCTAACACTCTAAGTTGAG
BG49 (F)	TGACGAATTGTCAAAATG
BG144 (R)	AGACAGACACGCAAACACTG
BG148 (R)	ATGTTTTCGTTATCAGGAAC

TABLE A1.5: Sequencing and Additional Primers

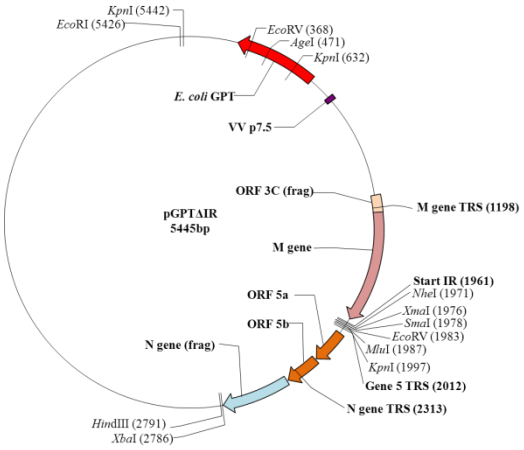
Primer (F/R)	Sequence
BG2 (R)	AGTGCCAAACAACCCCTGA
BG48 (F)	CGCCATACAAGCAAATGC
BG50 (F)	TAGCGCTCCAACAACTAA
BG51 (F)	GCAGTTATTGTTAACGAG
BG54 (F)	ACCACCTAAAGTCGGTTC
BG56 (F)	CAACAGCGCCCAAAGAAG
BG69 (F)	GCTTTTGCCACTATTATCTTC
BG70 (F)	CAAATACGGACGATGAAATG
BG133 (R)	ATTGCCAGTATTATACTGGC
BG137 (R)	CTTGCATTTCTGCTGTTATG
BG140 (R)	TTGAATACGATCAATTTCAC
BG142 (R)	AGGGATCAAATACTTCTGTG
BG143 (R)	CAAGAGTACAATTTGTCTCG
BG145 (R)	CTGTTGCTACACTTTCTAGC
BG149 (R)	TTCCACTCCTACCACGGTTC
93/100 (R)	GCTCTAACTCTATACTAGCCT
eGFP-S (F)	CTACCCCGACCACATGAAGC
eGFP-S (R)	TGTTGTAGTTGTACTCCAGC
hRluc (F)	ATGGCTTCCAAGGTGTACGAC
hRluc (R)	TTATTACTGCTCGTTCTTCAGC
hRluc-S (F)	CTTTCACTACTCCTACGAGC
hRluc-S (R)	CAGTTTCCGCATGATCTTGC
IBVhRluc-S (F)	GATTGGGGTGCATGTCTTGC
IBVhRluc-S (R)	TCATAATTTTAGATGGAAGC
Leader1 (F)	CTATTACACTAGCCTTGCGC
IR (R)	GACCACATCCTACAACAACC
M13 (F)	GTAAAACGACGGCCAG
M13 (R)	CAGGAAACAGCTATGAC

Appendix 2: Plasmid Maps

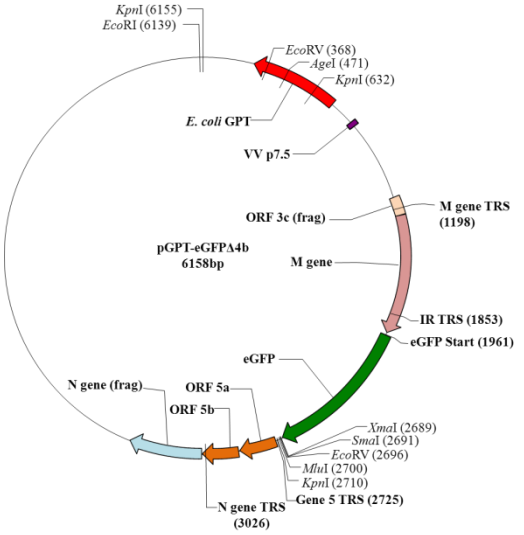
pGPTNEB193



pGPT-ΔIR



pGPT-eGFPΔ4b



Appendix 3: Recombinant IBV genome fragment sequences

BeauR-eGFPΔ5ab

*Asc*I ↓24979

GGCGCGCCGCTAAGTGTGAACCAGACCACTTGCCTAAAGATATATTTGTTTGTACA
CCGGATAGACGTAATATCTACCGTATGGTGCAGAAATATACTGGTGACCAAAGCGG
AAATAAGAAAAGGTTTGCTACGTTTGTCTATGCAAAGCAGTCAGTAGATACTGGCG
AGCTAGAAAGTGTAGCAACAGGAGGAAGTAGTCTTTACACATAAATGTGTGTGTGT
AGAGAGTATTTAAAATTATTCTTTAATAGCGCCTCTGTTTTAAGAGCGCATAAGAG
TATTTATTTTGAGGATACTAATATAAATCCTCTTTGTTTTATACTCTCCTTTCAAG
AGCTATTAACGGTGTTACCTTTCAAGTAGATAATGGAAAAGTCTACTACGAAGGAA
CACCAGTTTTCCAAAAAGGTTGTTGTAGGATGTGGTCCAATTATAAGAAAGAATAA
TTGAACCACCTACTACACTTATTTTTATAAGAGGTACCTTACTTAACAAAAACTTA
ACAAATACGGACG**ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCA**
TCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAG
GGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGT
GCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATG
CCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAA
GACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGA
AGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC
TACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGT
GAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACT
ACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTAC
CTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGT
CCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACA
AGTAATAACCTTTTCGCGGAGCAATAGCAAGAAAAGCTCGAATTTATCTGAGAGAA
GGATTAGATTGTGTTTACTTTCTTAACAAAGCAGGACAAGCAGAGCCTTGTCCCGC
GTGTACCTCTCTAGTATTCCAAGGGAAAACTTGTGAGGAACACATACATAATAATA
ATCTTTTGTCATGGCAAGCGGTAAAGCAGCTGGAAAAACAGACGCCCCAGCGCCAG
TCATTAAACTAGGAGGACCAAAACCACCTAAAGTCGGTTCTTCTGGAAATGCATCT
TGGTTTCAAGCAATAAAAGCCAAGAAGTTAAATACACCTCCGCCCAAGTTTGAAGG
TAGCGGTGTTCCTGATAACGAAAACATTAAGCCAAGCCAGCAACATGGATACTGGA
GACGCCAAGCCAGGTTTAAGCCAGGCAAAGGTGGAAGAAAACCAGTCCCAGATGCT
TGGTACTTTTACTATACTGGAACAGGACCTGCCGCTGACCTGAACTGGGGTGATAC
TCAAGATGGTATAGTGTGGGTTGCTGCTAAGGGTGCTGATACTAAATCTAGATCGA
C**TTAATTAA**
*Pac*I

M-IR-**eGFP**-5b-N

BeauR-hRlucΔ5ab

*Asc*I ↓24979

GGCGCGCCGCTAAGTGTGAACCAGACCACTTGCCTAAAGATATATTTGTTTGTACA
CCGGATAGACGTAATATCTACCGTATGGTGCAGAAATATACTGGTGACCAAAGCGG
AAATAAGAAAAGGTTTGCTACGTTTGTCTATGCAAAGCAGTCAGTAGATACTGGCG
AGCTAGAAAGTGTAGCAACAGGAGGAAGTAGTCTTTACACATAAATGTGTGTGTGT
AGAGAGTATTTAAAATTATTCTTTAATAGCGCCTCTGTTTTAAGAGCGCATAAGAG
TATTTATTTTGAGGATACTAATATAAATCCTCTTTGTTTTATACTCTCCTTTCAAG
AGCTATTAACGGTGTTACCTTTCAAGTAGATAATGGAAAAGTCTACTACGAAGGAA
CACCAGTTTTCCAAAAAGGTTGTTGTAGGATGTGGTCCAATTATAAGAAAGAATAA
TTGAACCACCTACTACACTTATTTTTATAAGAGGTACCTTACTTAACAAAAACTTA
ACAAATACGGACG**ATGGCTTCCAAGGTGTACGACCCCGAGCAACGCAAACGCATGA**
TCACTGGGCCTCAGTGGTGGGCTCGCTGCAAGCAAATGAACGTGCTGGACTCCTTC
ATCAACTACTATGATTCCGAGAAGCACGCCGAGAACGCCGTGATTTTTCTGCATGG
TAACGCTGCCTCCAGCTACCTGTGGAGGCACGTCGTGCCTCACATCGAGCCCGTGG
CTAGATGCATCATCCCTGATCTGATCGGAATGGGTAAGTCCGGCAAGAGCGGGAAT
GGCTCATATCGCCTCCTGGATCACTACAAGTACCTCACCGCTTGGTTCGAGCTGCT
GAACCTTCCAAAGAAAATCATCTTTGTGGGCCACGACTGGGGGGCTTGTCTGGCCT
TTCACTACTCCTACGAGCACCAAGACAAGATCAAGGCCATCGTCCATGCTGAGAGT
GTCGTGGACGTGATCGAGTCCTGGGACGAGTGGCCTGACATCGAGGAGGATATCGC
CCTGATCAAGAGCGAAGAGGGCGAGAAAATGGTGCTTGAGAATAACTTCTTCGTCG
AGACCATGCTCCCAAGCAAGATCATGCGGAAACTGGAGCCTGAGGAGTTCGCTGCC
TACCTGGAGCCATTCAAGGAGAAGGGCGAGGTTAGACGGCCTACCCTCTCCTGGCC
TCGCGAGATCCCTCTCGTTAAGGGAGGCAAGCCCGACGTCGTCCAGATTGTCCGCA
ACTACAACGCCTACCTTCGGGCCAGCGACGATCTGCCTAAGATGTTCATCGAGTCC
GACCCTGGGTTCTTTTCCAACGCTATTGTCGAGGGAGCTAAGAAGTTCCCTAACAC
CGAGTTCGTGAAGGTGAAGGGCCTCCACTTCAGCCAGGAGGACGCTCCAGATGAAA
TGGGTAAGTACATCAAGAGCTTCGTGGAGCGCGTGCTGAAGAACGAGCAGTAATAA
CCTTTTCGCGGAGCAATAGCAAGAAAAGCTCGAATTTATCTGAGAGAAGGATTAGA
TTGTGTTTACTTTCTTAACAAAGCAGGACAAGCAGAGCCTTGTCCCGCGTGTACCT
CTCTAGTATTCCAAGGGAAAACTTGTGAGGAACACATACATAATAATAATCTTTTG
TCATGGCAAGCGGTAAAGCAGCTGGAAAAACAGACGCCCCAGCGCCAGTCATTAAA
CTAGGAGGACCAAAACCACCTAAAGTCGGTTCTTCTGGAAATGCATCTTGGTTTCA
AGCAATAAAAGCCAAGAAGTTAAATACACCTCCGCCCAAGTTTGAAGGTAGCGGTG
TTCCTGATAACGAAAACATTAAGCCAAGCCAGCAACATGGATACTGGAGACGCCAA
GCCAGGTTTAAGCCAGGCAAAGGTGGAAGAAAACCAGTCCCAGATGCTTGGTACTT
TTACTATACTGGAACAGGACCTGCCGCTGACCTGAACTGGGGTGATACTCAAGATG
GTATAGTGTGGGTTGCTGCTAAGGGTGCTGATACTAAATCTAGATCGAC**TTAATTA**
A *Pac*I

M-IR-**hRluc**-**5b**-N

BeauR-eGFPΔIR

🡻24406

CATATGGTAGAAAACTTAACAATCCGGAATTAGAAGCAGTTATTGTTAACGAGTTT
CCTAAGAACGGTTGGAATAATAAAAATCCAGCAAATTTTCAAGATGCCCAACGAGA
CAAATTGTACTCTTGACTTTGAACAGTCAGTTCAGCTTTTTAAAGAGTATAATTTA
TTTATAACTGCATTCTTGTTGTTCTTAACCATAATACTTCAGTATGGCTATGCAAC
AAGAAGTAAGGTTATTTATACACTGAAAATGATAGTGTTATGGTGCTTTTGGCCCC
TTAACATTGCAGTAGGTGTAATTTCATGTACATACCCACCAAACACAGGAGGTCTT
GTCGCAGCGATAATACTTACAGTGTTTGCGTGTCTGTCTTTTGTAGGTTATTGGAT
CCAGAGTATTAGACTCTTTAAGCGGTGTAGGTCATGGTGGTCATTTAATCCAGAAT
CTAATGCCGTAGGTTCAATACTCCTAACTAATGGTCAACAATGTAATTTTGCTATA
GAGAGTGTGCCAATGGTGCTTTCTCCAATTATAAAGAATGGTGTTCTTTATTGTGA
GGGTCAGTGGCTTGCTAAGTGTGAACCAGACCACTTGCCTAAAGATATATTTGTTT
GTACACCGGATAGACGTAATATCTACCGTATGGTGCAGAAATATACTGGTGACCAA
AGCGGAAATAAGAAAAGGTTTGCTACGTTTGTCTATGCAAAGCAGTCAGTAGATAC
TGGCGAGCTAGAAAGTGTAGCAACAGGAGGAAGTAGTCTTTACACATAAATGTGTG
TG**GCTAGCGGCCAACTTAACAAATACGGACG**ATGGTGAGCAAGGGCGAGGAGCTGT
TCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAG
TTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAA
GTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCC
TGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGAC
TTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAA
GGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGG
TGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGG
CACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCA
GAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCG
TGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTG
CTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGA
GAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCG
GCATGGACGAGCTGTACAAGTAATAA**CCCGGGATATCACGCGT**GGTACCTTACTTA
ACAAAAACTTAACAAATACGGACG**ATGAAATGGCTGACTAGTTTTGGAAGAGCAGT**
TATTTCTTGTTATAAATCCCTACTATTAACTCAACTTAGAGTGTTAGATAGGTTAA
TTTTAGATCACGGACTACTACGCGTTTTAACGTGTAGTAGGCGCGTGCTTTTAGTT
CAATTAGATTTAGTTTATAGGTTGGCGTATACGCCCACCCAATCGCTGGCATGAAT
AATAGTAAAGATAATCCTTTTCGCGGAGCAATAGCAAGAAAAGCTCGAATTTATCT
GAGAGAAGGATTAGATTGTGTTTACTTTCTTAACAAAGCAGGACAAGCAGAGCCTT
GTCCCGCGTGTACCTCTCTAGTATTCCAAGGGAAAACTTGTGAGGAACACATACAT
AATAATAATCTTTTGTCATGGCAAGCGGTAAAGCAGCTGGAAAAACAGACGCCCCA
GCGCCAGTCATTAAACTAGGAGGACCAAAACCACCTAAAGTCGGTTCTTCTGGAAA
TGCATCTTGGTTTCAAGCAATAAAAGCCAAGAAGTTAAATACACCTCCGCCCAAGT
TTGAAGGTAGCGGTGTTCCTGATAACGAAAACATTAAGCCAAGCCAGCAACATGGA
TACTGGAGACGCCAAGCCAGGTTTAAGCCAGGCAAAGGTGGAAGAAAACCAGTCCC
AGATGCTTGGTACTTTTACTATACTGGAACAGGACCTGCCGCTGACCTGAACTGGG
GTGATACTCAAGATGGTATAGTGTGGGTTGCTGCTAAGGGTGCTGATACTAAATCT
AGA

3c-M-IR-**Restriction Sites and TRS-eGFP-Restriction Sites**-IR-5a-5b-N

BeauR-hRlucΔIR

↓24406

CATATGGTAGAAAACTTAACAATCCGGAATTAGAAGCAGTTATTGTTAACGAGTTT
CCTAAGAACGGTTGGAATAATAAAAATCCAGCAAATTTTCAAGATGCCCAACGAGA
CAAATTGTACTCTTGACTTTGAACAGTCAGTTCAGCTTTTTAAAGAGTATAATTTA
TTTATAACTGCATTCTTGTTGTTCTTAACCATAATACTTCAGTATGGCTATGCAAC
AAGAAGTAAGGTTATTTATACACTGAAAATGATAGTGTTATGGTGCTTTTGGCCCC
TTAACATTGCAGTAGGTGTAATTTCATGTACATACCCACCAAACACAGGAGGTCTT
GTCGCAGCGATAATACTTACAGTGTTTGCGTGTCTGTCTTTTGTAGGTTATTGGAT
CCAGAGTATTAGACTCTTTAAGCGGTGTAGGTCATGGTGGTCATTTAATCCAGAAT
CTAATGCCGTAGGTTCAATACTCCTAACTAATGGTCAACAATGTAATTTTGCTATA
GAGAGTGTGCCAATGGTGCTTTCTCCAATTATAAAGAATGGTGTTCTTTATTGTGA
GGGTCAGTGGCTTGCTAAGTGTGAACCAGACCACTTGCCTAAAGATATATTTGTTT
GTACACCGGATAGACGTAATATCTACCGTATGGTGCAGAAATATACTGGTGACCAA
AGCGGAAATAAGAAAAGGTTTGCTACGTTTGTCTATGCAAAGCAGTCAGTAGATAC
TGGCGAGCTAGAAAGTGTAGCAACAGGAGGAAGTAGTCTTTACACATAAATGTGTG
TG**GCTAGCGGCCAACTTAACAAATACGGACG**ATGGCTTCCAAGGTGTACGACCCCG
AGCAACGCAAACGCATGATCACTGGGCCTCAGTGGTGGGCTCGCTGCAAGCAAATG
AACGTGCTGGACTCCTTCATCAACTACTATGATTCCGAGAAGCACGCCGAGAACGC
CGTGATTTTTCTGCATGGTAACGCTGCCTCCAGCTACCTGTGGAGGCACGTCGTGC
CTCACATCGAGCCCGTGGCTAGATGCATCATCCCTGATCTGATCGGAATGGGTAAG
TCCGGCAAGAGCGGGAATGGCTCATATCGCCTCCTGGATCACTACAAGTACCTCAC
CGCTTGGTTCGAGCTGCTGAACCTTCCAAAGAAAATCATCTTTGTGGGCCACGACT
GGGGGGCTTGTCTGGCCTTTCACTACTCCTACGAGCACCAAGACAAGATCAAGGCC
ATCGTCCATGCTGAGAGTGTCGTGGACGTGATCGAGTCCTGGGACGAGTGGCCTGA
CATCGAGGAGGATATCGCCCTGATCAAGAGCGAAGAGGGCGAGAAAATGGTGCTTG
AGAATAACTTCTTCGTCGAGACCATGCTCCCAAGCAAGATCATGCGGAAACTGGAG
CCTGAGGAGTTCGCTGCCTACCTGGAGCCATTCAAGGAGAAGGGCGAGGTTAGACG
GCCTACCCTCTCCTGGCCTCGCGAGATCCCTCTCGTTAAGGGAGGCAAGCCCGACG
TCGTCCAGATTGTCCGCAACTACAACGCCTACCTTCGGGCCAGCGACGATCTGCCT
AAGATGTTCATCGAGTCCGACCCTGGGTTCTTTTCCAACGCTATTGTCGAGGGAGC
TAAGAAGTTCCCTAACACCGAGTTCGTGAAGGTGAAGGGCCTCCACTTCAGCCAGG
AGGACGCTCCAGATGAAATGGGTAAGTACATCAAGAGCTTCGTGGAGCGCGTGCTG
AAGAACGAGCAGTAATAA**CCCGGGATATCACGCGT**GGTACCTTACTTAACAAAAAC
TTAACAAATACGGACG**ATGAAATGGCTGACTAGTTTTGGAAGAGCAGTTATTTCTT**
GTTATAAATCCCTACTATTAACTCAACTTAGAGTGTTAGATAGGTTAATTTTAGAT
CACGGACTACTACGCGTTTTAACGTGTAGTAGGCGCGTGCTTTTAGTTCAATTAGA
TTTAGTTTATAGGTTGGCGTATACGCCCACCCAATCGCTGGCATGAATAATAGTAA
AGATAATCCTTTTCGCGGAGCAATAGCAAGAAAAGCTCGAATTTATCTGAGAGAAG
GATTAGATTGTGTTTACTTTCTTAACAAAGCAGGACAAGCAGAGCCTTGTCCCGCG
TGTACCTCTCTAGTATTCCAAGGGAAAACTTGTGAGGAACACATACATAATAATAA
TCTTTTGTCATGGCAAGCGGTAAAGCAGCTGGAAAAACAGACGCCCCAGCGCCAGT
CATTAAACTAGGAGGACCAAAACCACCTAAAGTCGGTTCTTCTGGAAATGCATCTT
GGTTTCAAGCAATAAAAGCCAAGAAGTTAAATACACCTCCGCCCAAGTTTGAAGGT
AGCGGTGTTCCTGATAACGAAAACATTAAGCCAAGCCAGCAACATGGATACTGGAG
ACGCCAAGCCAGGTTTAAGCCAGGCAAAGGTGGAAGAAAACCAGTCCCAGATGCTT
GGTACTTTTACTATACTGGAACAGGACCTGCCGCTGACCTGAACTGGGGTGATACT
CAAGATGGTATAGTGTGGGTTGCTGCTAAGGGTGCTGATACTAAATCTAGA

3c-M-IR-**Restriction Sites and TRS-** hRluc-**Restriction Sites**-IR-5a-5b-N

BeauR-IBVeGFPΔ5ab

*Asc*I ↓24979

GGCGCGCCGCTAAGTGTGAACCAGACCACTTGCCTAAAGATATATTTGTTTGTACA
CCGGATAGACGTAATATCTACCGTATGGTGCAGAAATATACTGGTGACCAAAGCGG
AAATAAGAAAAGGTTTGCTACGTTTGTCTATGCAAAGCAGTCAGTAGATACTGGCG
AGCTAGAAAGTGTAGCAACAGGAGGAAGTAGTCTTTACACATAAATGTGTGTGTGT
AGAGAGTATTTAAAATTATTCTTTAATAGCGCCTCTGTTTTAAGAGCGCATAAGAG
TATTTATTTTGAGGATACTAATATAAATCCTCTTTGTTTTATACTCTCCTTTCAAG
AGCTATTAACGGTGTTACCTTTCAAGTAGATAATGGAAAAGTCTACTACGAAGGAA
CACCAGTTTTCCAAAAAGGTTGTTGTAGGATGTGGTCCAATTATAAGAAAGAATAA
TTGAACCACCTACTACACTTATTTTTATAAGAGGTACCTTACTTAACAAAAACTTA
ACAAATACGGACG**ATGGTTTCTAAAGGTGAAGAACTTTTTACTGGTGTTGTTCCAA**
TTCTTGTTGAACTTGATGGTGATGTTAATGGTCATAAATTTTCTGTTTCTGGTGAA
GGTGAAGGTGATGCAACTTATGGTAAACTTACTCTTAAATTTATTTGTACTACTGG
TAAACTTCCAGTTCCATGGCCAACTCTTGTTACTACTCTTACTTATGGTGTTCAAT
GTTTTTCTAGATATCCAGATCATATGAAACAACATGATTTTTTTAAATCTGCAATG
CCAGAAGGTTATGTTCAAGAAAGAACTATTTTTTTTAAAGATGATGGTAATTATAA
AACTAGAGCAGAAGTTAAATTTGAAGGTGATACTCTTGTTAATAGAATTGAACTTA
AAGGTATTGATTTTAAAGAAGATGGTAATATTCTTGGTCATAAACTTGAATATAAT
TATAATTCTCATAATGTTTATATTATGGCAGATAAACAAAAAAATGGTATTAAAGT
TAATTTTAAAATTAGACATAATATTGAAGATGGTTCTGTTCAACTTGCAGATCATT
ATCAACAAAATACTCCAATTGGTGATGGTCCAGTTCTTCTTCCAGATAATCATTAT
CTTTCTACTCAATCTGCACTTTCTAAAGATCCAAATGAAAAAAGAGATCATATGGT
TCTTCTTGAATTTGTTACTGCAGCAGGTATTACTCTTGGTATGGATGAACTTTATA
AACCTTTTCGCGGAGCAATAGCAAGAAAAGCTCGAATTTATCTGAGAGAAGGATTA
GATTGTGTTTACTTTCTTAACAAAGCAGGACAAGCAGAGCCTTGTCCCGCGTGTAC
CTCTCTAGTATTCCAAGGGAAAACTTGTGAGGAACACATACATAATAATAATCTTT
TGTCATGGCAAGCGGTAAAGCAGCTGGAAAAACAGACGCCCCAGCGCCAGTCATTA
AACTAGGAGGACCAAAACCACCTAAAGTCGGTTCTTCTGGAAATGCATCTTGGTTT
CAAGCAATAAAAGCCAAGAAGTTAAATACACCTCCGCCCAAGTTTGAAGGTAGCGG
TGTTCCTGATAACGAAAACATTAAGCCAAGCCAGCAACATGGATACTGGAGACGCC
AAGCCAGGTTTAAGCCAGGCAAAGGTGGAAGAAAACCAGTCCCAGATGCTTGGTAC
TTTTACTATACTGGAACAGGACCTGCCGCTGACCTGAACTGGGGTGATACTCAAGA
TGGTATAGTGTGGGTTGCTGCTAAGGGTGCTGATACTAAATCTAGATCGAC**TTAAT**
TAA
*Pac*I

M-IR-**eGFP**-**5b**-**N**

BeauR-IBVhRlucΔ5ab

*Asc*I ↓24979

GGCGCGCCGCTAAGTGTGAACCAGACCACTTGCCTAAAGATATATTTGTTTGTACA
CCGGATAGACGTAATATCTACCGTATGGTGCAGAAATATACTGGTGACCAAAGCGG
AAATAAGAAAAGGTTTGCTACGTTTGTCTATGCAAAGCAGTCAGTAGATACTGGCG
AGCTAGAAAGTGTAGCAACAGGAGGAAGTAGTCTTTACACATAAATGTGTGTGTGT
AGAGAGTATTTAAAATTATTCTTTAATAGCGCCTCTGTTTTAAGAGCGCATAAGAG
TATTTATTTTGAGGATACTAATATAAATCCTCTTTGTTTTATACTCTCCTTTCAAG
AGCTATTAACGGTGTTACCTTTCAAGTAGATAATGGAAAAGTCTACTACGAAGGAA
CACCAGTTTTCCAAAAAGGTTGTTGTAGGATGTGGTCCAATTATAAGAAAGAATAA
TTGAACCACCTACTACACTTATTTTTATAAGAGGTACCTTACTTAACAAAAACTTA
ACAAATACGGACG**ATGGCATCTAAAGTTTATGATCCAGAACAAAGAAAAAGAATGA**
TTACTGGTCCACAATGGTGGGCAAGATGTAAACAAATGAATGTTCTTGATTCTTTT
ATTAATTATTATGATTCTGAAAAACATGCAGAAAATGCAGTTATTTTTCTTCATGG
TAATGCAGCATCTTCTTATCTTTGGAGACATGTTGTTCCACATATTGAACCAGTTG
CAAGATGTATTATTCCAGATCTTATTGGTATGGGTAAATCTGGTAAATCTGGTAAT
GGTTCTTATAGACTTCTTGATCATTATAAATATCTTACTGCATGGTTTGAACTTCT
TAATCTTCCAAAAAAAATTATTTTTGTTGGTCATGATTGGGGTGCATGTCTTGCAT
TTCATTATTCTTATGAACATCAAGATAAAATTAAAGCAATTGTTCATGCAGAATCT
GTTGTTGATGTTATTGAATCTTGGGATGAATGGCCAGATATTGAAGAAGATATTGC
ACTTATTAAATCTGAAGAAGGTGAAAAAATGGTTCTTGAAAATAATTTTTTTGTTG
AAACTATGCTTCCATCTAAAATTATGAGAAAACTTGAACCAGAAGAATTTGCAGCA
TATCTTGAACCATTTAAAGAAAAAGGTGAAGTTAGAAGACCAACTCTTTCTTGGCC
AAGAGAAATTCCACTTGTTAAAGGTGGTAAACCAGATGTTGTTCAAATTGTTAGAA
ATTATAATGCATATCTTAGAGCATCTGATGATCTTCCAAAAATGTTTATTGAATCT
GATCCAGGTTTTTTTTCTAATGCAATTGTTGAAGGTGCAAAAAAATTTCCAAATAC
TGAATTTGTTAAAGTTAAAGGTCTTCATTTTTCTCAAGAAGATGCACCAGATGAAA
TGGGTAAATATATTAAATCTTTTGTTGAAAGAGTTCTTAAAAATGAACAATAATAA
CCTTTTCGCGGAGCAATAGCAAGAAAAGCTCGAATTTATCTGAGAGAAGGATTAGA
TTGTGTTTACTTTCTTAACAAAGCAGGACAAGCAGAGCCTTGTCCCGCGTGTACCT
CTCTAGTATTCCAAGGGAAAACTTGTGAGGAACACATACATAATAATAATCTTTTG
TCATGGCAAGCGGTAAAGCAGCTGGAAAAACAGACGCCCCAGCGCCAGTCATTAAA
CTAGGAGGACCAAAACCACCTAAAGTCGGTTCTTCTGGAAATGCATCTTGGTTTCA
AGCAATAAAAGCCAAGAAGTTAAATACACCTCCGCCCAAGTTTGAAGGTAGCGGTG
TTCCTGATAACGAAAACATTAAGCCAAGCCAGCAACATGGATACTGGAGACGCCAA
GCCAGGTTTAAGCCAGGCAAAGGTGGAAGAAAACCAGTCCCAGATGCTTGGTACTT
TTACTATACTGGAACAGGACCTGCCGCTGACCTGAACTGGGGTGATACTCAAGATG
GTATAGTGTGGGTTGCTGCTAAGGGTGCTGATACTAAATCTAGATCGAC**TTAATTA**
A *Pac*I

M-IR-**hRluc**-**5b**-**N**

BeauR-eGFPΔ3ab

*Asc*I ↓23361

GGCGCGCCAGATGTAGTTACGCTTACTTCTTGTCAAGCAAATTATGTAAGTGTAAA
TAAGACCGTCATTACTACATTCGTAGACAATGATGATTTTGATTTTAATGACGAAT
TGTCAAAATGGTGGAATGATACTAAGCATGAGCTACCAGACTTTGACAAATTCAAT
TACACAGTACCTATACTTGACATTGATAGTGAAATTGATCGTATTCAAGGCGTTAT
ACAGGGTCTTAATGACTCTCTAATAGACCTTGAAAAACTTTCAATACTCAAAACTT
ATATTAAGTGGCCTTGGTATGTGTGGTTAGCCATAGCTTTTGCCACTATTATCTTC
ATCTTAATACTAGGATGGGTTTTCTTCATGACTGGTTGTTGTGGTTGTTGTTGTGG
ATGCTTTGGCATTATGCCTCTAATGAGTAAGTGTGGTAAGAAATCTTCTTATTACA
CGACTTTTGATAACGATGTGGTAACTGAACAATACAGACCTAAAAAGTCTGTTTGA
ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCT
GGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTG
CCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTA
CCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACG
TCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG
GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACA
ACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATC
CGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACAC
CCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGT
CCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTC
GTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAATAAGTTTT
ACTTAACAAATACGGACGATGATGAATTTATTGAATAAGTCGCTAGAGGAGAATGG
AAGTTTTCTAACAGCGCTTTACATAATTGTAGGATTTTTAGCACTTTATCTTCTAG
GTAGAGCACTTCAAGCATTTGTACAGGCTGCTGATGCTTGTTGTTTATTTTGGTAT
ACATGGGTAGTAATTCCAGGAGCTAAGGGTACAGCCTTTGTATACAAGTATACATA
TGGTAGAAAACTTAACAATCCGGAATTAGAAGCAGTTATTGTTAACGAGTTTCCTA
AGAACGGTTGGAATAATAAAAATCCAGCAAATTTTCAAG**ATGCCCAACGAGACAAA**
TTGTACTCTTGACTTTGAACAGTCAGTTCAGCTTTTTAAAGAGTATAATTTATTTA
TAACTGCATTCTTGTTGTTCTTAACCATAATACTTCAGTATGGCTATGCAACAAGA
AGTAAGGTTATTTATACACTGAAAATGATAGTGTTATGGTGCTTTTGGCCCCTTAA
CATTGCAGTAGGTGTAATTTCATGTACATAC**TTAATTAA**

*Pac*I

S-**eGFP**-TRS plus flanking nucleotides-**3C**-**M**

BeauR-hRlucΔ3ab

*Asc*I ↓23361

GGCGCGCCAGATGTAGTTACGCTTACTTCTTGTCAAGCAAATTATGTAAGTGTAAA
TAAGACCGTCATTACTACATTCGTAGACAATGATGATTTTGATTTTAATGACGAAT
TGTCAAAATGGTGGAATGATACTAAGCATGAGCTACCAGACTTTGACAAATTCAAT
TACACAGTACCTATACTTGACATTGATAGTGAAATTGATCGTATTCAAGGCGTTAT
ACAGGGTCTTAATGACTCTCTAATAGACCTTGAAAAACTTTCAATACTCAAAACTT
ATATTAAGTGGCCTTGGTATGTGTGGTTAGCCATAGCTTTTGCCACTATTATCTTC
ATCTTAATACTAGGATGGGTTTTCTTCATGACTGGTTGTTGTGGTTGTTGTTGTGG
ATGCTTTGGCATTATGCCTCTAATGAGTAAGTGTGGTAAGAAATCTTCTTATTACA
CGACTTTTGATAACGATGTGGTAACTGAACAATACAGACCTAAAAAGTCTGTTTGA
ATGGCTTCCAAGGTGTACGACCCCGAGCAACGCAAACGCATGATCACTGGGCCTCA
GTGGTGGGCTCGCTGCAAGCAAATGAACGTGCTGGACTCCTTCATCAACTACTATG
ATTCCGAGAAGCACGCCGAGAACGCCGTGATTTTTCTGCATGGTAACGCTGCCTCC
AGCTACCTGTGGAGGCACGTCGTGCCTCACATCGAGCCCGTGGCTAGATGCATCAT
CCCTGATCTGATCGGAATGGGTAAGTCCGGCAAGAGCGGGAATGGCTCATATCGCC
TCCTGGATCACTACAAGTACCTCACCGCTTGGTTCGAGCTGCTGAACCTTCCAAAG
AAAATCATCTTTGTGGGCCACGACTGGGGGGCTTGTCTGGCCTTTCACTACTCCTA
CGAGCACCAAGACAAGATCAAGGCCATCGTCCATGCTGAGAGTGTCGTGGACGTGA
TCGAGTCCTGGGACGAGTGGCCTGACATCGAGGAGGATATCGCCCTGATCAAGAGC
GAAGAGGGCGAGAAAATGGTGCTTGAGAATAACTTCTTCGTCGAGACCATGCTCCC
AAGCAAGATCATGCGGAAACTGGAGCCTGAGGAGTTCGCTGCCTACCTGGAGCCAT
TCAAGGAGAAGGGCGAGGTTAGACGGCCTACCCTCTCCTGGCCTCGCGAGATCCCT
CTCGTTAAGGGAGGCAAGCCCGACGTCGTCCAGATTGTCCGCAACTACAACGCCTA
CCTTCGGGCCAGCGACGATCTGCCTAAGATGTTCATCGAGTCCGACCCTGGGTTCT
TTTCCAACGCTATTGTCGAGGGAGCTAAGAAGTTCCCTAACACCGAGTTCGTGAAG
GTGAAGGGCCTCCACTTCAGCCAGGAGGACGCTCCAGATGAAATGGGTAAGTACAT
CAAGAGCTTCGTGGAGCGCGTGCTGAAGAACGAGCAGTAATAAGTTTTACTTAACA
AATACGGACG**ATG**ATGAATTTATTGAATAAGTCGCTAGAGGAGAATGGAAGTTTTC
TAACAGCGCTTTACATAATTGTAGGATTTTTAGCACTTTATCTTCTAGGTAGAGCA
CTTCAAGCATTTGTACAGGCTGCTGATGCTTGTTGTTTATTTTGGTATACATGGGT
AGTAATTCCAGGAGCTAAGGGTACAGCCTTTGTATACAAGTATACATATGGTAGAA
AACTTAACAA*TCCGGA*ATTAGAAGCAGTTATTGTTAACGAGTTTCCTAAGAACGGT
TGGAATAATAAAAATCCAGCAAATTTTCAAG**ATGCCCAACGAGACAAATTGTACTC**
TTGACTTTGAACAGTCAGTTCAGCTTTTTAAAGAGTATAATTTATTTATAACTGCA
TTCTTGTTGTTCTTAACCATAATACTTCAGTATGGCTATGCAACAAGAAGTAAGGT
TATTTATACACTGAAAATGATAGTGTTATGGTGCTTTTGGCCCCTTAACATTGCAG
TAGGTGTAATTTCATGTACATACTTAATTAA

*Pac*I

S-hRluc-TRS plus flanking nucleotides-3C-M

BeauR-5a/eGFP

*Asc*I ↓24979

GGCGCGCCGCTAAGTGTGAACCAGACCACTTGCCTAAAGATATATTTGTTTGTACA
CCGGATAGACGTAATATCTACCGTATGGTGCAGAAATATACTGGTGACCAAAGCGG
AAATAAGAAAAGGTTTGCTACGTTTGTCTATGCAAAGCAGTCAGTAGATACTGGCG
AGCTAGAAAGTGTAGCAACAGGAGGAAGTAGTCTTTACACATAAATGTGTGTGTGT
AGAGAGTATTTAAAATTATTCTTTAATAGCGCCTCTGTTTTAAGAGCGCATAAGAG
TATTTATTTTGAGGATACTAATATAAATCCTCTTTGTTTTATACTCTCCTTTCAAG
AGCTATTAACGGTGTTACCTTTCAAGTAGATAATGGAAAAGTCTACTACGAAGGAA
CACCAGTTTTCCAAAAAGGTTGTTGTAGGATGTGGTCCAATTATAAGAAAGAATAA
TTGAACCACCTACTACACTTATTTTTATAAGAGGTACCTTACTTAACAAAAACTTA
ACAAATACGGACG**ATGAAATGGCTGACTAGTTTTGGAAGAGCAGTTATTTCTTGTT**
ATAAATCCCTACTATTAACTCAACTTAGAGTGTTAGATAGGTTAATTTTAGATCAC
GGACTACTACGCGTTTTAACGTGTAGTAGGCGCGTGCTTTTAGTTCAATTAGATTT
AGTTTATAGGTTGGCGTATACGCCCACCCAATCGCTGGCAGTGAGCAAGGGCGAGG
AGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGAC
CCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGA
CCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAG
CACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTT
CTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACA
CCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATC
CTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGA
CAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACG
GCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
GTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCC
CAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCA
CTCTCGGCATGGACGAGCTGTACAAGTAATAAATGAATAATAGTAAAGATAATCCT
TTTCGCGGAGCAATAGCAAGAAAAGCTCGAATTTATCTGAGAGAAGGATTAGATTG
TGTTTACTTTCTTAACAAAGCAGGACAAGCAGAGCCTTGTCCCGCGTGTACCTCTC
TAGTATTCCAAGGGAAAACTTGTGAGGAACACATACATAATAATAATCTTTTGTCA
TGGCAAGCGGTAAAGCAGCTGGAAAAACAGACGCCCCAGCGCCAGTCATTAAACTA
GGAGGACCAAAACCACCTAAAGTCGGTTCTTCTGGAAATGCATCTTGGTTTCAAGC
AATAAAAGCCAAGAAGTTAAATACACCTCCGCCCAAGTTTGAAGGTAGCGGTGTTC
CTGATAACGAAAACATTAAGCCAAGCCAGCAACATGGATACTGGAGACGCCAAGCC
AGGTTTAAGCCAGGCAAAGGTGGAAGAAAACCAGTCCCAGATGCTTGGTACTTTTA
CTATACTGGAACAGGACCTGCCGCTGACCTGAACTGGGGTGATACTCAAGATGGTA
TAGTGTGGGTTGCTGCTAAGGGTGCTGATACTAAATCTAGATCGAC**TTAATTAA**

*Pac*I

M-IR-5a/**eGFP**-5b-N

BeauR-eGFPΔ4b

AscI ↓24406

GGCGCGCCCATATGGTAGAAAACTTAACAATCCGGAATTAGAAGCAGTTATTGTTA
ACGAGTTTCCTAAGAACGGTTGGAATAATAAAAATCCAGCAAATTTTCAAGATGCC
CAACGAGACAAATTGTACTCTTGACTTTGAACAGTCAGTTCAGCTTTTTAAAGAGT
ATAATTTATTTATAACTGCATTCTTGTTGTTCTTAACCATAATACTTCAGTATGGC
TATGCAACAAGAAGTAAGGTTATTTATACACTGAAAATGATAGTGTTATGGTGCTT
TTGGCCCCTTAACATTGCAGTAGGTGTAATTTCATGTACATACCCACCAAACACAG
GAGGTCTTGTCGCAGCGATAATACTTACAGTGTTTGCGTGTCTGTCTTTTGTAGGT
TATTGGATCCAGAGTATTAGACTCTTTAAGCGGTGTAGGTCATGGTGGTCATTTAA
TCCAGAATCTAATGCCGTAGGTTCAATACTCCTAACTAATGGTCAACAATGTAATT
TTGCTATAGAGAGTGTGCCAATGGTGCTTTCTCCAATTATAAAGAATGGTGTTCTT
TATTGTGAGGGTCAGTGGCTTGCTAAGTGTGAACCAGACCACTTGCCTAAAGATAT
ATTTGTTTGTACACCGGATAGACGTAATATCTACCGTATGGTGCAGAAATATACTG
GTGACCAAAGCGGAAATAAGAAAAGGTTTGCTACGTTTGTCTATGCAAAGCAGTCA
GTAGATACTGGCGAGCTAGAAAGTGTAGCAACAGGAGGAAGTAGTCTTTACACATA
AATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGC
TGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT
GCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGT
GCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCT
ACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTAC
GTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGA
GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACT
TCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCAC
AACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGAT
CCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACA
CCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAG
TCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTT
CGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAATAATTGA
ACCACCTACTACACTTATTTTTATAAGAGGTGTTTTACTTAACAAAAACTTAACAA
ATACGGACG**ATGAAATGGCTGACTAGTTTTGGAAGAGCAGTTATTTCTTGTTATAA**
ATCCCTACTATTAACTCAACTTAGAGTGTTAGATAGGTTAATTTTAGATCACGGAC
TACTACGCGTTTTAACGTGTAGTAGGCGCGTGCTTTTAGTTCAATTAGATTTAGTT
TATAGGTTGGCGTATACGCCCACCCAATCGCTGGCATGAATAATAGTAAAGATAAT
CCTTTTCGCGGAGCAATAGCAAGAAAAGCTCGAATTTATCTGAGAGAAGGATTAGA
TTGTGTTTACTTTCTTAACAAAGCAGGACAAGCAGAGCCTTGTCCCGCGTGTACCT
CTCTAGTATTCCAAGGGAAAACTTGTGAGGAACACATACATAATAATAATCTTTTG
TCATGGCAAGCGGTAAAGCAGCTGGAAAAACAGACGCCCCAGCGCCAGTCATTAAA
CTAGGAGGACCAAAACCACCTAAAGTCGGTTCTTCTGGAAATGCATCTTGGTTTCA
AGCAATAAAAGCCAAGAAGTTAAATACACCTCCGCCCAAGTTTGAAGGTAGCGGTG
TTCCTGATAACGAAAACATTAAGCCAAGCCAGCAACATGGATACTGGAGACGCCAA
GCCAGGTTTAAGCCAGGCAAAGGTGGAAGAAAACCAGTCCCAGATGCTTGGTACTT
TTACTATACTGGAACAGGACCTGCCGCTGACCTGAACTGGGGTGATACTCAAGATG
GTATAGTGTGGGTTGCTGCTAAGGGTGCTGATACTAAATCTAGA**GTCGAC**
SalI

3c-M-eGFP-IR-5a-5b-N

Appendix 4: Codon optimized reporter gene sequences

Codon Optimized eGFP

```
                10        20        30        40        50        60
           ....|....|....|....|....|....|....|....|....|....|....|....|
pEGFP-C1   ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGAC
Emboss     ATGGTTTCTAAAGGTGAAGAACTTTTTACTGGTGTTGTTCCAATTCTTGTTGAACTTGAT
Manual     ATGGTTTCTAAAGGTGAAGAACTTTTTACTGGTGTTGTTCCAATTCTTGTTGAACTTGAT
            M  V  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D

                70        80        90       100       110       120
           ....|....|....|....|....|....|....|....|....|....|....|....|
pEGFP-C1   GGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAC
Emboss     GGTGATGTTAATGGTCATAAATTTTCTGTTTCTGGTGAAGGTGAAGGTGATGCAACTTAT
Manual     GGTGATGTTAATGGTCATAAATTTTCTGTTTCTGGTGAAGGTGAAGGTGATGCAACTTAT
            G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y

               130       140       150       160       170       180
           ....|....|....|....|....|....|....|....|....|....|....|....|
pEGFP-C1   GGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACC
Emboss     GGTAAACTTACTCTTAAATTTATTTGTACTACTGGTAAACTTCCAGTTCCATGGCCAACT
Manual     GGTAAACTTACTCTTAAATTTATTTGTACTACTGGTAAACTTCCAGTTCCATGGCCAACT
            G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T

               190       200       210       220       230       240
           ....|....|....|....|....|....|....|....|....|....|....|....|
pEGFP-C1   CTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAG
Emboss     CTTGTTACTACTCTTACTTATGGTGTTCAATGTTTTTCTAGATATCCAGATCATATGAAA
Manual     CTTGTTACTACTCTTACTTATGGTGTTCAATGTTTTTCTAGATATCCAGATCATATGAAA
            L  V  T  T  L  T  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K

               250       260       270       280       290       300
           ....|....|....|....|....|....|....|....|....|....|....|....|
pEGFP-C1   CAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTC
Emboss     CAACATGATTTTTTTAAATCTGCAATGCCAGAAGGTTATGTTCAAGAAAGAACTATTTTT
Manual     CAACATGATTTTTTTAAATCTGCAATGCCAGAAGGTTATGTTCAAGAAAGAACTATTTTT
            Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F

               310       320       330       340       350       360
           ....|....|....|....|....|....|....|....|....|....|....|....|
pEGFP-C1   TTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTG
Emboss     TTTAAAGATGATGGTAATTATAAAACTAGAGCAGAAGTTAAATTTGAAGGTGATACTCTT
Manual     TTTAAAGATGATGGTAATTATAAAACTAGAGCAGAAGTTAAATTTGAAGGTGATACTCTT
            F  K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L

               370       380       390       400       410       420
           ....|....|....|....|....|....|....|....|....|....|....|....|
pEGFP-C1   GTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC
Emboss     GTTAATAGAATTGAACTTAAAGGTATTGATTTTAAAGAAGATGGTAATATTCTTGGTCAT
Manual     GTTAATAGAATTGAACTTAAAGGTATTGATTTTAAAGAAGATGGTAATATTCTTGGTCAT
            V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H

               430       440       450       460       470       480
           ....|....|....|....|....|....|....|....|....|....|....|....|
pEGFP-C1   AAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAAC
Emboss     AAACTTGAATATAATTATAATTCTCATAATGTTTATATTATGGCAGATAAACAAAAAAAT
Manual     AAACTTGAATATAATTATAATTCTCATAATGTTTATATTATGGCAGATAAACAAAAAAAT
            K  L  E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N

               490       500       510       520       530       540
           ....|....|....|....|....|....|....|....|....|....|....|....|
pEGFP-C1   GGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCC
Emboss     GGTATTAAAGTTAATTTTAAAATTAGACATAATATTGAAGATGGTTCTGTTCAACTTGCA
Manual     GGTATTAAAGTTAATTTTAAAATTAGACATAATATTGAAGATGGTTCTGTTCAACTTGCA
            G  I  K  V  N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A
```

```
                 550        560        570        580        590        600
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
pEGFP-C1  GACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCAC
Emboss    GATCATTATCAACAAAATACTCCAATTGGTGATGGTCCAGTTCTTCTTCCAGATAATCAT
Manual    GATCATTATCAACAAAATACTCCAATTGGTGATGGTCCAGTTCTTCTTCCAGATAATCAT
           D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H

                 610        620        630        640        650        660
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
pEGFP-C1  TACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTC
Emboss    TATCTTTCTACTCAATCTGCACTTTCTAAAGATCCAAATGAAAAAAGAGATCATATGGTT
Manual    TATCTTTCTACTCAATCTGCACTTTCTAAAGATCCAAATGAAAAAAGAGATCATATGGTT
           Y  L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V

                 670        680        690        700        710
          ....|....|....|....|....|....|....|....|....|....|....|..
pEGFP-C1  CTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAG
Emboss    CTTCTTGAATTTGTTACTGCAGCAGGTATTACTCTTGGTATGGATGAACTTTATAAA
Manual    CTTCTTGAATTTGTTACTGCAGCAGGTATTACTCTTGGTATGGATGAACTTTATAAA
           L  L  E  F  V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K
```

Codon Optimized hRluc

```
                  10         20         30         40         50         60
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
pGL4.75   ATGGCTTCCAAGGTGTACGACCCCGAGCAACGCAAACGCATGATCACTGGGCCTCAGTGG
Emboss    ATGGCATCTAAAGTTTATGATCCAGAACAAAGAAAAAGAATGATTACTGGTCCACAATGG
           M  A  S  K  V  Y  D  P  E  Q  R  K  R  M  I  T  G  P  Q  W

                  70         80         90         100        110        120
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
pGL4.75   TGGGCTCGCTGCAAGCAAATGAACGTGCTGGACTCCTTCATCAACTACTATGATTCCGAG
Emboss    TGGGCAAGATGTAAACAAATGAATGTTCTTGATTCTTTTATTAATTATTATGATTCTGAA
           W  A  R  C  K  Q  M  N  V  L  D  S  F  I  N  Y  Y  D  S  E

                 130        140        150        160        170        180
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
pGL4.75   AAGCACGCCGAGAACGCCGTGATTTTTCTGCATGGTAACGCTGCCTCCAGCTACCTGTGG
Emboss    AAACATGCAGAAAATGCAGTTATTTTTCTTCATGGTAATGCAGCATCTTCTTATCTTTGG
           K  H  A  E  N  A  V  I  F  L  H  G  N  A  A  S  S  Y  L  W

                 190        200        210        220        230        240
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
pGL4.75   AGGCACGTCGTGCCTCACATCGAGCCCGTGGCTAGATGCATCATCCCTGATCTGATCGGA
Emboss    AGACATGTTGTTCCACATATTGAACCAGTTGCAAGATGTATTATTCCAGATCTTATTGGT
           R  H  V  V  P  H  I  E  P  V  A  R  C  I  I  P  D  L  I  G

                 250        260        270        280        290        300
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
pGL4.75   ATGGGTAAGTCCGGCAAGAGCGGGAATGGCTCATATCGCCTCCTGGATCACTACAAGTAC
Emboss    ATGGGTAAATCTGGTAAATCTGGTAATGGTTCTTATAGACTTCTTGATCATTATAAATAT
           M  G  K  S  G  K  S  G  N  G  S  Y  R  L  L  D  H  Y  K  Y

                310        320        330        340        350        360
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
pGL4.75   CTCACCGCTTGGTTCGAGCTGCTGAACCTTCCAAAGAAAATCATCTTTGTGGGCCACGAC
Emboss    CTTACTGCATGGTTTGAACTTCTTAATCTTCCAAAAAAAATTATTTTTGTTGGTCATGAT
           L  T  A  W  F  E  L  L  N  L  P  K  K  I  I  F  V  G  H  D

                 370        380        390        400        410        420
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|
pGL4.75   TGGGGGGCTTGTCTGGCCTTTCACTACTCCTACGAGCACCAAGACAAGATCAAGGCCATC
Emboss    TGGGGTGCATGTCTTGCATTTCATTATTCTTATGAACATCAAGATAAAATTAAAGCAATT
           W  G  A  C  L  A  F  H  Y  S  Y  E  H  Q  D  K  I  K  A  I
```

```
                430       440       450       460       470       480
           ....|....|....|....|....|....|....|....|....|....|....|....|
pGL4.75    GTCCATGCTGAGAGTGTCGTGGACGTGATCGAGTCCTGGGACGAGTGGCCTGACATCGAG
Emboss     GTTCATGCAGAATCTGTTGTTGATGTTATTGAATCTTGGGATGAATGGCCAGATATTGAA
            V   H   A   E   S   V   V   D   V   I   E   S   W   D   E   W   P   D   I   E

                490       500       510       520       530       540
           ....|....|....|....|....|....|....|....|....|....|....|....|
pGL4.75    GAGGATATCGCCCTGATCAAGAGCGAAGAGGGCGAGAAAATGGTGCTTGAGAATAACTTC
Emboss     GAAGATATTGCACTTATTAAATCTGAAGAAGGTGAAAAAATGGTTCTTGAAAATAATTTT
            E   D   I   A   L   I   K   S   E   E   G   E   K   M   V   L   E   N   N   F

                550       560       570       580       590       600
           ....|....|....|....|....|....|....|....|....|....|....|....|
pGL4.75    TTCGTCGAGACCATGCTCCCAAGCAAGATCATGCGGAAACTGGAGCCTGAGGAGTTCGCT
Emboss     TTTGTTGAAACTATGCTTCCATCTAAAATTATGAGAAAACTTGAACCAGAAGAATTTGCA
            F   V   E   T   M   L   P   S   K   I   M   R   K   L   E   P   E   E   F   A

                610       620       630       640       650       660
           ....|....|....|....|....|....|....|....|....|....|....|....|
pGL4.75    GCCTACCTGGAGCCATTCAAGGAGAAGGGCGAGGTTAGACGGCCTACCCTCTCCTGGCCT
Emboss     GCATATCTTGAACCATTTAAAGAAAAAGGTGAAGTTAGAAGACCAACTCTTTCTTGGCCA
            A   Y   L   E   P   F   K   E   K   G   E   V   R   R   P   T   L   S   W   P

                670       680       690       700       710       720
           ....|....|....|....|....|....|....|....|....|....|....|....|
pGL4.75    CGCGAGATCCCTCTCGTTAAGGGAGGCAAGCCCGACGTCGTCCAGATTGTCCGCAACTAC
Emboss     AGAGAAATTCCACTTGTTAAAGGTGGTAAACCAGATGTTGTTCAAATTGTTAGAAATTAT
            R   E   I   P   L   V   K   G   G   K   P   D   V   V   Q   I   V   R   N   Y

                730       740       750       760       770       780
           ....|....|....|....|....|....|....|....|....|....|....|....|
pGL4.75    AACGCCTACCTTCGGGCCAGCGACGATCTGCCTAAGATGTTCATCGAGTCCGACCCTGGG
Emboss     AATGCATATCTTAGAGCATCTGATGATCTTCCAAAAATGTTTATTGAATCTGATCCAGGT
            N   A   Y   L   R   A   S   D   D   L   P   K   M   F   I   E   S   D   P   G

                790       800       810       820       830       840
           ....|....|....|....|....|....|....|....|....|....|....|....|
pGL4.75    TTCTTTTCCAACGCTATTGTCGAGGGAGCTAAGAAGTTCCCTAACACCGAGTTCGTGAAG
Emboss     TTTTTTTCTAATGCAATTGTTGAAGGTGCAAAAAAATTTCCAAATACTGAATTTGTTAAA
            F   F   S   N   A   I   V   E   G   A   K   K   F   P   N   T   E   F   V   K

                850       860       870       880       890       900
           ....|....|....|....|....|....|....|....|....|....|....|....|
pGL4.75    GTGAAGGGCCTCCACTTCAGCCAGGAGGACGCTCCAGATGAAATGGGTAAGTACATCAAG
Emboss     GTTAAAGGTCTTCATTTTTCTCAAGAAGATGCACCAGATGAAATGGGTAAATATATTAAA
            V   K   G   L   H   F   S   Q   E   D   A   P   D   E   M   G   K   Y   I   K

                910       920       930
           ....|....|....|....|....|...
pGL4.75    AGCTTCGTGGAGCGCGTGCTGAAGAACGAGCAG
Emboss     TCTTTTGTTGAAAGAGTTCTTAAAAATGAACAA
            S   F   V   E   R   V   L   K   N   E   Q
```

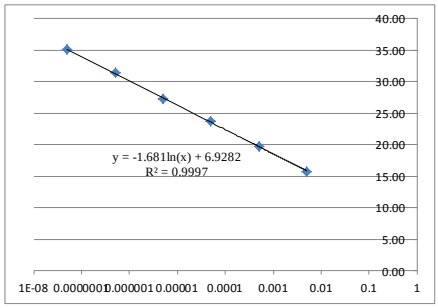
Appendix 5: Sample qRT-PCR data

Note to Appendix 5

This appendix shows sample data from one replicate of the qRT-PCR experiment to demonstrate the generation of standard curves, calculation of DNA amounts and normalisation of data to 28S housekeeping gene readings. The lower tables provide the values from each experiment used for the calculation of fold change values following additional normalisation to Gene 3 values to account for changes in the replication efficiencies of the rIBVs.

28S Data Replicate 1

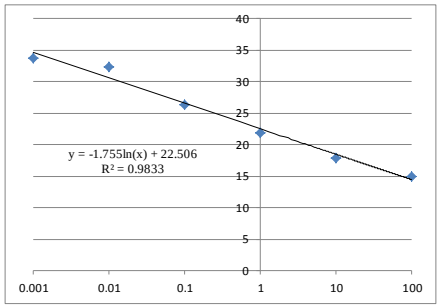
Sample		Ct Value			Average Ct Value
		1	2	3	
Standard Curve	0.5	13.75	13.78		13.77
	0.05	14.14	14.11		14.13
	0.005	15.76	15.76		15.76
	0.0005	19.75	19.65		19.70
	0.00005	23.73	23.84		23.79
	0.000005	27.3	27.3		27.30
	0.0000005	31.39	31.45		31.42
	0.00000005	35.02	35.23		35.13
Beau-R		18.03	18.09	18.03	18.05
BeauR-L-CTGAACAA		17.5	17.73	17.56	17.60
BeauR-L-CTTAACAT		17.88	17.78	17.7	17.79
BeauR-IR-CTGAACAA		17.51	17.45	17.5	17.49



Sample	Avg Ct Value	µg DNA in 10^{-4} cDNA	Total µg DNA
Beau-R	18.05	0.00133853	13.385333
BeauR-L-CTGAACAA	17.5966667	0.00175287	17.528713
BeauR-L-CTTAACAT	17.7866667	0.00156553	15.655344
BeauR-IR-CTGAACAA	17.4866667	0.00187141	18.714105

IR Gene Data Replicate 1

Sample		Ct Value			Average Ct Value
		1	2	3	
Standard Curve	100	14.91	14.97		14.94
	10	18.04	17.74		17.89
	1	21.97	21.94		21.955
	0.1	26.35	26.38		26.365
	0.01	32.25	32.4		32.325
	0.001	33.8	33.57		33.685
	0.0001	37.79	37.39		37.59
	0.00001	38.81	36.47		37.64
Beau-R		31.21	31.09	31.03	31.11
BeauR-L-CTGAACAA		29.21	29.16	29.04	29.1366667
BeauR-L-CTTAACAT		34.64	34.34	34.52	34.5
BeauR-IR-CTGAACAA		28.25	28.32	28.38	28.3166667



Sample	Avg Ct Value	pg DNA in 10^{-4} cDNA	Total pg DNA	Total µg DNA	Normalised to 28S
Beau-R	31.11	0.011324	113.240027	113240027	8460008.46
BeauR-L-CTGAACAA	29.1366667	0.03238476	323.847564	323847564	18475262.1
BeauR-L-CTTAACAT	34.5	0.0018623	18.6230094	18623009.4	1189562.47
BeauR-IR-CTGAACAA	28.3166667	0.05011511	501.151095	501151095	26779324.9

Fold Change Calculations

Replicate 1

Sample	Gene 3 Normalised to 28S	IR Normalised to 28S	IR Normalised to Gene 3	Fold Change to Beau-R
Beau-R	56573961.6	8460008.46	0.1495389	1
BeauR-L-CTGAACAA	118999569	18475262.1	0.1552549	1.03822387
BeauR-L-CTTAACAT	51003955.4	1189562.47	0.0233229	0.15596573
BeauR-IR-CTGAACAA	27581589	26779324.9	0.9709131	6.49271193

Replicate 2

Sample	Gene 3 Normalised to 28S	IR Normalised to 28S	IR Normalised to Gene 3	Fold Change to Beau-R
Beau-R	21629192	1701061.8	0.0786466	1
BeauR-L-CTGAACAA	74022863.4	6386331.47	0.0862751	1.09699786
BeauR-L-CTTAACAT	28318669.1	199630.101	0.0070494	0.08963413
BeauR-IR-CTGAACAA	15766390.1	6164373.24	0.3909819	4.9713789

References

Almazan, F., C. Galan, and L. Enjuanes. (2004). The nucleoprotein is required for efficient coronavirus genome replication. *J Virol.* **78**(22), 12683-12688.

Almazan, F., J. M. Gonzalez, Z. Penzes, A. Izeta, E. Calvo, J. Plana-Duran, and L. Enjuanes. (2000). Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc Natl Acad Sci U S A.* **97**(10), 5516-5521.

Alonso, S., A. Izeta, I. Sola, and L. Enjuanes. (2002). Transcription regulatory sequences and mRNA expression levels in the coronavirus transmissible gastroenteritis virus. *J Virol.* **76**(3), 1293-1308.

Armesto, M., D. Cavanagh, and P. Britton. (2009). The replicase gene of avian coronavirus infectious bronchitis virus is a determinant of pathogenicity. *PLoS ONE.* **4**(10), e7384.

Baliji, S., S. A. Cammer, B. Sobral, and S. C. Baker. (2009). Detection of nonstructural protein 6 in murine coronavirus-infected cells and analysis of the transmembrane topology by using bioinformatics and molecular approaches. *J Virol.* **83**(13), 6957-6962.

Baric, R. S., G. W. Nelson, J. O. Fleming, R. J. Deans, J. G. Keck, N. Casteel, and S. A. Stohlman. (1988). Interactions between coronavirus nucleocapsid protein and viral RNAs: implications for viral transcription. *J Virol.* **62**(11), 4280-4287.

Baric, R. S., S. A. Stohlman, and M. M. Lai. (1983). Characterization of replicative intermediate RNA of mouse hepatitis virus: presence of leader RNA sequences on nascent chains. *J Virol.* **48**(3), 633-640.

Baudoux, P., C. Carrat, L. Besnardeau, B. Charley, and H. Laude. (1998). Coronavirus pseudoparticles formed with recombinant M and E proteins induce alpha interferon synthesis by leukocytes. *J Virol.* **72**(11), 8636-8643.

Bhardwaj, K., L. Guarino, and C. C. Kao. (2004). The severe acute respiratory syndrome coronavirus Nsp15 protein is an endoribonuclease that prefers manganese as a cofactor. *J Virol.* **78**(22), 12218-12224.

Boltz, C. R., D. A. Boltz, D. Bunick, G. Scherba, and J. M. Bahr. (2007). Vaccination against the avian infectious bronchitis virus affects sperm concentration, sperm quality and blood testosterone concentrations in cockerels. *Br Poult Sci.* **48**(5), 617-624.

Boltz, D. A., M. Nakai, and J. M. Bahra. (2004). Avian infectious bronchitis virus: a possible cause of reduced fertility in the rooster. *Avian Dis.* **48**(4), 909-915.

Bosch, B. J., C. A. de Haan, and P. J. Rottier. (2004). Coronavirus spike glycoprotein, extended at the carboxy terminus with green fluorescent protein, is assembly competent. *J Virol.* **78**(14), 7369-7378.

Bosch, B. J., R. van der Zee, C. A. de Haan, and P. J. Rottier. (2003). The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J Virol.* **77**(16), 8801-8811.

Bouvet, M., I. Imbert, L. Subissi, L. Gluais, B. Canard, and E. Decroly. (2012). RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc Natl Acad Sci U S A.* **109**(24), 9372-9377.

Brierley, I., M. E. Boursnell, M. M. Binns, B. Bilimoria, V. C. Blok, T. D. Brown, and S. C. Inglis. (1987). An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J.* **6**(12), 3779-3785.

Britton, P., S. Evans, B. Dove, M. Davies, R. Casais, and D. Cavanagh. (2005). Generation of a recombinant avian coronavirus infectious bronchitis virus using transient dominant selection. *J Virol Methods.* **123**(2), 203-211.

Calvo, E., D. Escors, J. A. Lopez, J. M. Gonzalez, A. Alvarez, E. Arza, and L. Enjuanes. (2005). Phosphorylation and subcellular localization of transmissible gastroenteritis virus nucleocapsid protein in infected cells. *J Gen Virol.* **86**(Pt 8), 2255-2267.

Cao, J., C. C. Wu, and T. L. Lin. (2008). Complete nucleotide sequence of polyprotein gene 1 and genome organization of turkey coronavirus. *Virus Res.* **136**(1-2), 43-49.

Casais, R., M. Davies, D. Cavanagh, and P. Britton. (2005). Gene 5 of the avian coronavirus infectious bronchitis virus is not essential for replication. *J Virol.* **79**(13), 8065-8078.

Casais, R., V. Thiel, S. G. Siddell, D. Cavanagh, and P. Britton. (2001). Reverse genetics system for the avian coronavirus infectious bronchitis virus. *J Virol.* **75**(24), 12359-12369.

Cavanagh, D. (2007). Coronavirus avian infectious bronchitis virus. *Vet Res.* **38**(2), 281-297.

Cavanagh, D., P. J. Davis, J. H. Darbyshire, and R. W. Peters. (1986). Coronavirus IBV: virus retaining spike glycopolypeptide S2 but not S1 is unable to induce virus-neutralizing or haemagglutination-inhibiting antibody, or induce chicken tracheal protection. *J Gen Virol.* **67 (Pt 7)**, 1435-1442.

Cavanagh, D., M. M. Elus, and J. K. Cook. (1997). Relationship between sequence variation in the S1 spike protein of infectious bronchitis virus and the extent of cross-protection in vivo. *Avian Pathol.* **26**(1), 63-74.

Cervantes-Barragan, L., R. Zust, R. Maier, S. Sierro, J. Janda, F. Levy, D. Speiser, P. Romero, P. S. Rohrlich, B. Ludewig, and V. Thiel. (2010). Dendritic cell-specific antigen delivery by coronavirus vaccine vectors induces long-lasting protective antiviral and antitumor immunity. *MBio.* **1**(4), e00171.

Chang, C. K., S. C. Sue, T. H. Yu, C. M. Hsieh, C. K. Tsai, Y. C. Chiang, S. J. Lee, H. H. Hsiao, W. J. Wu, W. L. Chang, C. H. Lin, and T. H. Huang. (2006). Modular organization of SARS coronavirus nucleocapsid protein. *J Biomed Sci.* **13**(1), 59-72.

Chen, Y., H. Cai, J. Pan, N. Xiang, P. Tien, T. Ahola, and D. Guo. (2009). Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proc Natl Acad Sci U S A.* **106**(9), 3484-3489.

Chu, V. C., L. J. McElroy, V. Chu, B. E. Bauman, and G. R. Whittaker. (2006). The avian coronavirus infectious bronchitis virus undergoes direct low-pH-dependent fusion activation during entry into host cells. *J Virol.* **80**(7), 3180-3188.

Corse, E., and C. E. Machamer. (2003). The cytoplasmic tails of infectious bronchitis virus E and M proteins mediate their interaction. *Virology.* **312**(1), 25-34.

Corse, E., and C. E. Machamer. (2000). Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles. *J Virol.* **74**(9), 4319-4326.

Crinion, R. A. (1972). Egg quality and production following infectious bronchitis virus exposure at one day old. *Poult Sci.* **51**(2), 582-585.

Cruz, J. L., I. Sola, M. Becares, B. Alberca, J. Plana, L. Enjuanes, and S. Zuniga. (2011). Coronavirus gene 7 counteracts host defenses and modulates virus virulence. *PLoS Pathog.* **7**(6), e1002090.

Curtis, K. M., B. Yount, and R. S. Baric. (2002). Heterologous gene expression from transmissible gastroenteritis virus replicon particles. *J Virol.* **76**(3), 1422-1434.

de Haan, C. A., B. J. Haijema, D. Boss, F. W. Heuts, and P. J. Rottier. (2005). Coronaviruses as vectors: stability of foreign gene expression. *J Virol.* **79**(20), 12742-12751.

de Haan, C. A., L. Kuo, P. S. Masters, H. Vennema, and P. J. Rottier. (1998). Coronavirus particle assembly: primary structure requirements of the membrane protein. *J Virol.* **72**(8), 6838-6850.

de Haan, C. A., P. S. Masters, X. Shen, S. Weiss, and P. J. Rottier. (2002). The group-specific murine coronavirus genes are not essential, but their deletion, by reverse genetics, is attenuating in the natural host. *Virology.* **296**(1), 177-189.

de Haan, C. A., L. van Genne, J. N. Stoop, H. Volders, and P. J. Rottier. (2003). Coronaviruses as vectors: position dependence of foreign gene expression. *J Virol.* **77**(21), 11312-11323.

de Haan, C. A., H. Vennema, and P. J. Rottier. (2000). Assembly of the coronavirus envelope: homotypic interactions between the M proteins. *J Virol.* **74**(11), 4967-4978.

de Vries, A. A., A. L. Glaser, M. J. Raamsman, and P. J. Rottier. (2001). Recombinant equine arteritis virus as an expression vector. *Virology.* **284**(2), 259-276.

Decroly, E., I. Imbert, B. Coutard, M. Bouvet, B. Selisko, K. Alvarez, A. E. Gorbalenya, E. J. Snijder, and B. Canard. (2008). Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2'O)-methyltransferase activity. *J Virol.* **82**(16), 8071-8084.

DeDiego, M. L., E. Alvarez, F. Almazan, M. T. Rejas, E. Lamirande, A. Roberts, W. J. Shieh, S. R. Zaki, K. Subbarao, and L. Enjuanes. (2007). A severe acute respiratory syndrome coronavirus that lacks the E gene is attenuated in vitro and in vivo. *J Virol.* **81**(4), 1701-1713.

Delmas, B., and H. Laude. (1990). Assembly of coronavirus spike protein into trimers and its role in epitope expression. *J Virol.* **64**(11), 5367-5375.

Denison, M. R., R. L. Graham, E. F. Donaldson, L. D. Eckerle, and R. S. Baric. (2011). Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol.* **8**(2), 270-279.

DiNapoli, J. M., A. Kotelkin, L. Yang, S. Elankumaran, B. R. Murphy, S. K. Samal, P. L. Collins, and A. Bukreyev. (2007). Newcastle disease virus, a host range-restricted virus, as a vaccine vector for intranasal immunization against emerging pathogens. *Proc Natl Acad Sci U S A.* **104**(23), 9788-9793.

Donaldson, E. F., B. Yount, A. C. Sims, S. Burkett, R. J. Pickles, and R. S. Baric. (2008). Systematic assembly of a full-length infectious clone of human coronavirus NL63. *J Virol.* **82**(23), 11948-11957.

Dove, B., D. Cavanagh, and P. Britton. (2004). Presence of an encephalomyocarditis virus internal ribosome entry site sequence in avian infectious bronchitis virus defective RNAs abolishes rescue by helper virus. *J Virol.* **78**(6), 2711-2721.

Draper, S. J., and J. L. Heeney. (2010). Viruses as vaccine vectors for infectious diseases and cancer. *Nat Rev Micro.* **8**(1), 62-73.

Dufour, D., P. A. Mateos-Gomez, L. Enjuanes, J. Gallego, and I. Sola. (2011). Structure and functional relevance of a transcription-regulating sequence involved in coronavirus discontinuous RNA synthesis. *J Virol.* **85**(10), 4963-4973.

Eriksson, K. K., D. Makia, R. Maier, B. Ludewig, and V. Thiel. (2006). Towards a coronavirus-based HIV multigene vaccine. *Clin Dev Immunol.* **13**(2-4), 353-360.

Falkner, F. G., and B. Moss. (1990). Transient dominant selection of recombinant vaccinia viruses. *J Virol.* **64**(6), 3108-3111.

Firth, A. E., J. C. Zevenhoven-Dobbe, N. M. Wills, Y. Y. Go, U. B. Balasuriya, J. F. Atkins, E. J. Snijder, and C. C. Posthuma. (2011). Discovery of a small

arterivirus gene that overlaps the GP5 coding sequence and is important for virus production. *J Gen Virol.* **92**(Pt 5), 1097-1106.

Fischer, F., C. F. Stegen, C. A. Koetzner, and P. S. Masters. (1997). Analysis of a recombinant mouse hepatitis virus expressing a foreign gene reveals a novel aspect of coronavirus transcription. *J Virol.* **71**(7), 5148-5160.

Freundt, E. C., L. Yu, C. S. Goldsmith, S. Welsh, A. Cheng, B. Yount, W. Liu, M. B. Frieman, U. J. Buchholz, G. R. Screaton, J. Lippincott-Schwartz, S. R. Zaki, X. N. Xu, R. S. Baric, K. Subbarao, and M. J. Lenardo. (2010). The open reading frame 3a protein of severe acute respiratory syndrome-associated coronavirus promotes membrane rearrangement and cell death. *J Virol.* **84**(2), 1097-1109.

Frieman, M., B. Yount, M. Heise, S. A. Kopecky-Bromberg, P. Palese, and R. S. Baric. (2007). Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane. *J Virol.* **81**(18), 9812-9824.

Frieman, M. B., B. Yount, A. C. Sims, D. J. Deming, T. E. Morrison, J. Sparks, M. Denison, M. Heise, and R. S. Baric. (2006). SARS coronavirus accessory ORFs encode luxury functions. *Adv Exp Med Biol.* **581**, 149-152.

Ge, J., X. Wang, L. Tao, Z. Wen, N. Feng, S. Yang, X. Xia, C. Yang, H. Chen, and Z. Bu. (2011). Newcastle disease virus-vectored rabies vaccine is safe, highly immunogenic, and provides long-lasting protection in dogs and cats. *J Virol.* **85**(16), 8241-8252.

Gomaa, M. H., J. R. Barta, D. Ojkic, and D. Yoo. (2008). Complete genomic sequence of turkey coronavirus. *Virus Res.* **135**(2), 237-246.

Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov. (1989). Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res.* **17**(12), 4847-4861.

Grossoehme, N. E., L. Li, S. C. Keane, P. Liu, C. E. Dann, 3rd, J. L. Leibowitz, and D. P. Giedroc. (2009). Coronavirus N protein N-terminal domain (NTD) specifically binds the transcriptional regulatory sequence (TRS) and melts TRS-cTRS RNA duplexes. *J Mol Biol.* **394**(3), 544-557.

Hahn, C. S., Y. S. Hahn, T. J. Braciale, and C. M. Rice. (1992). Infectious Sindbis virus transient expression vectors for studying antigen processing and presentation. *Proc Natl Acad Sci U S A.* **89**(7), 2679-2683.

Hewson, K. A., J. Ignjatovic, G. F. Browning, J. M. Devlin, and A. H. Noormohammadi. (2011). Infectious bronchitis viruses with naturally occurring genomic rearrangement and gene deletion. *Arch Virol.* **156**(2), 245-252.

Hodgson, T., P. Britton, and D. Cavanagh. (2006). Neither the RNA nor the proteins of open reading frames 3a and 3b of the coronavirus infectious bronchitis virus are essential for replication. *J Virol.* **80**(1), 296-305.

Huang, C., K. G. Lokugamage, J. M. Rozovics, K. Narayanan, B. L. Semler, and S. Makino. (2011). SARS coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs: viral mRNAs are resistant to nsp1-induced RNA cleavage. *PLoS Pathog.* **7**(12), e1002433.

Hughes, A. L. (2011). Recombinational histories of avian infectious bronchitis virus and turkey coronavirus. *Arch Virol.* **156**(10), 1823-1829.

Ignjatovic, J., and L. Galli. (1994). The S1 glycoprotein but not the N or M proteins of avian infectious bronchitis virus induces protection in vaccinated chickens. *Arch Virol.* **138**(1-2), 117-134.

Iqbal, M. (2012). Progress toward the development of polyvalent vaccination strategies against multiple viral infections in chickens using herpesvirus of turkeys as vector. *Bioengineered.* **3**(4), 222-226.

Ivanov, K. A., V. Thiel, J. C. Dobbe, Y. van der Meer, E. J. Snijder, and J. Ziebuhr. (2004). Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J Virol.* **78**(11), 5619-5632.

Johnson, C. R., T. F. Griggs, J. Gnanandarajah, and M. P. Murtaugh. (2011). Novel structural protein in porcine reproductive and respiratory syndrome virus encoded by an alternative ORF5 present in all arteriviruses. *J Gen Virol.* **92**(Pt 5), 1107-1116.

Kanjanahaluethai, A., Z. Chen, D. Jukneliene, and S. C. Baker. (2007). Membrane topology of murine coronavirus replicase nonstructural protein 3. *Virology.* **361**(2), 391-401.

Keane, S. C., P. Liu, J. L. Leibowitz, and D. P. Giedroc. (2012). Functional Transcriptional Regulatory Sequence (TRS) RNA Binding and Helix Destabilizing Determinants of Murine Hepatitis Virus (MHV) Nucleocapsid (N) Protein. *J Biol Chem.* **287**(10), 7063-7073.

Klumperman, J., J. K. Locker, A. Meijer, M. C. Horzinek, H. J. Geuze, and P. J. Rottier. (1994). Coronavirus M proteins accumulate in the Golgi complex beyond the site of virion budding. *J Virol.* **68**(10), 6523-6534.

Knoops, K., M. Kikkert, S. H. Worm, J. C. Zevenhoven-Dobbe, Y. van der Meer, A. J. Koster, A. M. Mommaas, and E. J. Snijder. (2008). SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS Biol.* **6**(9), e226.

Koetzner, C. A., L. Kuo, S. J. Goebel, A. B. Dean, M. M. Parker, and P. S. Masters. (2010). Accessory Protein 5a Is a Major Antagonist of the Antiviral Action of Interferon against Murine Coronavirus. *J Virol.* **84**(16), 8262-8274.

Kottier, S. A., D. Cavanagh, and P. Britton. (1995). Experimental evidence of recombination in coronavirus infectious bronchitis virus. *Virology.* **213**(2), 569-580.

Kuo, L., and P. S. Masters. (2003). The small envelope protein E is not essential for murine coronavirus replication. *J Virol.* **77**(8), 4597-4608.

Ladman, B. S., C. R. Pope, A. F. Ziegler, T. Swieczkowski, C. J. Callahan, S. Davison, and J. Gelb, Jr. (2002). Protection of chickens after live and inactivated virus vaccination against challenge with nephropathogenic infectious bronchitis virus PA/Wolgemuth/98. *Avian Dis.* **46**(4), 938-944.

Lai, M. M. C., and K. V. Holmes. 2001. *Coronaviridae*:The Viruses and Their Replication, p. 1163-1185. *In* D. M. Knipe and P. M. Howley (ed.), Fields Virology, 4 ed. Lippincott, Williams & Wilkins, Philadelphia.

Lau, S. K., P. Lee, A. K. Tsang, C. C. Yip, H. Tse, R. A. Lee, L. Y. So, Y. L. Lau, K. H. Chan, P. C. Woo, and K. Y. Yuen. (2011). Molecular epidemiology of human coronavirus OC43 reveals evolution of different genotypes over time and recent emergence of a novel genotype due to natural recombination. *J Virol.* **85**(21), 11325-11337.

Lee, C. W., and M. W. Jackwood. (2000). Evidence of genetic diversity generated by recombination among avian coronavirus IBV. *Arch Virol.* **145**(10), 2135-2148.

Lin, K. Y., H. C. Wang, and C. H. Wang. (2005). Protective effect of vaccination in chicks with local infectious bronchitis viruses against field virus challenge. *J Microbiol Immunol Infect.* **38**(1), 25-30.

Liu, D. X., D. Cavanagh, P. Green, and S. C. Inglis. (1991). A polycistronic mRNA specified by the coronavirus infectious bronchitis virus. *Virology.* **184**(2), 531-544.

Liu, D. X., and S. C. Inglis. (1992). Identification of two new polypeptides encoded by mRNA5 of the coronavirus infectious bronchitis virus. *Virology.* **186**(1), 342-347.

Liu, D. X., and S. C. Inglis. (1992). Internal entry of ribosomes on a tricistronic mRNA encoded by infectious bronchitis virus. *J Virol.* **66**(10), 6143-6154.

Liu, S., X. Zhang, Y. Wang, C. Li, Q. Liu, Z. Han, Q. Zhang, X. Kong, and G. Tong. (2009). Evaluation of the protection conferred by commercial vaccines and attenuated heterologous isolates in China against the CK/CH/LDL/97I strain of infectious bronchitis coronavirus. *Vet J.* **179**(1), 130-136.

Lorenz, W. W., R. O. McCann, M. Longiaru, and M. J. Cormier. (1991). Isolation and expression of a cDNA encoding Renilla reniformis luciferase. *Proc Natl Acad Sci U S A.* **88**(10), 4438-4442.

Makino, S., N. Fujioka, and K. Fujiwara. (1985). Structure of the intracellular defective viral RNAs of defective interfering particles of mouse hepatitis virus. *J Virol.* **54**(2), 329-336.

Makino, S., J. G. Keck, S. A. Stohlman, and M. M. Lai. (1986). High-frequency RNA recombination of murine coronaviruses. *J Virol.* **57**(3), 729-737.

Mateos-Gomez, P. A., S. Zuniga, L. Palacio, L. Enjuanes, and I. Sola. (2011). Gene N proximal and distal RNA motifs regulate coronavirus nucleocapsid mRNA transcription. *J Virol.* **85**(17), 8968-8980.

Miknis, Z. J., E. F. Donaldson, T. C. Umland, R. A. Rimmer, R. S. Baric, and L. W. Schultz. (2009). Severe acute respiratory syndrome coronavirus nsp9 dimerization is essential for efficient viral growth. *J Virol.* **83**(7), 3007-3018.

Moreno, J. L., S. Zuniga, L. Enjuanes, and I. Sola. (2008). Identification of a coronavirus transcription enhancer. *J Virol.* **82**(8), 3882-3893.

Nakaya, T., J. Cros, M. S. Park, Y. Nakaya, H. Zheng, A. Sagrera, E. Villar, A. Garcia-Sastre, and P. Palese. (2001). Recombinant Newcastle disease virus as a vaccine vector. *J Virol.* **75**(23), 11868-11873.

Narayanan, K., A. Maeda, J. Maeda, and S. Makino. (2000). Characterization of the coronavirus M protein and nucleocapsid interaction in infected cells. *J Virol.* **74**(17), 8127-8134.

Nelson, G. W., S. A. Stohlman, and S. M. Tahara. (2000). High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA. *J Gen Virol.* **81**(Pt 1), 181-188.

Oostra, M., E. G. te Lintelo, M. Deijs, M. H. Verheije, P. J. Rottier, and C. A. de Haan. (2007). Localization and membrane topology of coronavirus nonstructural protein 4: involvement of the early secretory pathway in replication. *J Virol.* **81**(22), 12323-12336.

Ortego, J., I. Sola, F. Almazan, J. E. Ceriani, C. Riquelme, M. Balasch, J. Plana, and L. Enjuanes. (2003). Transmissible gastroenteritis coronavirus gene 7 is not essential but influences in vivo virus replication and virulence. *Virology.* **308**(1), 13-22.

Ozdarendeli, A., S. Ku, S. Rochat, G. D. Williams, S. D. Senanayake, and D. A. Brian. (2001). Downstream sequences influence the choice between a naturally occurring noncanonical and closely positioned upstream canonical heptameric fusion motif during bovine coronavirus subgenomic mRNA synthesis. *J Virol.* **75**(16), 7362-7374.

Pasternak, A. O., E. van den Born, W. J. Spaan, and E. J. Snijder. (2001). Sequence requirements for RNA strand transfer during nidovirus discontinuous subgenomic RNA synthesis. *EMBO J.* **20**(24), 7220-7228.

Pasternak, A. O., E. van den Born, W. J. Spaan, and E. J. Snijder. (2003). The stability of the duplex between sense and antisense transcription-regulating sequences is a crucial factor in arterivirus subgenomic mRNA synthesis. *J Virol.* **77**(2), 1175-1183.

Penzes, Z., K. Tibbles, K. Shaw, P. Britton, T. D. Brown, and D. Cavanagh. (1994). Characterization of a replicating and packaged defective RNA of avian coronavirus infectious bronchitis virus. *Virology.* **203**(2), 286-293.

Prasher, D. C., V. K. Eckenrode, W. W. Ward, F. G. Prendergast, and M. J. Cormier. (1992). Primary structure of the Aequorea victoria green-fluorescent protein. *Gene.* **111**(2), 229-233.

Pugachev, K. V., P. W. Mason, R. E. Shope, and T. K. Frey. (1995). Double-subgenomic Sindbis virus recombinants expressing immunogenic proteins of Japanese encephalitis virus induce significant protection in mice against lethal JEV infection. *Virology.* **212**(2), 587-594.

Pugachev, K. V., W. P. Tzeng, and T. K. Frey. (2000). Development of a rubella virus vaccine expression vector: use of a picornavirus internal ribosome entry site increases stability of expression. *J Virol.* **74**(22), 10811-10815.

Putics, A., W. Filipowicz, J. Hall, A. E. Gorbalenya, and J. Ziebuhr. (2005). ADP-ribose-1"-monophosphatase: a conserved coronavirus enzyme that is dispensable for viral replication in tissue culture. *J Virol.* **79**(20), 12721-12731.

Raaben, M., Prins, H. J., Martens, A. C., Rottier, P. J., de Haan, C. A. (2009). Non-invasive imaging of mouse hepatitis coronavirus infection reveals determinants of viral replication and spread *in vivo*. *Cellular Microbiology.* **11**(5), 825-841.

Raamsman, M. J., J. K. Locker, A. de Hooge, A. A. de Vries, G. Griffiths, H. Vennema, and P. J. Rottier. (2000). Characterization of the coronavirus mouse hepatitis virus strain A59 small membrane protein E. *J Virol.* **74**(5), 2333-2342.

Raj, G. D., and R. C. Jones. (1997). Infectious bronchitis virus: Immunopathogenesis of infection in the chicken. *Avian Pathol.* **26**(4), 677-706.

Rayner, J. O., S. A. Dryga, and K. I. Kamrud. (2002). Alphavirus vectors and vaccination. *Rev Med Virol.* **12**(5), 279-296.

Ribes, J. M., J. Ortego, J. Ceriani, R. Montava, L. Enjuanes, and J. Buesa. (2011). Transmissible gastroenteritis virus (TGEV)-based vectors with engineered murine tropism express the rotavirus VP7 protein and immunize mice against rotavirus. *Virology.* **410**(1), 107-118.

Rota, P. A., M. S. Oberste, S. S. Monroe, W. A. Nix, R. Campagnoli, J. P. Icenogle, S. Penaranda, B. Bankamp, K. Maher, M. H. Chen, S. Tong, A. Tamin, L. Lowe, M. Frace, J. L. DeRisi, Q. Chen, D. Wang, D. D. Erdman, T. C. Peret, C. Burns, T. G. Ksiazek, P. E. Rollin, A. Sanchez, S. Liffick, B. Holloway, J. Limor, K. McCaustland, M. Olsen-Rasmussen, R. Fouchier, S. Gunther, A. D. Osterhaus, C. Drosten, M. A. Pallansch, L. J. Anderson, and W. J. Bellini. (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science.* **300**(5624), 1394-1399.

Sawicki, S. G., and D. L. Sawicki. (1995). Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands. *Adv Exp Med Biol.* **380**, 499-506.

Schaad, M. C., and R. S. Baric. (1994). Genetics of mouse hepatitis virus transcription: evidence that subgenomic negative strands are functional templates. *J Virol.* **68**(12), 8169-8179.

Shen, H., S. G. Fang, B. Chen, G. Chen, F. P. Tay, and D. X. Liu. (2009). Towards construction of viral vectors based on avian coronavirus infectious bronchitis virus for gene delivery and vaccine development. *J Virol Methods.* **160**(1-2), 48-56.

Shi, X. M., Y. Zhao, H. B. Gao, Z. Jing, M. Wang, H. Y. Cui, G. Z. Tong, and Y. F. Wang. (2011). Evaluation of recombinant fowlpox virus expressing infectious bronchitis virus S1 gene and chicken interferon-gamma gene for immune protection against heterologous strains. *Vaccine.* **29**(8), 1576-1582.

Shimomura, O., F. H. Johnson, and Y. Saiga. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *J Cell Comp Physiol.* **59**, 223-239.

Skinner, M. A., S. M. Laidlaw, I. Eldaghayes, P. Kaiser, and M. G. Cottingham. (2005). Fowlpox virus as a recombinant vaccine vector for use in mammals and poultry. *Expert Rev Vaccines.* **4**(1), 63-76.

Sola, I., S. Alonso, S. Zuniga, M. Balasch, J. Plana-Duran, and L. Enjuanes. (2003). Engineering the transmissible gastroenteritis virus genome as an expression vector inducing lactogenic immunity. *J Virol.* **77**(7), 4357-4369.

Sola, I., C. Galan, P. A. Mateos-Gomez, L. Palacio, S. Zuniga, J. L. Cruz, F. Almazan, and L. Enjuanes. (2011). The Polypyrimidine Tract-Binding Protein Affects Coronavirus RNA Accumulation Levels and Relocalizes Viral RNAs to Novel Cytoplasmic Domains Different from Replication-Transcription Sites. *J Virol.* **85**(10), 5136-5149.

Sola, I., P. A. Mateos-Gomez, F. Almazan, S. Zuniga, and L. Enjuanes. (2011). RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA Biol.* **8**(2), 237-248.

Sola, I., J. L. Moreno, S. Zuniga, S. Alonso, and L. Enjuanes. (2005). Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis. *J Virol.* **79**(4), 2506-2516.

Stadler, K., V. Masignani, M. Eickmann, S. Becker, S. Abrignani, H.-D. Klenk, and R. Rappuoli. (2003). SARS - beginning to understand a new virus. *Nat Rev Micro.* **1**(3), 209-218.

Stern, D. F., L. Burgess, and B. M. Sefton. (1982). Structural analysis of virion proteins of the avian coronavirus infectious bronchitis virus. *J Virol.* **42**(1), 208-219.

Stern, D. F., and S. I. Kennedy. (1980). Coronavirus multiplication strategy. I. Identification and characterization of virus-specified RNA. *J Virol.* **34**(3), 665-674.

Stirrups, K., K. Shaw, S. Evans, K. Dalton, R. Casais, D. Cavanagh, and P. Britton. (2000). Expression of reporter genes from the defective RNA CD-61 of the coronavirus infectious bronchitis virus. *J Gen Virol.* **81**(Pt 7), 1687-1698.

Sutton, G., E. Fry, L. Carter, S. Sainsbury, T. Walter, J. Nettleship, N. Berrow, R. Owens, R. Gilbert, A. Davidson, S. Siddell, L. L. Poon, J. Diprose, D. Alderton, M. Walsh, J. M. Grimes, and D. I. Stuart. (2004). The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure.* **12**(2), 341-353.

Tanaka, T., W. Kamitani, M. L. Dediego, L. Enjuanes, and Y. Matsuura. (2012). SARS coronavirus nsp1 facilitates an efficient propagation in cells through a specific translational shutoff of host mRNA. *J Virol.* doi.10.1128/JVI.01700-12

te Velthuis, A. J., S. H. van den Worm, and E. J. Snijder. (2012). The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic Acids Res.* **40**(4), 1737-1747.

Tekes, G., R. Hofmann-Lehmann, I. Stallkamp, V. Thiel, and H. J. Thiel. (2008). Genome organization and reverse genetic analysis of a type I feline coronavirus. *J Virol.* **82**(4), 1851-1859.

Thiel, V., J. Herold, B. Schelle, and S. G. Siddell. (2001). Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J Gen Virol.* **82**(Pt 6), 1273-1281.

Thiel, V., N. Karl, B. Schelle, P. Disterer, I. Klagge, and S. G. Siddell. (2003). Multigene RNA vector based on coronavirus transcription. *J Virol.* **77**(18), 9790-9798.

Tooze, J., S. A. Tooze, and S. D. Fuller. (1987). Sorting of progeny coronavirus from condensed secretory proteins at the exit from the trans-Golgi network of AtT20 cells. *J Cell Biol.* **105**(3), 1215-1226.

Tsang, K. Y., C. Palena, J. Yokokawa, P. M. Arlen, J. L. Gulley, G. P. Mazzara, L. Gritz, A. G. Yafal, S. Ogueta, P. Greenhalgh, K. Manson, D. Panicali, and J. Schlom. (2005). Analyses of recombinant vaccinia and fowlpox vaccine vectors expressing transgenes for two human tumor antigens and three human costimulatory molecules. *Clin Cancer Res.* **11**(4), 1597-1607.

Ulasli, M., M. H. Verheije, C. A. de Haan, and F. Reggiori. (2010). Qualitative and quantitative ultrastructural analysis of the membrane rearrangements induced by coronavirus. *Cell Microbiol.* **12**(6), 844-861.

van den Born, E., C. C. Posthuma, K. Knoops, and E. J. Snijder. (2007). An infectious recombinant equine arteritis virus expressing green fluorescent protein from its replicase gene. *J Gen Virol.* **88**(Pt 4), 1196-1205.

van den Worm, S. H., K. K. Eriksson, J. C. Zevenhoven, F. Weber, R. Zust, T. Kuri, R. Dijkman, G. Chang, S. G. Siddell, E. J. Snijder, V. Thiel, and A. D. Davidson. (2012). Reverse Genetics of SARS-Related Coronavirus Using Vaccinia Virus-Based Recombination. *PLoS ONE.* **7**(3), e32857.

van der Most, R. G., L. Heijnen, W. J. Spaan, and R. J. de Groot. (1992). Homologous RNA recombination allows efficient introduction of site-specific mutations into the genome of coronavirus MHV-A59 via synthetic co-replicating RNAs. *Nucleic Acids Res.* **20**(13), 3375-3381.

van Marle, G., J. C. Dobbe, A. P. Gultyaev, W. Luytjes, W. J. Spaan, and E. J. Snijder. (1999). Arterivirus discontinuous mRNA transcription is guided by base pairing between sense and antisense transcription-regulating sequences. *Proc Natl Acad Sci U S A.* **96**(21), 12056-12061.

Varshney, B., S. Agnihotram, Y. J. Tan, R. Baric, and S. K. Lal. (2012). SARS Coronavirus 3b Accessory Protein Modulates Transcriptional Activity of RUNX1b. *PLoS ONE.* **7**(1), e29542.

Weli, S. C., and M. Tryland. (2011). Avipoxviruses: infection biology and their use as vaccine vectors. *Virol J.* **8**, 49.

Wilson, L., P. Gage, and G. Ewart. (2006). Hexamethylene amiloride blocks E protein ion channels and inhibits coronavirus replication. *Virology.* **353**(2), 294-306.

Winter, C., G. Herrler, and U. Neumann. (2008). Infection of the tracheal epithelium by infectious bronchitis virus is sialic acid dependent. *Microbes Infect.* **10**(4), 367-373.

Winter, C., C. Schwegmann-Wessels, D. Cavanagh, U. Neumann, and G. Herrler. (2006). Sialic acid is a receptor determinant for infection of cells by avian Infectious bronchitis virus. *J Gen Virol.* **87**(Pt 5), 1209-1216.

Woo, J., M. H. Howell, and A. G. von Arnim. (2008). Structure-function studies on the active site of the coelenterazine-dependent luciferase from Renilla. *Protein Sci.* **17**(4), 725-735.

Xiao, Y., Q. Ma, T. Restle, W. Shang, D. I. Svergun, R. Ponnusamy, G. Sczakiel, and R. Hilgenfeld. (2012). Nonstructural proteins 7 and 8 of feline coronavirus form a 2:1 heterotrimer that exhibits primer-independent RNA polymerase activity. *J Virol.* **86**(8), 4444-4454.

Yamada, Y., X. B. Liu, S. G. Fang, F. P. Tay, and D. X. Liu. (2009). Acquisition of cell-cell fusion activity by amino acid substitutions in spike protein determines the infectivity of a coronavirus in cultured cells. *PLoS ONE.* **4**(7), e6130.

Yang, D., P. Liu, D. P. Giedroc, and J. Leibowitz. (2011). Mouse hepatitis virus stem-loop 4 functions as a spacer element required to drive subgenomic RNA synthesis. *J Virol.* **85**(17), 9199-9209.

Youn, S., J. L. Leibowitz, and E. W. Collisson. (2005). In vitro assembled, recombinant infectious bronchitis viruses demonstrate that the 5a open reading frame is not essential for replication. *Virology.* **332**(1), 206-215.

Yount, B., K. M. Curtis, and R. S. Baric. (2000). Strategy for systematic assembly of large RNA and DNA genomes: transmissible gastroenteritis virus model. *J Virol.* **74**(22), 10600-10611.

Yount, B., K. M. Curtis, E. A. Fritz, L. E. Hensley, P. B. Jahrling, E. Prentice, M. R. Denison, T. W. Geisbert, and R. S. Baric. (2003). Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. *Proc Natl Acad Sci U S A.* **100**(22), 12995-13000.

Yount, B., R. S. Roberts, L. Lindesmith, and R. S. Baric. (2006). Rewiring the severe acute respiratory syndrome coronavirus (SARS-CoV) transcription circuit: engineering a recombination-resistant genome. *Proc Natl Acad Sci U S A.* **103**(33), 12546-12551.

Yount, B., R. S. Roberts, A. C. Sims, D. Deming, M. B. Frieman, J. Sparks, M. R. Denison, N. Davis, and R. S. Baric. (2005). Severe acute respiratory syndrome coronavirus group-specific open reading frames encode nonessential functions for replication in cell cultures and mice. *J Virol.* **79**(23), 14909-14922.

Zhang, G., V. Gurtu, and S. R. Kain. (1996). An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. *Biochem Biophys Res Commun.* **227**(3), 707-711.

Zhang, X., and M. M. Lai. (1995). Interactions between the cytoplasmic proteins and the intergenic (promoter) sequence of mouse hepatitis virus RNA: correlation with the amounts of subgenomic mRNA transcribed. *J Virol.* **69**(3), 1637-1644.

Zhuang, Y., B. Butler, E. Hawkins, A. Paguio, L. Orr, M. G. Wood, and K. V. Wood. (2001). New Synthetic *Renilla* Gene and Assay System Increase Expression, Reliability and Sensitivity. *Promega Notes.* **79**, 6-11.

Zolotukhin, S., M. Potter, W. W. Hauswirth, J. Guy, and N. Muzyczka. (1996). A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells. *J Virol.* **70**(7), 4646-4654.

Zuniga, S., J. L. Cruz, I. Sola, P. A. Mateos-Gomez, L. Palacio, and L. Enjuanes. (2010). Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription. *J Virol.* **84**(4), 2169-2175.

Zuniga, S., I. Sola, S. Alonso, and L. Enjuanes. (2004). Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *J Virol.* **78**(2), 980-994.