THE UNIVERSITY OF READING

Coronavirus secondary RNA structures as regulators of the virus life cycle and virus pathogenicity

A thesis submitted for the degree of Doctor of Philosophy in

Microbiology / Virology

By

SAAD T. MUTLK

Supervisors:

Prof. Ian M. Jones

Dr. Benjamin W. Neuman

School of Biological Sciences

The University of Reading, UK

February 2018

Declaration

I confirm that this is my own work and that it has not been previously submitted for any dissertation or PhD degree at any university or institution. The use of all materials from other sources has been properly and fully acknowledged. All drawings not cited are original artwork by the author.

Saad T. MUTLK

Abstract

Human coronaviruses are large enveloped viruses with a single stranded, positive-sense RNA genome, the largest described among all RNA viruses. Infection gives rise to a number of pathologies ranging from mild upper respiratory infections (the common cold) to acute respiratory syndromes, with documented evidence of additional involvement of the gut and renal systems. Human coronavirus HKU1, classified as a group 2 coronavirus, was first isolated in Hong Kong in 2003 and generally causes mild infections. Although the structure of the virus and its genome are typical of coronaviruses there has been limited experimental confirmation of some of the key features of the virus life cycle and its pathogenicity. Coronavirus genomes contain extensive secondary structure elements, for example at the 5' and 3' extremities as well as internally at the junctions between open reading frames but their role in HKU1 has not been documented. This project aims to investigate the RNA secondary structures of human HKU1 coronavirus to determine how these structures affect virus transcription and translation.

Bioinformatics analysis identified potential RNA secondary structures in the 5'-UTR, 3'-UTR and ORF1a/1b ribosomal frameshift regions of HKU1. Subsequently, the identified 5'-UTR, 3'-UTR and frameshift sequences were constructed *de novo* and introduced into appropriate vectors in order to transcribe RNA *in vitro* and *in vivo*. Cloned HKU1 sequences were then flanked by two reporter genes, mCherry and eGFP, in the same vectors to provide an easily assayed readout of the function of the cloned segments and mutants thereof.

In vitro and *in vivo* analysis using a cell free coupled transcription-and-

translation system and expression in insect cells respectively revealed protein expression that was dependent on the sequence appended to the reporter. In particular, the results showed that conserved stem loops in the 5'-UTR of HKU1 stimulated translation while those present in the 3'-UTR had no effect. Time course experiments suggested a more efficient initiation of translation consistent with the presence of an internal ribosome binding site (IRES) and mutational analysis using 12 structure designed nucleotide changes highlighted critical residues, notably those indicated by mutant 10, which essentially abolished the stimulatory activity shown by the parental sequence.

Analysis of the 1ab junction sequence of HKU1, predicted to be a ribosomal frameshift sequence, using similar genetic constructs, confirmed for the first time a *bona fide* frameshift function in HKU1. Further, the rate of shift was formally measured as approximatively 25%, that is about 1 in 4 messenger RNAs encoding the 1ab junction led to translation of the downstream sequence while the majority of transcripts ceased translation within the 1ab region. The deduced rate is similar to that reported for other coronaviruses confirming that the experimental set-up provides an accurate model for measuring HKU1 folded RNA function. A limited mutational analysis of the folded structure abolished the frameshift, mapping some of the critical nucleotides concerned.

Together the experimental evidence obtained provides the first formal demonstration of activity for folded RNA structures found in the genome of HKU1 and indicated some of the critical residues involved. Further study focussing on these areas might include antisense or other small molecule antiviral therapy targeting the 5'-UTR and 1abFS and apply these methods on other human coronaviruses.

Dedication

This thesis is dedicated to my father's memory, Taha, who always wanted us to be the best and get the highest degree and to my mother Fatima for her endless love, support and encouragement through all my life. I am faithfully thankful for having them in my life. As well, my wife Nawal, my son Muataz and my new baby girl Malk. This work is also dedicated to my brothers and sister, Yassin, Daher, Khalid and Wazeri, who stood by me when things looked bleak and to all the people in my life who touched my heart.

Acknowledgments

It is very difficult to sufficiently express my sincere appreciation to my supervisors Prof. Ian Jones and Dr. Ben Neuman for their constant support, advice and valuable scientific discussions. Their guidance helped me a lot during the time of my research. I could not imagine how the research could have been done if I did not have such advisors and mentors for my PhD study. I specially thank Prof. Ian Jones for all the correction and assistance that he offered starting from gifting the plasmid vector to guidelines during writing of this thesis. Exceptional thanks to Drs. Silvia Loureiro and Sinéad Lyons who always kindly advised and helped me. In addition, all the staff in the Knight building for their kindness and help, with special thanks to Mrs Jessica del Rio and Mrs Jane Clark.

In addition, I would like to thank Ban Abdulsattar for her help and support, with very special thanks to Mohammed M. Dakheel and his family who always stood side by side with me even in difficult situations, also to the wonderful family: Prof. Qusai, Isaa Bader and their lovely kids for supporting me and encouraging me with their best wishes. Very special gratitude to **Sam**, **Sarah** and all the Oasis members for admitting me and my family to their houses and living there for the time I like even if we are from different religious. I cannot imagine how I'm happy during all that time until today. My sincere gratitude to my sponsor MOHESR in Iraq, Iraqi cultural attaché in London and University of Al-Anbar; I recognize that this research would not have been possible without their financial support. Finally, I would like to thank my parents, uncles, brothers, sister, my wife and lovely kids. They were always supporting me and encouraging me with their best wishes.

And.... to all those I have not mentioned, **thanks a lot again**.

Presentations

1- Mfolding web server for RNA secondary structures predication of the 5'-UTR, 3'-UTR and ORF1a/1b ribosomal frameshift regions in HCOV-HKU1. The annual symposium at School of Biological Sciences University of Reading 2015. Presentation.

2- Identifying potential RNA secondary structures in the 5'-UTR, 3'-UTR and ORF1a/1b ribosomal frameshift regions in HCOV-HKU1. The annual symposium at School of Biological Sciences University of Reading 2016. Presentation

3- Conserved RNA signals of Human coronavirus HKU1 5'end involved in translation. Doctoral Research Conference, University of Reading 2017. Presented poster.

Table of contents

List of Figures

List of Tables

Abbreviations

Ab Antibody

Ag Antigen

Amp Ampicillin

ASOs Antisense oligonucleotides

BSA Bovine serum albumin

CoV Coronavirus

Cryo-EM Electron cryomicroscopy

ddiH_2O Distilled water

DMSO Dimethyl sulfoxide

dNTP Deoxy ribonucleotide triphosphate

DPP4 dipeptidyl peptidase 4

dTTP Deoxy thymidine triphosphate

ECL Enhanced chemiluminescence

EDTA Ethylene Diamine Tetra Acetic acid

EM Electron microscopy

EMSA Electro mobility shift assay

PCR Polymerase chain reaction

FBS Fetal bovine serum

Fs Frameshift

Fw Forward primer

eGFP enhanced Green Florescent Protein

HAE human airway epithelial cell

HACE2 Human viruses angiotensin converting enzyme 2

HCoV-HKU1 Human coronavirus- Hong Kong University strain1

HRP Horse reddish peroxidase

IBV Infectious Bronchitis Virus

IDT Integrated DNA Technology

JEV Japanese encephalitis virus

LB Luria- Bertani

LNA locked nucleic acid

MERS Middle Eastern respiratory syndrome

MHV Mouse Hepatitis Virus

M protein Membrane protein

N N protein

NMR Nuclear magnetic resonance

nsp Non structural protein

NTP Nucleotides (A, G, C and U) triphosphate

OD Optical density

oligo Oligonucleotides

ORF Open Reading Frame

PBS Phosphate Buffer Saline

Plpro Papain-like proteinases

PMOs Phosphonodiamidite morpholino oligomers

PNA peptide nucleic acid

PVDF Polyvinylidene fluoride

RdRp RNA dependent RNA polymerase

RLP Ribosome landing pad

RNP Ribonucleoprotein complex

rpm Revolutions per minute

Rv Reverse primer

SARS Sever Acute Respiratory Syndrome

SDS-PAGE Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

SOC Super Optimal Broth

ssDNA Single stranded deoxyribonucleic acid

ssRNA Single stranded ribonucleic acid

Taq-Pol Thermus aquaticus polymerase

TBST Tris-buffered saline Tween20

TGEV Transmissible gastroenteritis virus

TEM transmission electron microscopy

TRS Transcriptional regulating sequence

UTR Untranslated region

VLPs Virus-like particles

WT Wild type

WNV West Nile virus

YFV Yellow fever virus

1.1. General introduction and characteristics of Coronaviruses.

1.1.1. Coronaviruses, taxonomy and defining characteristics.

Coronaviruses are classified under the Coronavirinae subfamily along with the Torovirinae which make up the Coronaviridae family. The Coronaviridae, Roniviridae, Mesoniviridae and Arteriviridae are members of the order *Nidovirales* (Gorbalenya *et al*., 2006, Hulo *et al.*, 2011). This order name derives from "*nidus"* which means nest in Latin, due to the synthesis of a nested set of subgenomic RNAs during the virus replication cycle. The Coronavirinae are classified into four genera, Alphacoronavirus, Betacoronavirus, Gamacoronavirus and Deltacoronavirus. Whereas, the Torovirinae include Bafinivirus and Torovirus **Figure 1** (Hulo *et al.*, 2011).

Structurally, coronaviruses are spherical or pleomorphic enveloped viruses that have been measured as being approximately 120 – 160 nm in diameter by negative stain transmission electron microscopy (Davies and Macnaughton, 1979), or 50-150 nm by cryo-electron microscopy (Neuman *et al.*, 2006b). Virions enclose a helical nucleocapsid surrounded by the envelope (Davies and Macnaughton, 1979, Barcena *et al.*, 2009); characteristic peplomers, 'spikes' which are club-shaped glycoprotein projections, 12–24 nm long, extend from the outer surface of the envelope and resemble the solar corona or crown from which the name coronavirus was derived (Neuman *et al.*, 2006b). The genome is a single stranded, non-segmented, positive-sense RNA typically of molecular weight approximately 6–7 $\times 10^6$ Da. Proteins associated with the virion include (at least) the spike glycoprotein (S), the integral membrane glycoprotein (M) and the nucleocapsid phosphoprotein (N) (Sharmin and Islam, 2014).

Subsets of CoV (BCV, HEV, Turkey CV) possess a third glycoprotein on the virion surface, the hemagglutinin-esterase dimer (HE). These proteins can be identified and quantified using pooled monoclonal antibodies to distinct epitopes found on each protein by ELISA. Nucleoprotein molecules (N) and the positive-stranded genome form the ribonucleoprotein core (RNP) **Figure 2** (Weiss and Navas-Martin, 2005). As it is expressed at such high levels, the nucleoprotein is a source of antigen for the detection of CoV infection and the presence of N is also critical for the rescue of viable coronavirus RNA following transfection in the process of reverse genetics (Grossoehme *et al.*, 2009, Casais *et al.*, 2001).

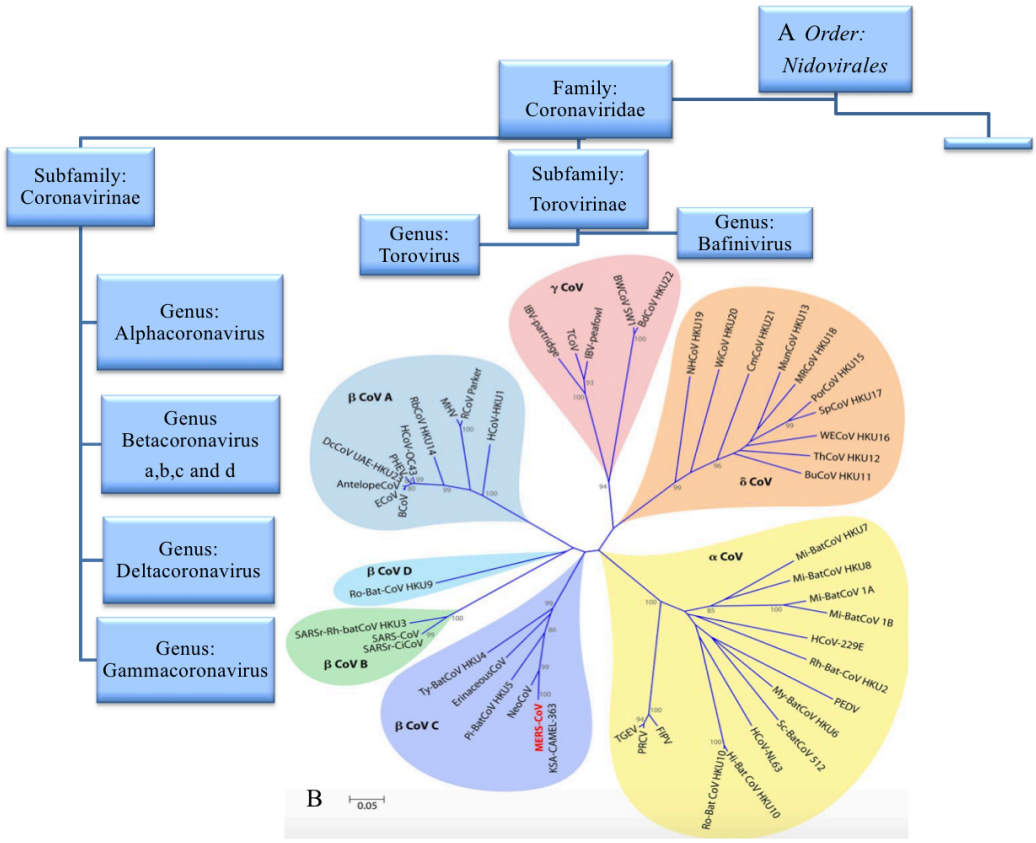

Figure 1. Coronaviridae Taxonomy and phylogenetic tree of 50 coronaviruses. (A) Coronaviridae Taxonomy according to the International Committee on Viruses Taxonomy (Adapted from (Hulo *et al.*, 2011)). (B) Phylogenetic tree of 50 coronaviruses with partial nucleotide sequences of RNA-dependent RNA polymerase, MEGA 5.0 was used for tree construction according to the neighbour-joining method. The scale bar indicates the estimated number of substitutions per 20 nucleotides (Adapted from (Chan JFW *et al.*, 2015)).

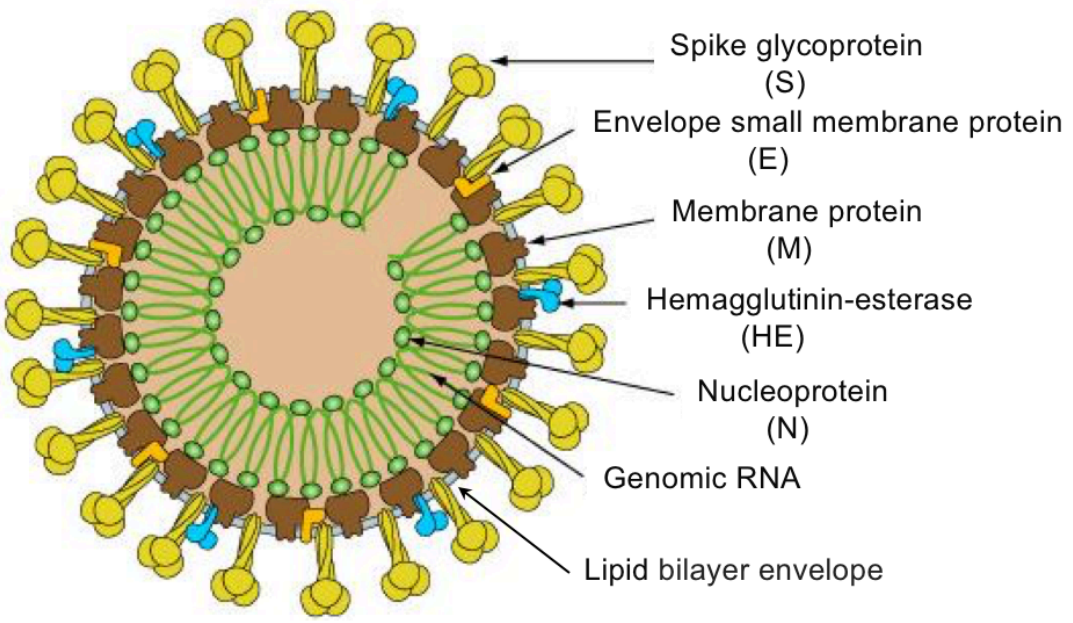

Figure 2. Schematic diagram of Coronavirus using Murine Hepatitis Virus (MHV). Coloured indicates spike (S; yellow), envelope (E; orange), membrane (M; brawn), nucleocapsid (N; green), accessory protein hemagglutinin esterase (HE; blue) and spiral genomic RNA (green) (Adapted from (Hulo *et al.*, 2011)).

Coronaviruses have the largest genomes among known RNA viruses, typically around 30 kb (van der Hoek *et al.*, 2004). A leader RNA is present in both the genomic and subgenomic RNA followed in the 5' two-thirds of the genome by the 1a and 1b genes which encode the non-structural polyproteins. The 3' third of the genome includes the genes encoding the four structural proteins: spike (S), envelope (E), membrane (M) and nucleocapsid (N). It also contains a variable number of unique characteristic open reading frames (ORFs) encoding non-structural proteins situated either between the 1b and S genes, between the S and E genes, between the M and N genes, or downstream of the N gene **Figure 3** (Hulo *et al.*, 2011).

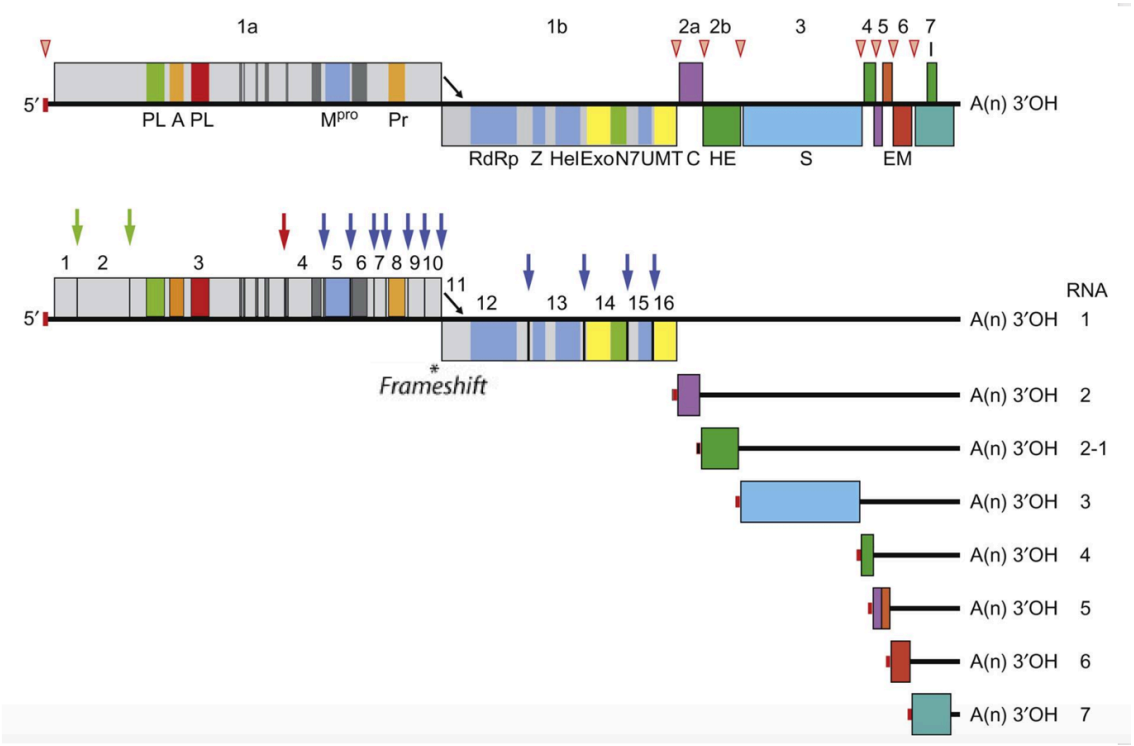

Figure 3. Coronavirus genomic structure and gene expression by using Murine Hepatitis Virus (MHV). The upper panel is a schematic representation of the genome of mouse hepatitis virus (MHV). Open reading frames (ORFs) are represented by boxes, indicated by number (above) and encoded protein (acronyms below). Regions encoding key domains in replicase polyproteins pp1a and pp1ab are color-coded with hydrophobic segments shown in dark grey. The 5′ leader sequence is depicted by a small red box. The arrow between ORF 1a and 1b represents the ribosomal frameshifting site. The poly (A) tail is indicated by "A(n)." Red arrowheads indicate the locations of transcription-regulating sequences (TSRs). PL (green) papain-like proteinase 1 (PL1pro); PL (red), papain-like proteinase 2 (PL2pro); A, ADP-reibose-1" phosphatase (macrodomain); Mpro, 3C-like main protease; Pr, noncanonical RNA-dependent RNA polymerase, putative primase; RdRp, RNA-dependent RNA polymerase; Z, zinc-binding domain; Hel, helicase domain; Exo, 3′–5′ exoribonuclease domain; N7, guanine-N7-methyltransferase; U, nidoviral uridylate-specific endoribonuclease (NendoU); MT, ribose-2′-O-methyltransferase domain; HE, hemagglutinin-esterase; S, spike protein; E, envelope protein; M, membrane protein; N, nucleocapsid protein; I, internal ORF. (Lower panel) Processing of the replicase polyproteins and structural relationship between the genomic RNA and subgenomic mRNAs of coronaviruses. Arrows indicate cleavage sites for PL1pro (green), PL2pro (red) and Mpro (blue). The locations of the nonstructural proteins (nsp's) are indicated by their number. mRNA species are numbered as by convention on the basis of their size, from large to small, with the genome designated as RNA1. For the sg mRNAs only ORFs that are translated are shown (Adapted from (King *et al.*, 2012)).

1.1.2. Coronaviruses pathogenicity.

Coronaviruses have the ability to cause infection in mammals and birds depending on the virus species and strain. The diseases caused range from mild to severe, for example infections such as avian infectious bronchitis virus is acute (IBV; type species) whereas most human coronavirus infections are generally only mild upper respiratory infections (part of the range of viruses that cause the common cold). Others such as HKU1, severe acute respiratory syndrome (SARS), murine hepatitis virus (MHV), porcine haemagglutinating encephalitis virus, and porcine transmissible gastroenteritis virus also cause a range of pathologies. Toroviruses includes Breda virus and Berne virus cause enteric infections in cattle and horses. The use of lipid solvents and detergents inactivates all these viruses through dissolution of their membranes (David Cavanagh 1990, Weiss and Navas-Martin, 2005).

1.1.3. Human Coronaviruses.

There are six coronaviruses currently infecting humans: HCoVs 229E and OC43 are both one of the causal agents of the common cold (McIntosh *et al.*, 1974) but there is some long held speculation about the association of these HCoVs with more serious human diseases such as multiple sclerosis (Burks *et al.*, 1980), hepatitis (Zuckerman *et al.*, 1970), or enteric disease in newborns (Resta *et al.*, 1985). However, it was SARS-CoV (Severe Acute Respiratory Syndrome), identified in March 2003 during the SARS epidemic and shown to cause a severe acute respiratory syndrome which was the first example of a serious illness in humans caused by a coronavirus (Rota *et al.*, 2003).

Since these isolations there have been reports of two more new human coronaviruses associated with respiratory disease. The HCoV HKUI was isolated from an elderly patient with pneumonia (Woo *et al.*, 2005b) but the virus has been difficult to recover in cell culture. In addition, HCoV NL63 was isolated from a 7-month-old child in The Netherlands who was suffering from bronchiolitis and conjunctivitis (Fouchier *et al.*, 2004). HCoV-NL63 has now been found to be associated with serious respiratory symptoms, including upper respiratory infection, bronchiolitis, and pneumonia (Ebihara *et al.*, 2005). The strong correlation of the presence of NL63 with croup in children with lower respiratory infections has also suggested a causal relationship between the virus and croup (van der Hoek *et al.*, 2005).

While primarily associated with infections of children, NL63 has also been detected in immunocompromised adults with respiratory tract infections. HCoV-NH virus was independently isolated in New Haven (Esper *et al.*, 2005) but appears to be the same as HCoV-HKU1. The most serious current CoV is Middle East Respiratory Syndrome coronavirus (MERS-CoV), a newly identified human pathogen associated with severe acute respiratory disease and occasionally accompanied by renal failure. MERS-CoV has claimed 693 lives (35.5%) among the **1,952** laboratory-confirmed cases reported to the World Health Organization as of 20 May 2017.

Although most cases of MERS originated in the Middle East, it can be transmitted from human-to-human, particularly during hospital outbreaks, and the virus has emerged as a threat to public health worldwide (Memish *et al.*, 2013). An outbreak in South Korea in 2015 killed 36 people before coming under control.

As described for the family in general, HCoVs are characterised by their characteristic spike protein resemblance to the solar corona or crown, cytoplasmic replication, budding into the ER and Golgi complex, linear, non-segmented, positive-sense, ssRNA genome and a genome length of approximately 27.6-31kb (Neuman *et al.*, 2006a).

The development of Human coronaviruses occurs through a number of features: there is a quasi-species structure to the viral population which results in multiple mutants being present during virus replication and the high frequency of mutation results in a genome flexibility which underpins the ability to jump the barrier species and to adapt to a new environment. The distribution of the different HCoV species differs according to the geographic area and season and the routine detection of these viruses is difficult due to a requirement for specialised facilities and expertise. This mainly depends on molecular assays such as RT-PCR and even if successfully diagnosed there is no established specific therapy for any CoV infection to date (Vabret *et al.*, 2001).

1.1.4. Human coronavirus HKU1.

In January 2005, a novel human coronavirus was reported and isolated from a Chinese patient with pneumonia. The 71 years old patient was admitted to hospital in January 2004 due to symptoms of high temperature and a productive cough with purulent sputum for at least 2 days. The patient had a previous history of pulmonary tuberculosis more than 40 years ago and he had fibrous tissue in the right upper lung lobe and bronchiectasis with chronic *Pseudomonas aeruginosa* colonization of the airways.

In addition, he was a chronic smoker and had chronic obstructive airway disease (COPD), hyperlipidaemia, and asymptomatic abdominal aortic aneurysm. His admission to hospital occurred 3 days after his return from a visit to Shenzhen and China. The radiograph result for his chest showed patchy infiltrates over the left lower zone (Woo *et al.*, 2005b, Wan-Ji Lee *et al.*, 2013).

A nasopharyngeal aspirate was taken in order to perform a direct antigen detection for common respiratory viruses which include influenza A virus, human metapneumovirus and SARS-CoV along with viral cultures. When the virus was determined to be a coronavirus, the nasopharyngeal aspirates were inoculated into human rhabdomyosarcoma (RD), murine macrophage (I13.35), murine fibroblast (L929), colorectal adenocarcinoma (HRT-18), marmoset B-lymblastoid (B95a) cell lines and mixed neuron-glia culture. The observation for cytopathic effect was negative. Furthermore, suckling mice stayed healthy 14 days after intracerebral inoculation. Sputum was negative for bacterial and mycobacterial pathogens and tests with Mycoplasma, Chlamydia, Legionella, and SARS-CoV sera were also negative.

The patient was discharged from hospital after 5 days following improvement in his symptoms. However, the culture supernatants and cell lysates were used for quantitative RT-PCR using conserved primers to the CoV polymerase and a complete novel coronavirus genome sequence was recovered. The negative results for culturing HCoV-HKU1 in different cell line indicated the limitation for recovery in cell cultures and this can be associated with the lack of a susceptible cell line for HCoV-HKU1 and the inherently low recovery rate of some coronaviruses.

Many decades after the identification of HCoV-229E and HCoV-OC43 it is now recognised that other non-SARS human respiratory coronaviruses can cause pneumonia at low frequencies (Peiris *et al.*, 2003, Vabret *et al.*, 2003, Woo *et al.*, 2005b) and HCoV-HKU1 has now been reported in USA in hepatitis patients (Esper *et al.*, 2006) (Kistler *et al.*, 2007), Germany (Kupfer *et al.*, 2007, de Souza Luna L. K *et al.*, 2007), Italy (Bosis *et al.*, 2007), France (Vabret *et al.*, 2006), Australia (Sloots *et al.*, 2006) and Malaysia (Amini *et al.*, 2012).

Currently, HCoV-HKU1 is diagnosed from patient Naso-Pharyngeal Aspirate (NPA) samples by RT-PCR and DNA sequencing using HCoV-HKU1-specific primers LPW1926 as a forward primer (AAAGGATGTTGACAACCCTGTT) and LPW1927 as a reverse primer (ATCATCAT- ACTAAAATGCTTACA) (Lau *et al.*, 2006). In addition, an ELISA-based antibody test with recombinant N protein has been developed in order to detect specific antibodies against the N protein of HCoV-HKU1 and a modified immunofluorescence assay is used with NPA samples using a monoclonal antibody specific for HCoV-HKU1 (Gerna *et al.*, 2007).

1.2. Coronavirus replication.

Coronaviruses use a special strategy for genome replication and transcription as can be seen in **Figure 4**. The process of infection is initiated by the attachment of the S protein to a variety of receptors depending on the coronavirus species. Viruses are made up of proteins, nucleic acids and sometimes lipids and sugars. They are metastable and poised for structural changes. These features allow viruses to communicate with host cells during entry, and to release the viral genome, a process known as uncoating. Studies

have shown that hundreds of host factors directly or indirectly support this process. The cell provides molecules that promote stepwise virus uncoating and direct the virus to the site of replication. It acts akin to a snooker player who delivers accurate and timely shots (cues) to the ball (virus) to score. The viruses, on the other hand, trick (snooker) the host, hijack its homeostasis systems, and dampen innate immune responses directed against danger signals (Yamauchi and Greber, 2016). A key step in uncoating is the acidification of the content of the endosome to a pH of about 5, owing to the activity of a proton pump present in the membrane. The low pH causes rearrangement of coat components, which then expose normally hidden hydrophobic sites. They bind to the lipid bilayer of the membrane, causing the extrusion of the viral core into the cytosol. For influenza virus, the acid-sensitive component is the core HA2 unit of the haemagglutinin, for adenoviruses, it is the penton base (Blaas, 2016).

The coronavirus nucleocapsid (N) protein packages viral genomic RNA into a ribonucleoprotein complex. Interactions between N proteins and RNA are thus crucial for the assembly of infectious virus particles (Fan *et al.*, 2005). The nucleocapsid protein of IBV (IBV-N) is a phosphoprotein of 409 amino acids that is well-conserved across various IBV strains (Williams *et al.*, 1992) and The 45 kDa recombinant nucleocapsid N protein of coronavirus infectious bronchitis virus (IBV) is highly sensitive to proteolysis (Fan *et al.*, 2005). It forms a protective shell that packages the viral genomic RNA of 27.6 kb and is also thought to participate in viral RNA replication and transcription. Specific packaging of viral genetic material is usually performed via the recognition of a particular nucleotide sequence by a nucleocapsid protein. Such “packaging signals” have been identified at the 3′ end of the viral genomes of mouse hepatitis virus (MHV)

(Fosmire *et al.*, 1992) and bovine coronavirus (BCV) (Cologna and Hogue, 2000) and at the 5′ end of the TGEV genome (Escors *et al.*, 2003), but not unambiguously for the IBV genome. In elegant structural studies performed in other viral families with RNA genomes, such as HIV (De Guzman *et al.*, 1998) and the MS2 bacteriophage (Valegard *et al.*, 1997), the packaging signals were seen to form a stem-loop structure that is recognized by the nucleocapsid protein. In the case of the IBV genome, this special RNA structure has not been determined with certainty, although previous studies demonstrated that the IBV-N protein interacts specifically with RNA sequences located at the 3′ noncoding region of the viral genome (Zhou and Collisson, 2000). Both the N- and C-terminal domains of IBV-N, but not its middle region, bind to an oligoribonucleotide of 155 nucleotides, located at the 3′ end of the viral genome nontranslated region, however, there is a limited knowledge about the details of this interaction and how it relates to virus assembly (Zhou and Collisson, 2000).

As a result, there is a virus-cell fusion, either at the cell membrane or following uptake into the cell via a vesicle, and subsequent release of the virus genome into the cytoplasm (Risco *et al.*, 1996). Following formal entry, direct translation of the positive-strand genomic RNA gives rise to a large polyprotein that undergoes proteolytic processing to generate an RNA-dependent RNA polymerase along with associated non-structural proteins. Only the first two thirds of the genome is translated, the OFR 1a and 1b products and through the action of one of these products, the RNA polymerase, a full-length, antisense negative-strand template is generated.

Subgenomic mRNAs are synthesized using the negative strand as template and translation of these subgenomic mRNAs gives rise to the viral

structural proteins (Lai, 1990). In addition, there is formation of both genomic and multiple subgenomic RNA species. This is followed by assembly of helical capsids around the new full length ssRNA (+) genome and budding of these complexes at membranes of the endoplasmic reticulum (ER), the intermediate compartments, and/or the Golgi complex. This is followed by final release of new virions (Gorbalenya *et al.*, 2006, Hulo *et al.*, 2011).

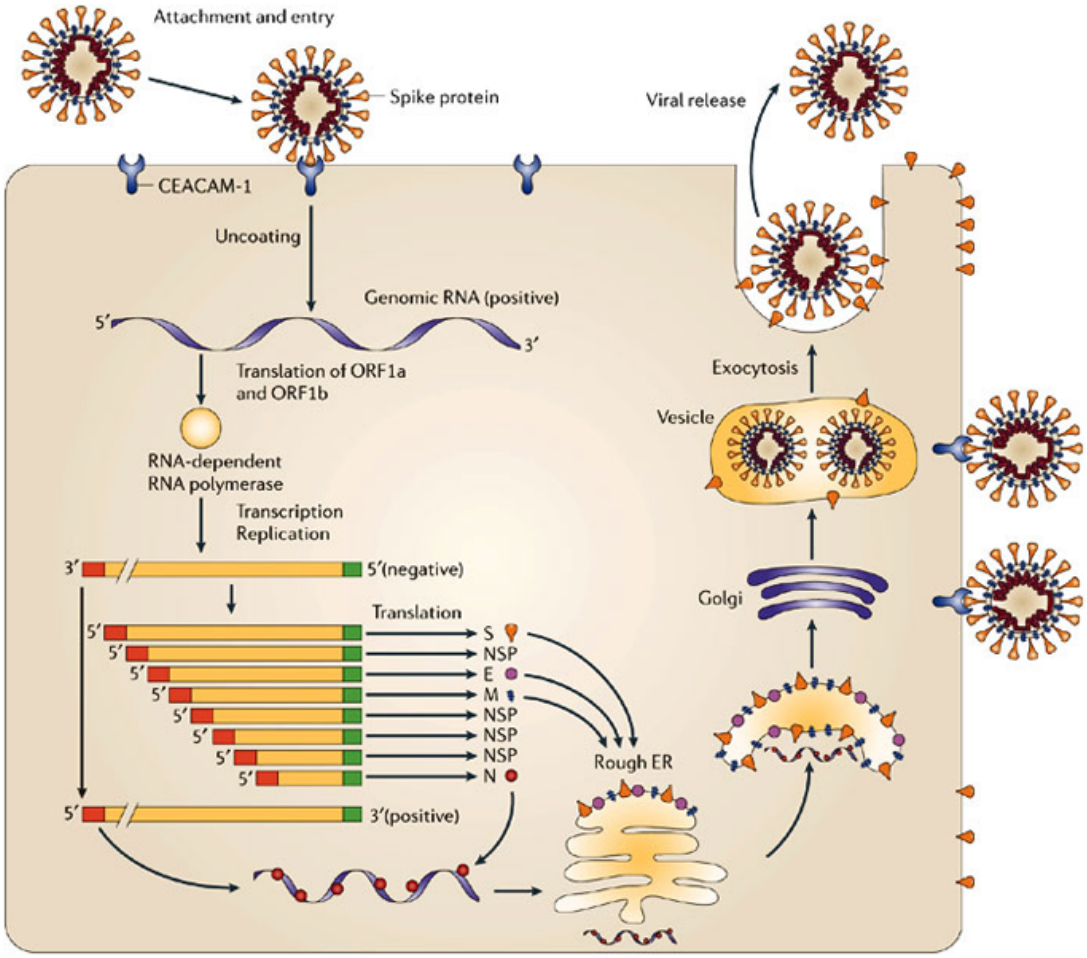

Figure 4. Summary of mouse hepatitis virus (MHV) replication. Host-cell receptor CEACAM-1. E, envelope protein; ER, endoplasmic reticulum; M, membrane protein; N, nucleocapsid protein; ORF, open reading frame (Adapted from (Bergmann *et al.*, 2006)).

1.3. Human airway epithelium cells.

The difficulty in obtaining high titre clinical specimens, and the inefficiency of propagating potential viruses in recognised cell lines, can limit the possibilities for discovering novel viruses. In recent years, viral respiratory pathogens such as human metapneumovirus (van den Hoogen *et al.*, 2001) and human coronavirus NL63 (HCoV-NL63) (van der Hoek *et al.*, 2004), both of which propagate slowly in standard cell lines, have been identified using molecular approaches to amplify novel sequences from infected cells. While this is an effective tool for virus discovery, the lack of a suitable cell culture system for a virus severely restricts molecular studies on virus replication and pathogenicity.

As a result, human airway epithelial (HAE) cells have been evaluated as a culture system for the initial propagation of novel human respiratory viruses, followed by visualization of the virus by transmission electron microscopy (TEM). As noted for HKU1, the use of an amplification approach with conserved primers can detect viral sequences but HAE cell infection provides excellent confirmatory evidence for viable virus replication **Figure 5** (Fulcher *et al.*, 2005, Pyrc *et al.*, 2010). HAE cultures are derived from primary bronchial epithelial cells isolated from the airways of human lung donors (Berube *et al.*, 2010) or patients with chronic diseases such as asthma and cystic fibrosis (Ostrowski *et al.*, 2012); and have been used widely to study the cell biology of the respiratory epithelium (Pickles *et al.*, 1998, Prytherch *et al.*, 2011), modelling respiratory disease (Ostrowski *et al.*, 2012), viral respiratory infection (Matrosovich *et al.*, 2004) and toxicity studies (Wu *et al.*, 1986).

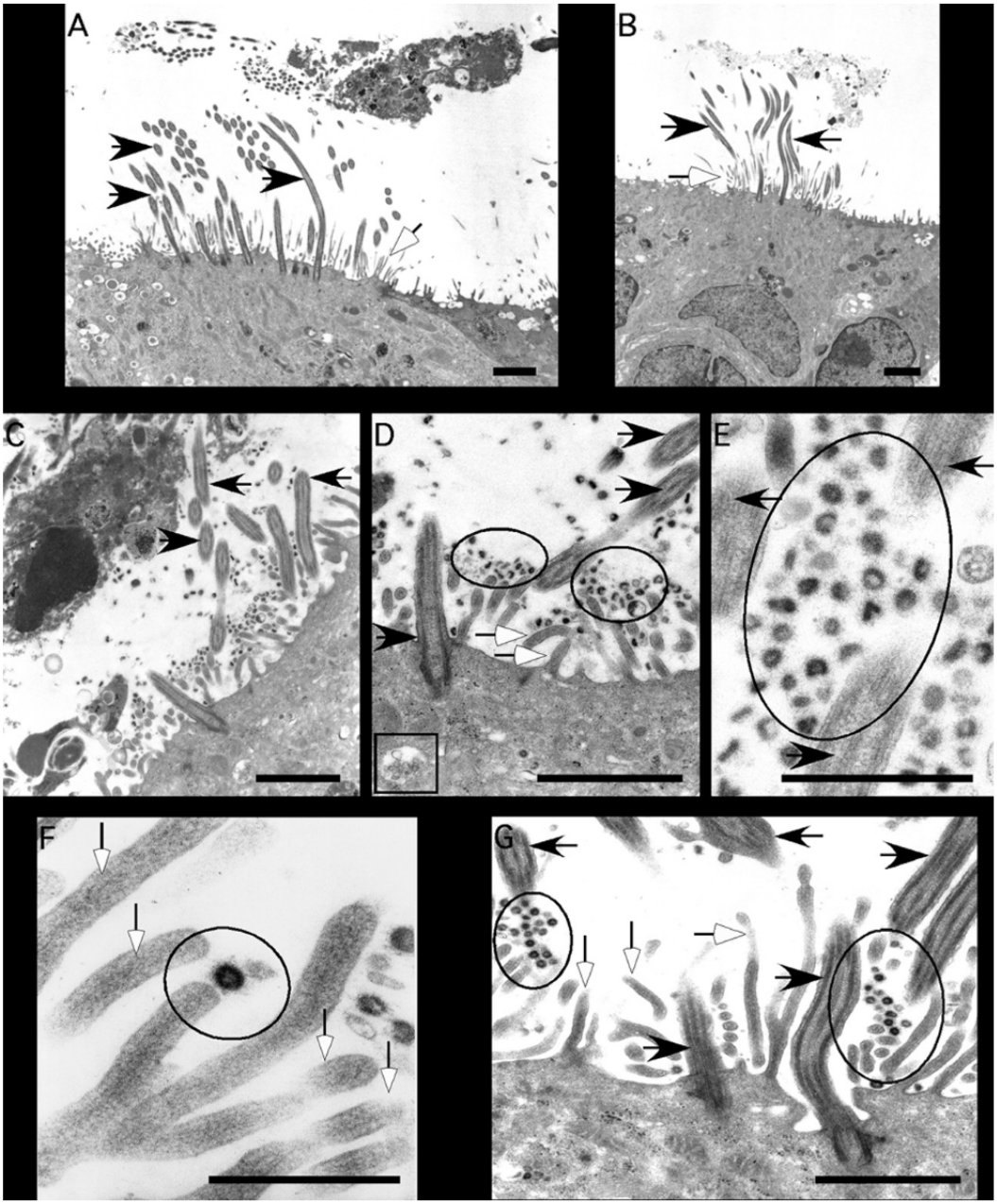

Figure 5. Transmission electron microscopy of HCoV-HKU1 in HAE. (A and B) HAE mock inoculated with the vehicle alone, showing the standard morphological structures of the apical surfaces of HAE with cilia (black arrows) and microvilli (white arrows). (C to G) HAE infected with HCoV-HKU1 for 96 h showing the presence of the large numbers of virions (circled) associated with the surfaces of ciliated cells or shed into pericilial areas (black arrows, cilia; white arrows, microvilli). Intracellular virions were also noted inside vesicular structures in the cytoplasm of ciliated cells (D, box). (F) magnified image of a virion connected to the tip of a microvillus. Scale bars are shown in the lower right of each panel and represent 2 μm in panels A to D and G and 1 μm in panels E and F (Adapted from (Pyrc *et al*., 2010)).

1.4. RNA secondary structure prediction.

All RNA molecules have a net negative charge due to the presence of the phosphate groups in the oligoribonucleic acid chain (Draper, 2004). As a result, these charged RNA molecules are not stable inside living cells and in order to gain more stability, some parts of a single-stranded RNA folds back on itself forming RNA secondary structure [stem loop of RNA and RNA with pseudoknots] with an increase in structural stability with minimum free energy (Borkar and Mahajan, 2014). In other words, base pairing interactions occur within a single molecule or set of interacting molecules.

RNA secondary structure originates from the pairing up of the four nucleotides according to the rules of Watson-Crick and Wobble pairs [(A, U), (U, A), (G, C), (C, G), (G, U), (U, G)] (Dirks *et al.*, 2004). These nested base pairings result into two-dimensional structures called secondary structures. Different kinds of loops result from pairings among these nucleotides and can be classified according to the number of branches present in them. Essentially ssRNA secondary structure forms two large groups of stem-loops and pseudoknots in which these loops includes hairpin loop, interior loop, multi loop, stack loop, bulge loop and external loop **Figure 6** (Borkar and Mahajan, 2014).

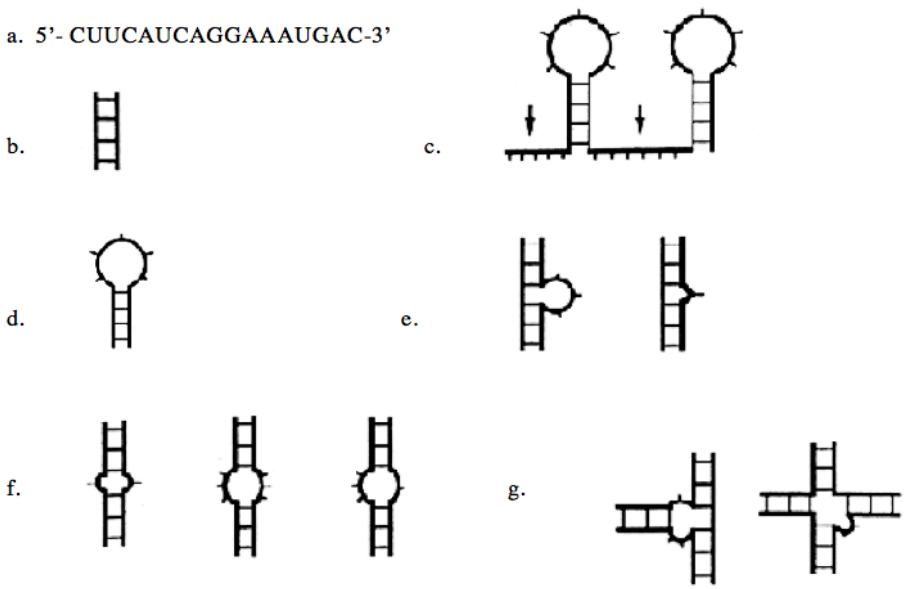

Figure 6. RNA secondary structure motifs. (a) Primary structure; (b) Duplexes; (c) Single stranded regions; (d) Hairpins; (e) Bulges and Single base bulge; (f) Internal loops: Mismatch, Symmetric and Asymmetric; (g) Three and Four stem junction (Adapted from (Borkar and Mahajan, 2014)).

The prediction of RNA secondary structure is important for understanding their potential roles and for both *in silico* and *in vitro* studies (Laing and Schlick, 2011). There are several methods that can be used to predict the RNA secondary structure and each one has its own merits, strengths and weaknesses. Such methods are based on a thermodynamic energy minimization model (Hor *et al.*, 2013), genetic algorithm (GA) (Borkar and Mahajan, 2014), comparative approaches from multiple homologous sequences and a comparative prediction consensus of the RNA structures by using SCFG methods (Hor *et al.*, 2013). In addition, many computational methods have been introduced to reduce the computational complexity of prediction by using the parallel multicore architecture, graphics hardware like GPU, message passing interface (MPI) library and the fine-grained implementation on FPGA Tables 1 and 2 (Reuter and Mathews, 2010).

Table 1. Single sequence methods for RNA secondary structure prediction (Reuter and Mathews, 2010).

Feature:	Text Interface Program:	JAVA/Windows GUI Menu Item:	Class Library and Function Name:
Free energy minimization structure prediction [14]	Fold	Fold RNA Single Strand	RNA::FoldSingleStrand
Maximum expected accuracy structure prediction [18]	MaxExpect	Predict Maximum Expected Accuracy Structure	RNA:MaximizeExpectedAccuracy
Partition function [22]	partition	Partition Function RNA	RNA::PartitionFunction
Efn2 (energy calculator) [13]	efn2	Efn2 RNA	RNA:: CalculateFreeEnergy
Free energy minimization and generation of all suboptimal structures [28,47]	AllSub	Generate All Suboptimal RNA Structures	RNA::GenerateAllSuboptimalStructures
Stochastic sampling of structures [17]	stochastic	Stochastic RNA Sampling	RNA::Stochastic
Remove Pseudoknots [49]	RemovePseudoknots	Break Pseudoknots	RNA::BreakPeudoknots
Prediction of structures with pairs above specified pairing probability threshold [22]	ProbablePair	Output Probable Structure	RNA::PredictProbablePairs
Drawing secondary structure diagrams	draw	Draw	RNA::DetermineDrawingCoordinates
NAPSS [30]	NAPSS	-	-

Table 2. Multiple-sequence methods for RNA secondary structure prediction (Reuter and Mathews, 2010).

Feature:	Text Interface Program:	JAVA/Windows GUI Menu Item:	Class Library and Function Name:
Dynalign [23,41,46]	dynalign	Dynalign RNA	Dynalign_object::Dynalign
OligoWalk [31-33]	OligoWalk	OligoWalk	Oligowalk_object::Oligowalk
OligoScreen [55]	oligoscreen	OligoScreen	Oligowalk_object::OligoScreen
Bimolecular structure prediction with intramolecular pairs [33]	bifold	Fold RNA Bimolecular	HybridRNA::FoldBimolecular
PARTS [24,48]	PARTS	-	-
Bimolecular partition function (no intramolecular pairs) [31]	bipartition	Partition Function RNA Bimolecular	HybridRNA::PartitionFunctionBimolecular
Bimolecular structure prediction without intramolecular pairs	DuplexFold	-	HybridRNA::FoldDuplex

1.5. Coronavirus RNA secondary structures.

1.5.1. Coronavirus 5' untranslated region (5′-UTR).

These are sequences at the 5' end of mRNA that are not translated into protein. They extend from the transcription start site to just before the AUG translation initiation codon **Figure 7** (Araujo *et al.*, 2012). The 5'-UTR of mouse hepatitis virus (MHV) includes a higher-order structural sequence that functions as a cis-active element that is vital for viral genome transcription and replication (Liu *et al.*, 2007). The 5'-UTR of porcine transmissible gastroenteritis virus (TGEV) and bovine coronavirus (BCoV) has been extensively studied and has mapped four stems–loops (I, II, III, and IV) within the 5'-UTR (210-nt) **Figure 8** (Raman and Brian, 2005). The predicted SL-I has 11–42 nucleotides, contains just three contiguous Watson–Crick base pairs and has a large 16-nucleotides loop that is not conserved among group 2 coronaviruses.

The predicted SL-II has 51–84 nucleotides and is an A-U base-pair-rich hairpin that folds the transcription regulatory sequence (TRS, the core motif described above that causes an RNA-dependent RNA polymerase (RdRp) template switch site) into the terminal loop (Liu *et al*., 2007). On other hand, SL-III is phylogenetically conserved among group 2 coronaviruses and seems to have homologues in coronavirus groups 1 and 3 (Raman *et al.*, 2003). The predicted SL-IV of 186–215 nucleotides is also conserved among group 2 coronaviruses (Raman and Brian, 2005).

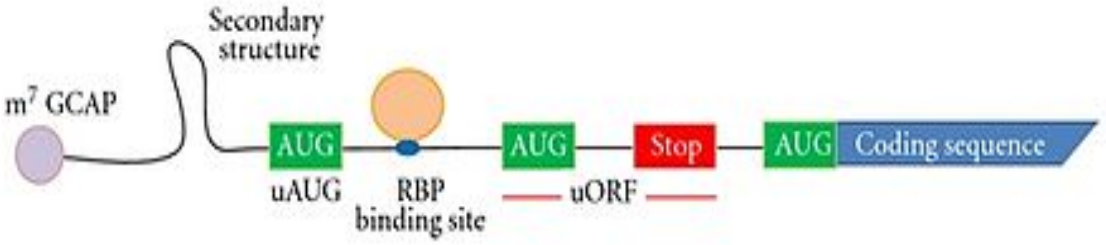

Figure 7. Regulatory elements present in 5'-UTR (Araujo *et al.,* 2012). Key features of the regulatory elements are marked, gray colored circle refer to the 5′ cap (cap-0), found on the 5′ end of an mRNA molecule, consists of a guanine nucleotide connected to mRNA via an unusual 5′ to 5′ triphosphate linkage. This guanosine is methylated on the 7 position directly after capping in vivo by a methyltransferase, uAUG: upstream initiation site, RBP is a RNA binding protein site, uORF is an upstream open reading frame.

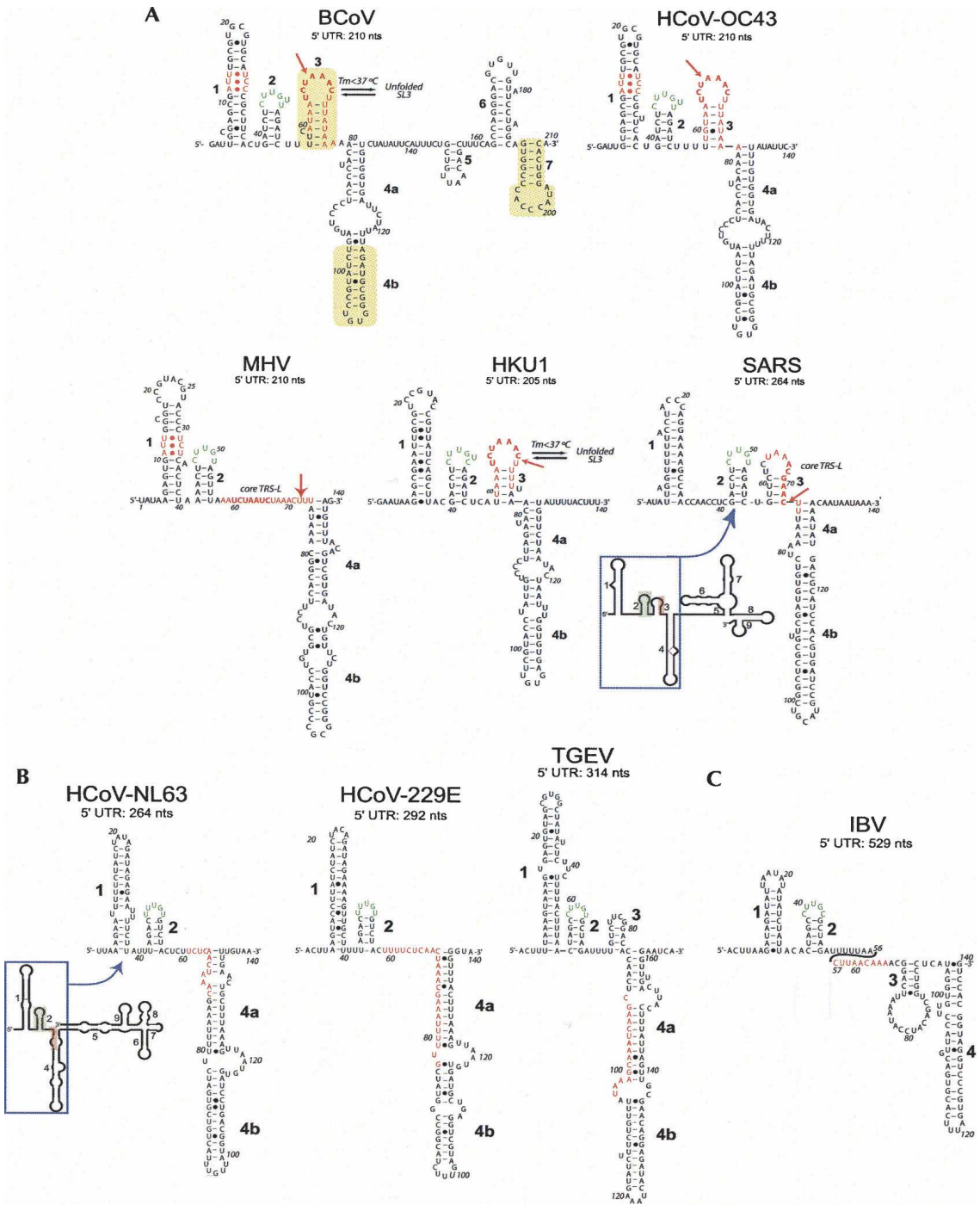

Figure 8. Prediction of the secondary sequence elements include in the 5'-UTR coronaviruses. (A) Prediction and comparison of the entire 5'-UTR of BCoV with the 5' 140 nt of selected group 2 coronaviruses. (B) Prediction of the secondary structure models for three group 1 coronaviruses. (C) Prediction of the secondary structure model of a group 3 coronavirus. (Bold numbers) Predicted stem–loops SL1, SL2, SL3, and SL4 (4a and 4b), (bold red letters) leader TRS-L sequences, (yellow) SL-II, SL-III, and SL-IV (Adapted from (Raman and Brian, 2005)).

1.5.2. Coronavirus transcriptional regulating sequence (TRS).

The TRS is an intragenic sequence or leader-body junction site that is rich in Adenine and Uracil elements (Pasternak *et al.*, 2006). It is a short-repeated sequence of about 7-18 nucleotides found in the leader TRS region and at many other sites, termed body sites, throughout the genome, all preceding translated sequences. The leader–body junction occurs at numerous sites and includes conserved 39-proximal nucleotides within the main TRS (Smits *et al.*, 2005) and the TRS immediately upstream of the initiating AUG for the ORF that follows (Sawicki *et al.*, 2007). The MHV TRSs are located at the 3' end of the leader (TRS-L), which contains two to four 5'-UCUAA-3'repeats, and the body (TRS-B) of each gene which contains 5'-AAUCUAAAC-3' (or a closely related sequence depending on the virus strain) (Sola *et al.*, 2011).

These TRSs at the 5' end of each gene represents signals that regulate the discontinuous transcription of subgenomic mRNAs (sgmRNAs) and include a core sequence (CS; 5'-CUA AAC-3') and the 5' and 3' flanking sequences (5' TRS and 3'TRS, respectively) that modulate transcription **Figure 9** (Zuniga *et al.*, 2004, Dufour *et al.*, 2011).

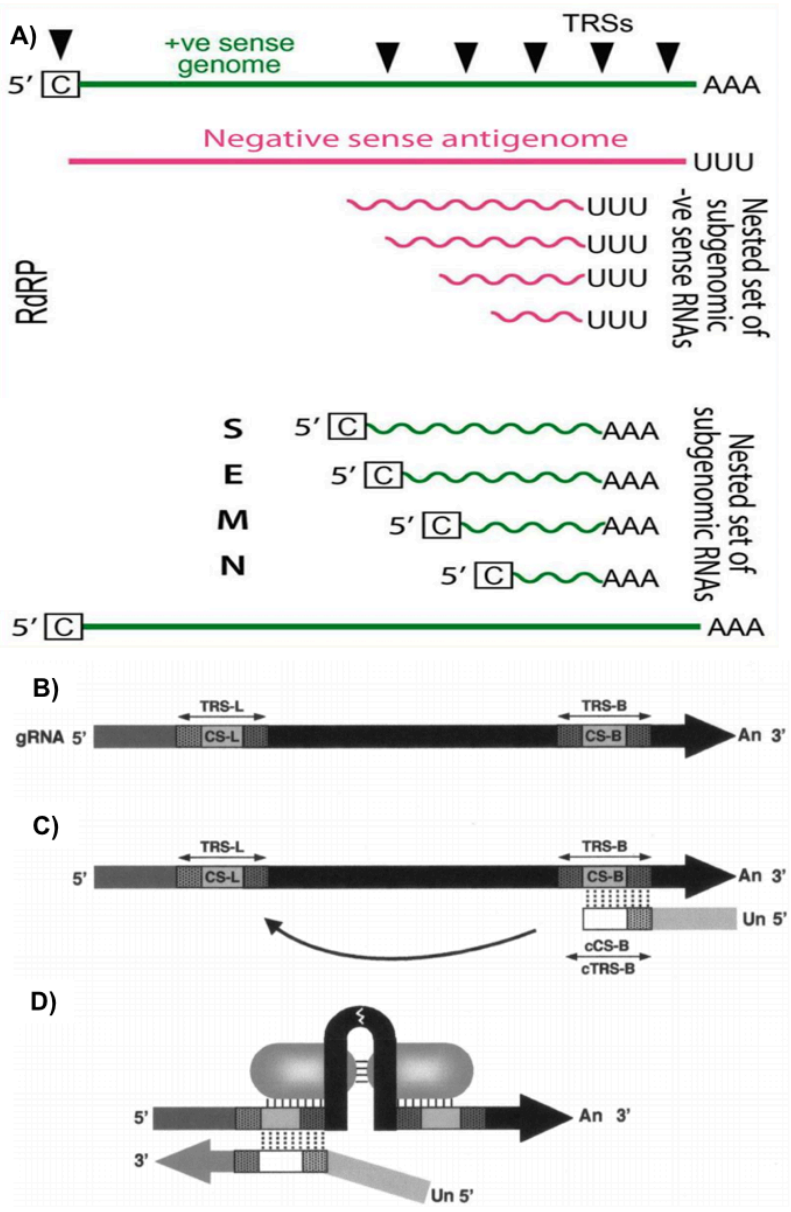

Figure 9. The sequence elements include in coronavirus transcription. (A) the continuous and discontinuous transcription negative-strand and positive-strand synthesis model (B) All the sequence elements probably involved in the discontinuous negative-strand synthesis model. CS-L; core sequence leader, CS-B; core sequence body, TRS-L and TRS-B, transcription regulating sequences from the leader and body, respectively. An, poly(A). (C) Representation of the discontinuous transcription during negative-strand synthesis. cCS-B and cTRS-B represent the CS-B and TRS-B complementary sequences, respectively. Un, poly(U). (D) Leader and body sequences are probably located close to one another in higher-order structures maintained by RNA-protein and protein-protein interactions (Adapted from (Zuniga *et al.*, 2004)).

1.5.3. Coronavirus 3' untranslated region (3′-UTR) pseudoknot.

Members of coronaviruses such as mouse hepatitis virus (MHV), bovine coronavirus (BCoV) and HCoV-HKU1 have a pseudoknot structure in the 3' untranslated region 63-115 nucleotides downstream of the N gene (nucleotide position 29708-29760) that is essential for viral replication **Figure 10** (Woo *et al.*, 2005b). A bulge occurs due to an overlap between the last segment of the stem-loop and stem 1 in the pseudoknot **Figure 11** and the structure has a conserved counterpart in every group 1 and group 2 coronavirus (Goebel *et al.*, 2004). A pseudoknot is a type of tertiary interaction that includes base pairing between nucleotides in a loop with nucleotides in a single-stranded region outside the loop. It has two double-helical stem regions and two loop regions **Figure 11**.

The common type of pseudoknot is the H-type which consists of a hairpin stem (stem 1) with a second stem formed by the downstream region those base pairs with the loop of stem 1 (Giedroc *et al.*, 2000). Pseudoknots usually serve as structural elements helping to stabilize complex 3D structures and perform an active role as vital elements in the regulation of several biological processes such as binding of ribosomal proteins to RNA (Giedroc *et al.*, 2000), initiation of internal ribosome entry translation (Wang *et al.*, 1995), and controlling the translational frame during protein synthesis (Green *et al.*, 2008).

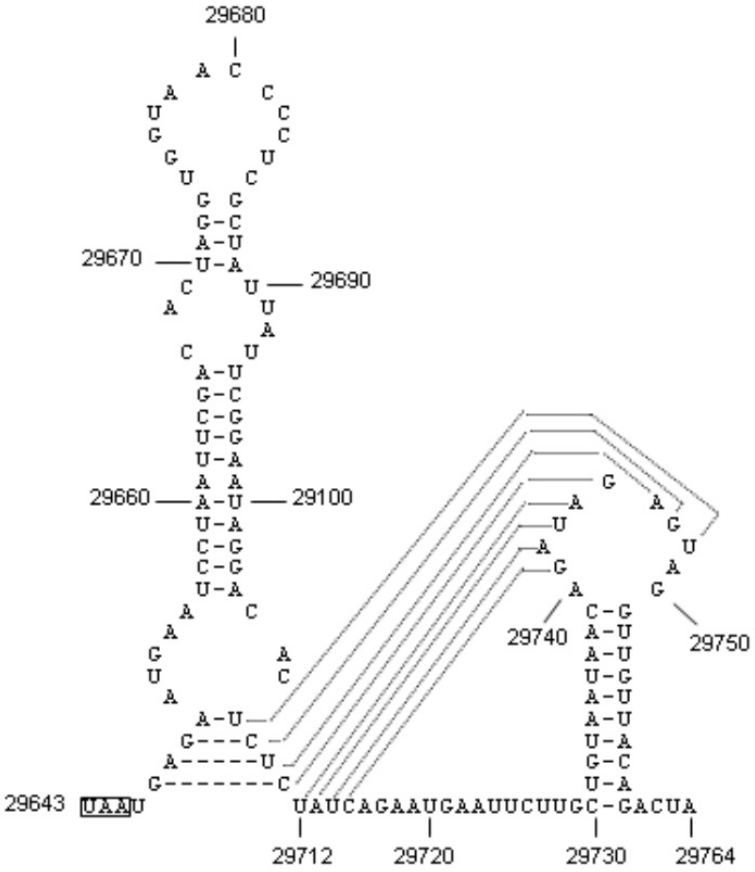

Figure 10. Predicted bulged stem-loop structure and pseudoknot in HCoV-HKU1 3' untranslated region (Woo *et al.*, 2005a).

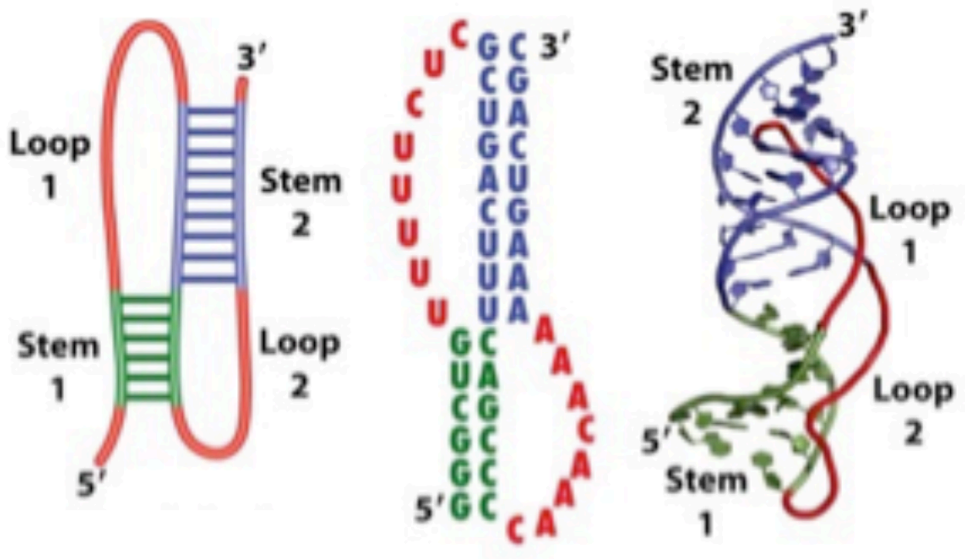

Figure 11. Pseudoknot (the tertiary structure). Represented by green double helical stem 1 region, mauve double helical stem 2 regions, and red signal loop 1and 2 regions (Adapted from (Giedroc *et al.*, 2000)).

1.5.4. Coronavirus 3' stem-loop II-like motif (s2m).

The 3' stem-loop II-like motif is an RNA secondary structure motif recognized in the 3’ untranslated region (3'-UTR) of astrovirus, coronavirus and equine rhinovirus genomes (Jonassen *et al.*, 1998, Robertson *et al.*, 2005). It has an important role in viral replication whereby it might bind to the host’s eukaryotic translation initiation factor 1A (eIF- 1A) to take-over the host translational machinery for use by the virus, or to bind other translational regulation proteins having similar folds for similar purposes (Hor *et al.*, 2013). It seems to be conserved at both nucleotide sequence and secondary structure folding levels indicating a strong evolutionary selection for its conservation **Figure 12**. The presence of this conserved motif in three different viral families is suggested to be the result of at least two separate recombination events and they have been important for the design of anti-viral therapeutic agents (Hor *et al.*, 2013).

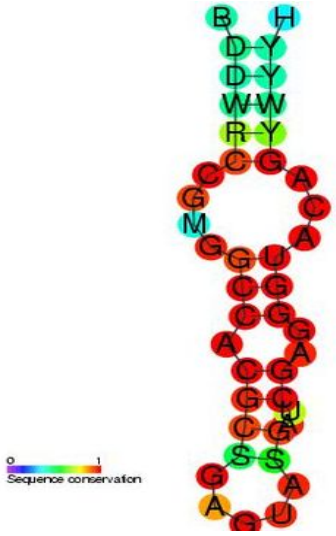

Figure 12. Coronavirus 3' stem-loop II-like motif (s2m) (Griffiths-Jones *et al.*, 2005). Red colored circle refers to conserved base (A or U or G or C), D: (G or A or T), W: (A or T), R: (G or A), M: (A or C), S: (G or C), Y: (T or C) and H: (A or C or T).

1.5.5. Pseudoknot ribosomal frameshifting.

Programmed -1 frameshifting is a directed change in the translational reading frame that allows the production of a single protein from two or more overlapping open reading frames **Figure 13** (Bailey *et al.*, 2014). This -1 programmed ribosomal frameshifting happens in mammals (Clark *et al.*, 2007), yeast (Dinman *et al.*, 1991) and bacteria (Tsuchihashi and Kornberg, 1990). The process is most studied in plant and animal RNA viruses, including HIV-1 and related retroviruses and coronaviruses including SARS-CoV (Brierley and Dos Ramos, 2006). This process is programmed by the nucleotide sequence of the mRNA which consists of two parts; one is the called the heptanucleotide “slippery sequence” with a general sequence (X-XXY-YYZ) over which the ribosome

pauses and the other is the pseudoknot just downstream that causes the ribosome to stutter during translation, stimulating the frameshift event.

For example, in the coronavirus infectious bronchitis virus [IBV] the slippery sequence UUUAAAC occurs (Farabaugh, 1996) and in Coronavirus HKU1 it is UUUAAAC (Woo *et al.*, 2005b) and is sometimes affected by the secondary or tertiary mRNA structure **Figure 14** (Dos Ramos *et al.*, 2004).

Frameshifting does not occur on every RNA, most translation therefore stops at or just after the frameshift site. However, the occasional frameshift fuses the 3' downstream ORF with the 5' upstream component leading to a defined molar ratio between the upstream and downstream produced proteins, as required during virus assembling. It has been found that lowering the frameshifting efficiency results in reduction of virus infectivity during frameshift use for HIV-1 and Murine Maloney leukaemia virus (Biswas *et al.*, 2004, Dulude *et al.*, 2006) showing that the evolved ratio of upstream and downstream sequence translation is critical for viral infectivity (Shehu-Xhilaga *et al.*, 2002, Dos Ramos *et al.*, 2004).

In coronaviruses a frameshift signal occurs at the junction of ORF1a and 1b. As a consequence, the 1b reading frame is translated only about 25% of the frequency of the 1a reading frame from the same positive (+ve) strand genomic template. Thus, two polyproteins are produced in different amounts in infected cells without the need to regulate two different ORFs. As noted similar mechanisms are used in many other viruses, notably retroviruses, to effect the same differential regulation of protein expression.

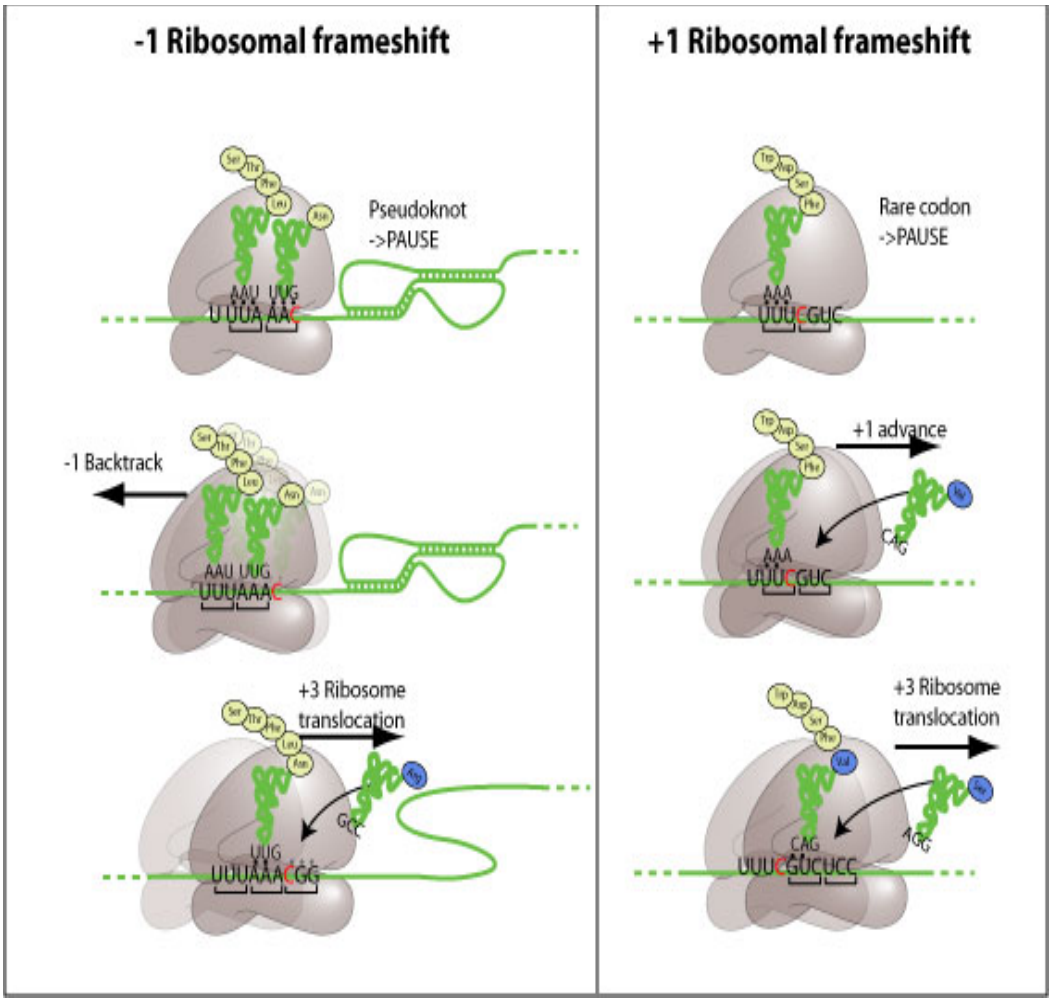

Figure 13. Ribosomal frameshifting programmed. Cis-acting mRNA elements essential for −1 programmed frameshifting: (gray) large and small subunit of the ribosome move along (green) mRNA toward the 3' end and pause at the (black) slippery sequence when encountering a downstream pseudoknot (green). (Green) tRNAs are shown in the zero reading frames before the frameshift has occurred (Adapted from (Hulo *et al.*, 2011)).

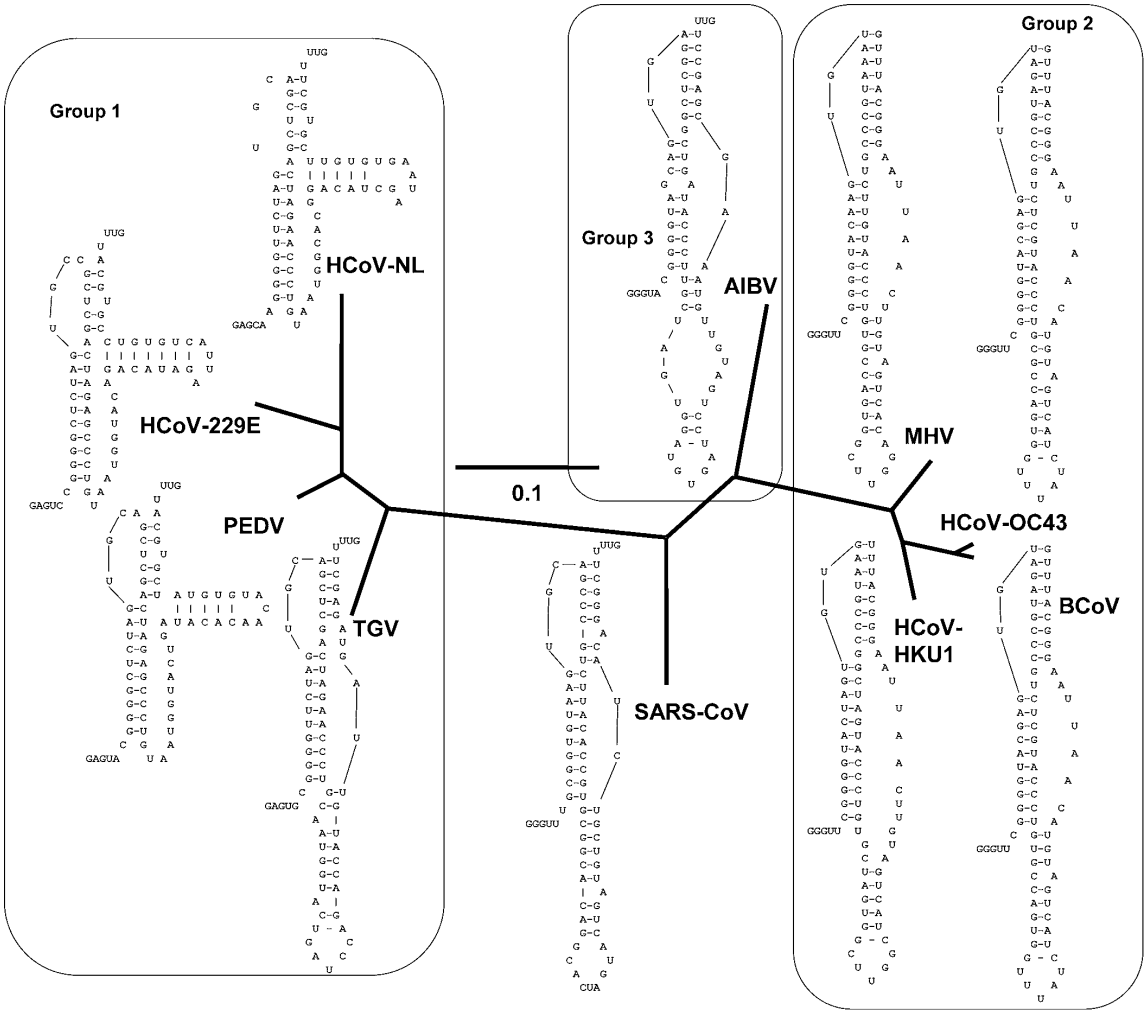

Figure 14. Prediction of 1a/1b ribosomal frameshifting signals from group 1, 2 and 3 coronaviruses (Plant *et al.*, 2005).

1.6. Coronavirus RNA secondary structure regulating proteins.

1.6.1. Nucleoproteins as RNA chaperones.

Nucleoproteins are relatively nonspecific nucleic acid binding proteins with long disordered regions that assist RNA molecules to assume their functional conformation or prevent RNA molecule from degradation in the cellular environment (Semrad, 2011). Use of coronavirus nucleocapsid proteins (N) from transmissible gastroenteritis coronavirus (TGEV) and severe acute respiratory syndrome (SARS)-CoV) as models for RNA chaperone activity has been reported (Zuniga *et al.*, 2007, Zuniga *et al.*, 2009). The *in silico* prediction of N protein structure suggests the same function for all other coronaviruses (Zuniga *et al*., 2007).

Nucleoprotein binding to RNA (RNA chaperone activity) is important to increase ribozyme cleavage, to allow rapid and accurate RNA annealing, and to simplify strand transfer and exchange. However, the ability of this chaperone activity to increase ribozyme cleavage is the most significant (Rajkowitsch *et al.*, 2005). Several methods have been used to assay all types of RNA chaperone activity *in vitro* including using annealing, strand displacement, cis-splicing, trans-splicing and hammerhead ribozyme cleavage assays and *in vivo* by using intron folding trap and transcription anti-termination assays (Rajkowitsch *et al*., 2007). Generally, these assays involve labelling of the RNA and gel electrophoresis to study the proportion of single- and double-stranded fractions (Rajkowitsch *et al*., 2005). Alternatively, the RNA may be labelled with direct reporters, for example FRET compatible dyes, whose excitation or quenching occurs when strands are annealed or separated. The activity of an RNA chaperone added to these assays can then be directly read by the measured emissions **Figure 15**.

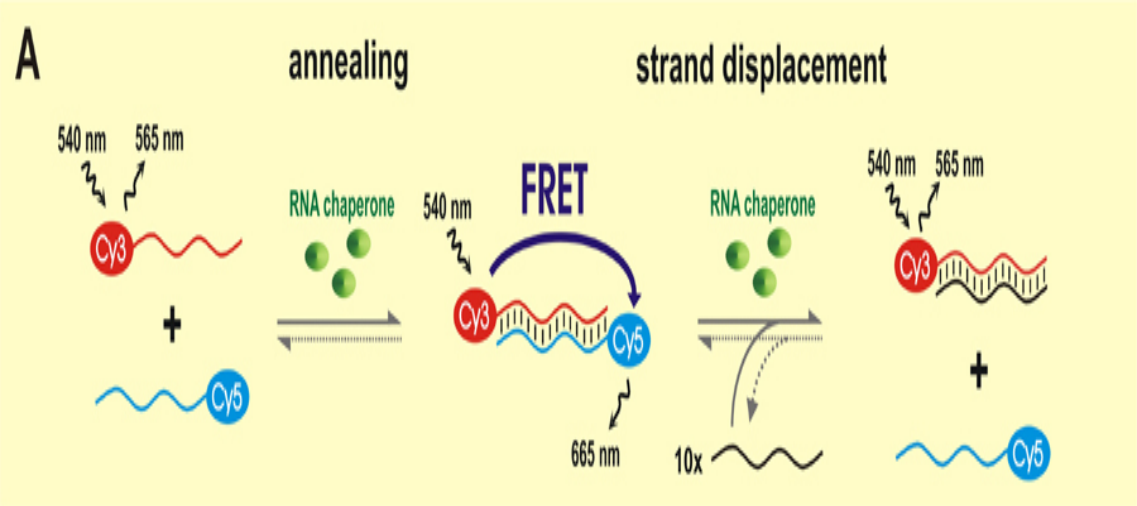

Figure 15. In vitro assays for annealing and strand displacement. (A) Using of fluorophore-labelled RNAs. Hybridisation of (red + blue) CyDye-labelled short ribooligonucleotides results in fluorescence resonance energy transfer (FRET), and this reaction is accelerated in the presence of (green) RNA chaperones. For strand displacement, a double-labelled RNA duplex is incubated with an excess of (black) non-labelled competitor strand. RNA chaperone activity is required to facilitate strand dissociation and formation of single-labelled, FRET-inactive RNA duplexes. On the other hand, radioactively labelled RNAs have to be gel-separated at distinct time point's prior quantification; fluorescence signals can be measured in real time in solution (Adapted from(Rajkowitsch *et al.*, 2007)).

1.6.2. Non-structural protein 3 (nsp3) as a RNA chaperone.

Mouse hepatitis virus (MHV), as a member of coronaviruses, encodes non-structural protein 3 (nsp3) which is a ~200kDa protein cleaved from the polyprotein 1a or 1ab by the protease PLpro and is essential for coronavirus RNA synthesis and virogenesis. Nsp3 has multiple large domains within the coding

sequence which include two papain-like protease domains, PLP1 and PLP2, and a predicted trans membrane (TM) domain (Kanjanahaluethai *et al.*, 2007).

In addition, nsp3 has two additional RNA-binding domains, one of which is a canonical nucleic acid chaperone-like domain located immediately downstream of the papain-like proteinase domain while the second domain is a novel cysteine-coordinated metal ion-binding domain. Nsp3, nsp4, nsp5 and nsp6 are important coronavirus proteins as they form a conserved block that are involved in forming the double-membrane vesicles (DVMs) that are the sites of viral RNA synthesis in infected cells (Neuman *et al.*, 2008).

1.6.3. Non-structural protein 13 (nsp13) as a helicase.

Nsp13 is one of 16 non-structural proteins that are encoded by ORF 1b and is processed from pp1ab by the 3C-like proteinase. It has the ability to uncoil both RNA and DNA duplexes in the 5' to 3' direction with high processivity (Ivanov *et al.*, 2004). It hydrolyses all typical nucleotides and deoxynucleotides by possessing a deoxynucleoside triphosphatase (dNTPase) activity and also has RNA 5′-triphosphatase activity, which may be involved in the first step of formation of the 5′ cap structure of the viral mRNAs (Ivanov *et al*., 2004). These two-hydrolase activities seem to be have a common active site, which includes a canonical Walker A NTPase-like motif which is a phosphate-binding P-loop (Ivanov *et al.*, 2004). As these NTPase/helicase proteins are considered essential for the viral life cycle (Kadare and Haenni, 1997) they have been considered as novel potential targets for the treatment of viral infection (Anand *et al.*, 2003).

1.7. Validation methods for RNA secondary structure.

Several approaches have been used to validate RNA secondary structure experimentally and these techniques includes:

1.7.1. RNAse fingerprinting.

Fingerprinting provides evidence of the RNA secondary structure by using digestion of target folded RNA structures with ribonucleases. These ribonucleases (RNases) recognise specific sites in the single-stranded RNA and cleave them to oligonucleotides whose length can then predict where the ssRNA region occurred in the folded molecule. Several types of ribonucleases, with different specificities, have been used for this purpose such as RNase T1 (cleavage at the 3'-end of unpaired G residues) (Pomerantz *et al.*, 1993), RNase U2 (cleavage at the 3'-end of unpaired A residues) (Pomerantz *et al.*, 1993), RNase T2 (cleavage at the 3'-end of all four residues but preferentially 3'-end of A residues), RNase CL3 (cleaves 39 nts 3' to C-residues), RNase A (cleaves 39 nts 3' to C- or U-residues) and RNase J1 (cleaves ssRNA, at AU-rich regions). All have been used for probing the interaction between, for example, RNA and the 30S ribosomal subunit (Daou-Chabo and Condon, 2009).

RNase P1 has a broad specificity for single-stranded nucleotides while RNase (II) and RNase R are also used to induce cleavage in ssRNA regions (Cheng and Deutscher, 2005). RNase V1 specially cleaves phosphodiester bonds of double-stranded RNA, whereas S1 nuclease specially cleaves single-stranded RNA (Kertesz *et al.*, 2010). *Neurospora crassa* nuclease (*N.c.* nuclease) has been also used for the same purpose **Table 3** (Ehresmann *et al.*, 1987).

Experientially, the process of fingerprinting includes mixing of specific ribonuclease (RNase) with purified viral RNA and allowing cleavage to occur based on the level of single-stranded (ss) or double-stranded (ds) RNA present in the structure. The resulting ribo-oligonucleotides, preferably labelled at one end are resolved by gel electrophoresis to provide a fingerprint of the target sequence which can be used to identify the site of cleavage, often with single nucleotide resolution. Increasingly, deep sequencing is replacing gel electrophoresis to define the population of nucleic acids present following digestion. Fingerprints can also be obtained by comparing with fingerprints of known strains that have been predicted by computer modelling to define the most probable structure. This method is efficient for genomes with more than 95 % homology (De Wit, 2000) but a combination of laboratory and the theoretical approaches provide the most secure RNA secondary structure predictions overall (Daou-Chabo and Condon, 2009).

Table 3. The cleavage specificity of several enzymatic probes on RNA molecule (Ehresmann *et al.*, 1987).

Probes	Specificity	Reference
RNase T1	3'-end of unpaired G residues.	(Pomerantz *et al.*, 1993)
RNase T2	Cleavage 3'-end of all four residues but preferentially 3'-end of A residues.	(Daou-Chabo and Condon, 2009)
RNase U2	Cleavage 3'-end of unpaired A residues.	(Daou-Chabo and Condon, 2009)
RNase CL3	Unpaired C>>A>U or cleaves 39 to C-residues.	(Daou-Chabo and Condon, 2009)
Nuclease S1	Unpaired N.	(Daou-Chabo and Condon, 2009)
RNase A	Cleaves 39 to C- or U-residues.	(Daou-Chabo and Condon, 2009)
RNase (II)	Unpaired N.	(Cheng and Deutscher, 2005)
RNase R	Unpaired N.	(Cheng and Deutscher, 2005)
N.c. nuclease	Unpaired N.	(Ehresmann *et al.*, 1987)
RNase V1	Paired or stacked N.	(Ehresmann *et al.*, 1987)

1.7.2. Nuclear magnetic resonance (NMR).

NMR results from atomic nuclei absorbing energy and re-emitting it as detectable radio waves when a sample is placed in a strong and stable magnetic field. The orientation of the nuclei, which respond to the magnetic field like small-scale bar magnets (Hine and Martin, 2015) changes when magnetised and certain chemical bonds have specific resonance frequencies whose activity changes depending on their microenvironment allowing a map of residues that are close together in solution. As a result, nuclear magnetic resonance (NMR) can be used for determining the molecular structure of small sized ribonucleic acids (~20kDa) as well as their dynamics in solution (Flinders and Dieckmann, 2006). Quantitative and qualitative analyses have been performed by NMR

technique of specific sites on certain RNA secondary structures such as domain II of the 5' untranslated region of the HCV genome, which has a molecular size of 25kDa and binds to the 40S ribosomal subunit (Lukavsky *et al.*, 2003).

Moreover, a combination of NMR with differential RNA labelling (13C- and 15N-labeled) methods can be an effective procedure for examining of the conformations of smaller labelled RNA portions within the context of a large RNA (Tzakos *et al.*, 2007). NMR can be used to demonstrate molecular conformation in solution and is used for studying physical properties at the molecular level, such as conformational exchange, phase changes, solubility, and diffusion (Mohammed and Segni, 2014).

1.7.3. Electron microscopy.

An electron microscope is used to image structures down to an atomic scale (Sousa and Leapman, 2012). The electron beam used as a source of illumination has a much greater resolving power than the ordinary light microscope as the wavelength of the beam is much smaller. This technique can be used for identifying the causative agent of an unknown infection as no specific reagent is needed to act as a probe (Goldsmith and Miller, 2009). Knoll and Ruska are credited with the development of the first electron microscope in 1932 (Dowle and Miriam, 2014).

Developments in the field have resulted in the production of several types of EM such as scanning electron microscope (SEM), which depends on analyses of the signals that are emitted from the surface of the specimen, whereas in the transmission electron microscope (TEM) the signals for analysis are those that are transmitted through the specimen. Scanning transmission electron

microscope (STEM) is a development of TEM in which the electron beam is scanned across the sample and the scintillators, positioned beneath the sample in the microscope column, are used for recognition of the transmitted, scattered electrons pixel-by-pixel.

This has been followed by development of other EMs such as aberration-corrected transmission electron microscopy, environmental transmission electron microscopy and cryo-electron microscope (Dowle and Miriam, 2014). These types of EM have significant roles in virus investigation and identification as they allow direct visualisation of viral structures and of viral replication complexes in cells (Vale F. *et al.*, 2010). Cryo-electron microscopy in particular has emerged as a very powerful technique for providing structural information of RNA folding as it can visualise individual molecules and average structurally identical molecules in a reconstruction process to yield a high resolution image **Figure 16** (Baird *et al.*, 2010).

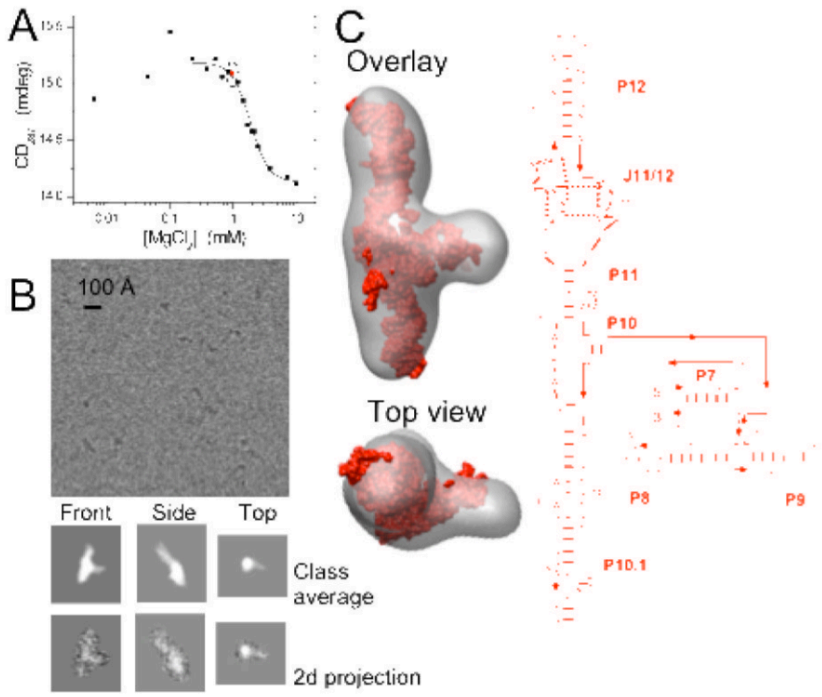

Figure 16. Cryo-EM and SPR of the S-domain intermediate structure (Baird *et al.,* 2010) . (A) Folding monitored by CD spectroscopy under conditions similar to cryo-EM studies. (B) Selected class averages of the intermediate compared with 2D projections from the 3D cryo- EM reconstruction. (C) SPR of the intermediate overlayed to an atomic model and secondary structure. The final 3D map was generated using 60 class averages from ~11,600 particles.

1.7.4. Mutagenesis.

Mutagenesis refers to the process in which a change of virus genetic information occurs. This change can be occur spontaneously in nature or be due to introduction of a mutagenic agent to plants, animals and humans. In other words, it can happen naturally due to exposure to a contaminating material or artificially by introducing mutagenic materials to the test sample. On a molecular basis these mutations includes nucleotide substitution, deletion, insertion and recombination (Beale, 1993, Liu and Naismith, 2008).

Mutagenesis has an important role in the determination of RNA secondary structures that are related to key biological processes be that process protein production or function. Common mutagenesis methods are the site-directed mutagenesis methods such as Kunkel mutagenesis (Kunkel, 1985), Quickchange (Liu and Naismith, 2008), modified PCR (Cadwell and Joyce, 1992) and inverse PCR (Dominy and Andrews, 2003).

These latter procedures significantly increase the efficiency of single mutations and also allow large single insertions, deletions/truncations and multiple mutations in only one experiment, which is generally not the case for the standard Quickchange methods (Liu and Naismith, 2008).For example, in viral replication (Ito and Lai, 1997) following the prediction of an RNA secondary structure, a site-direct mutagenesis can be used to change nucleotides predicted to be critical to loop or stem formation and their effect on biological activities can then be measured. For example, the 3'-terminal 42 nucleotides of the MHV genome has been analysed in this way (Johnson *et al.*, 2005) and the role of these changes in virus pathogenicity has also been assayed (Goebel *et al*., 2007).

1.8. Aim of the project.

Due to the limitation of knowledge about the role of RNA secondary structure in the virus replication cycle and pathogenicity, an investigation is made of critical folded structures found in the HKU1 coronavirus genome. To enable this several approaches have been used or developed. These include:

1- Construction of the full length 5' untranslated region, 3' untranslated region and frameshift element of HCoV-HKU1, addition of reporter genes, cloning

them into the appropriate plasmid vectors and introducing them into *E. coli* competent cells and other cell lines.

2- Prediction of secondary RNA structure for the unique 5' untranslated region, 3' untranslated region and frameshift element in all HCoV-HKU1 sequences.

3- RNA-ribosome interaction using electro mobility shift assay and kinetic protein expression.

4- Mutagenesis of in HCoV-HKU1 5'-UTR, 3'-UTR and the ORF1a-1b frameshift signal in order to test the importance and function of each of the described stem-loop structures.

The experiments described aim to address what these secondary structures do and to measure their efficiency so that their role in the virus life cycle can be deduced or improved. It is hoped that, in the longer term, a fuller description of the replication cycle may lead to new approaches to inhibition leading to a better control of coronavirus infection.

2. Materials and Methods.

2.1. Plasmid DNA vector.

The pTriEx1.1 vector (5301 bp) used in this study for cloning the full length of constructed 5'-UTR, 3'-UTR and FS of HCoV HKU1 was a gift from Prof. Ian Jones lab. The vector contains three promoters, bacterial, mammalian and insect upstream of the cloning cassette. In addition, this vector includes the ampicillin resistance gene for positive colony isolation, HSV tag sequence upstream of the cloning site which is derived from the glycoprotein D precursor envelope protein, so it is unlikely to interfere with protein structure or function and 8X His tag downstream of the cloning site which allowed the construction of N-terminal HSV-tagged and C-terminal His-tagged fusion proteins or both for protein expression detection **Figure 17**.

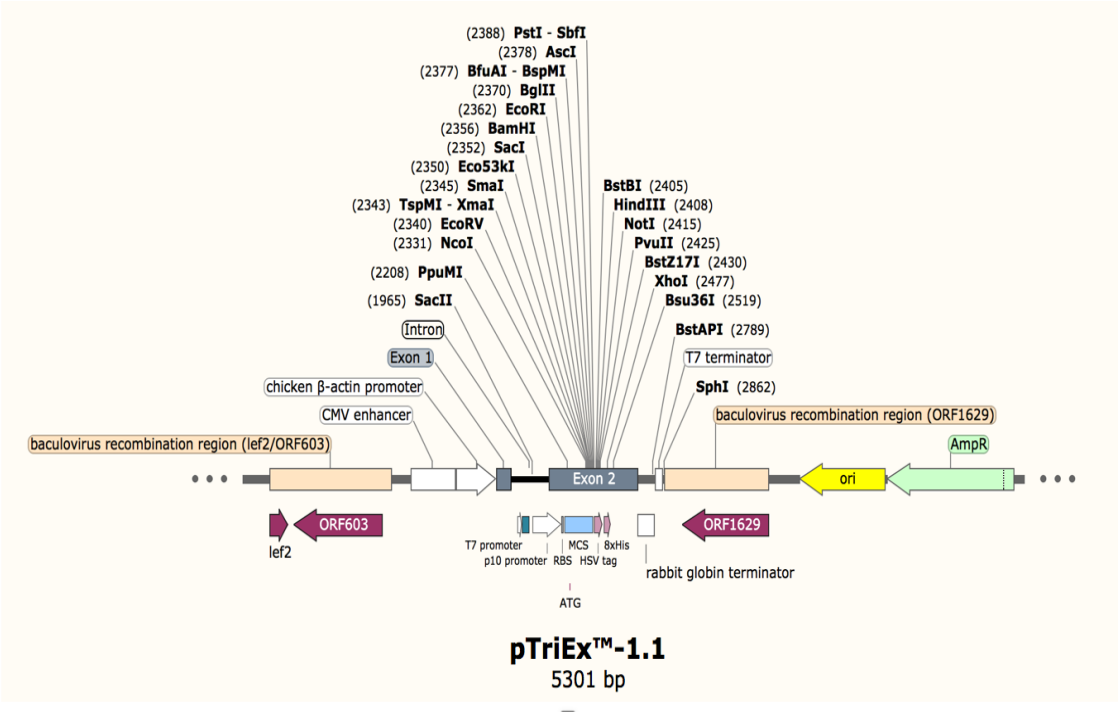

Figure 17. The vector map of pTriEx1.1 showing the multiple coning site and the most common structures. Key features of the plasmid are marked.

2.2. (100 mg/mL) Ampicillin stock solution.

For (100 mg/mL) Ampicillin stock solution, 1g of ampicillin sodium salt (Sigma) was weighed out and added to 10 mL of ddiH2O and allowed to stand for about 5 min until the Ampicillin was completely dissolved. The solution was filter sterilised with the 0.22 µm filter unit, aliquoted into 1.5 mL Eppendorf tubes, labelled and store at -20 °C.

2.3. (5X) TBE buffer stock solution.

This buffer was prepared by dissolving 54 g of Tris base, 27.5 g of boric acid and 20 mL of 0.5 M EDTA (pH 8.0). ddiH2O was added to make 1 litre of 5X stock solution with the adjustment of the pH of the solution to 8.3 using concentrated HCl, filtration with a 0.22-µm filter unit and stored at room temperature. This solution was diluted to 1X or 0.5X for electrophoresis or EMSA.

2.4. Sypro Ruby EMSA protein gel stain.

According to Invitrogen protocol 100 mL of the thawed Sypro Ruby EMSA protein gel stain (Component B) was added to TCA (Component C), and gently mixed for 5 min in order to ensure the dissolution of TCA. This was added to the bottle containing the remaining Sypro Ruby EMSA stain, mixed by inverting the bottle for ten times and stored at room temperature without exposure to direct sun light ready to use for at least 6 months.

2.5. Sybr green EMSA gel stain.

For the day's use of 1X Sybr green EMSA gel stain stock, 15µl of the 10,000X concentrate (Component A) (Invitrogen) was added to 150 mL of 0.5 X TBE buffer (89 mM Tris base, 89 mM boric acid, 1 mM EDTA, pH ~8.0) in a plastic bottle, sealed with an aluminium foil in order to protect it from direct light exposure and stored at 4 ^{o}C ready to use.

2.6. Complete genome sequences of HCoV-HKU1.

A total of 30 complete genome sequences of HCoV-HKU1 were downloaded from The National Centre for Biotechnology Information web site (NCBI) in order to predict the secondary and pseudoknot structures in the 5'-end UTR, 3'-end UTR and 1ABF. The GenBank accession number of all complete HCoV-HKU1 genomes used in this thesis has been list in the **Table 4**.

Table 4. The GenBank accession numbers for complete HCoV-HKU1 genomes.

Strain number	GenBank Accession No.
Reference (Ref)	NC_006577.2
Caen1	HM034837.1
N20 genotype C,	DQ41897.1
N21 genotype C,	DQ415898.1
N22 genotype C,	DQ415899.1
HKU1-5	KF686340.1
HKU1-10	KF686341.1
HKU1-11	KF686342.1
HKU1-12	KF686346.1
HKU1-13	KF686343.1
HKU1-15	KF686344.1
HKU1-18	KF430201.1
Genotype A	AY597011.2
N 3 genotype A,	DQ415903.1
N 6 genotype A,	DQ415904.1
N 7 genotype A,	DQ415905.1
N 9 genotype A,	DQ415906.1
N10 genotype A,	DQ415907.1
N11 genotype A,	DQ415908.1
N13 genotype A,	DQ415909.1
N14 genotype A,	DQ415910.1
N18 genotype A	DQ415914.1
N19 genotype A,	DQ415896.1
N 23 genotype A,	DQ415900.1
N 24 genotype A,	DQ415901.1
Genotype B,	AY884001.1
N15 genotype B,	DQ415911.1
N 25 genotype B,	DQ415902.1
N16 genotype C,	DQ415912.1
N17 genotype C,	DQ415913.1

2.7. Secondary and pseudoknot structure prediction.

After removing the identical sequences found for the 5'-end UTR, 3'-end UTR and 1ABF HCoV-HKU1 sequences, secondary structure prediction was performed depending on the unique sequences using of the Jalview, for alignments, and Mfold server web site (http://www.bioinfo.rpi.edu/applications/mfold) for RNA folding (Zuker, 2003). RNA secondary structure annotation of the frameshifting signal with slippery sequence of HCoV-HKU1 was performed manually based on a comparison of coronavirus frameshift signals in (Woo *et al.*, 2005b) and this can be seen in **Figures 22**, **23** and **24**.

2.8. Oligonucleotide primers.

Oligonucleotide primers were divided into two groups, the first groups were designed manually in order to construct by overlapping PCR the full sequence for the 5'-UTR, 3'-UTR and the frameshift regions. These oligonucleotides are 45bp in length with 15-nt as the overlapping part. The second group were designed using the Clontech web site (www.clontech.com) for infusion cloning of the overlapped fragments and as DNA sequencing primers. All these oligonucleotide primers were ordered from Integrated DNA Technology (IDT) as listed in the **Tables 5**, **6**, **7** and **8**.

Table 5. Oligonucleotide primers used for constructing the 5'-UTR, 3'-UTR and Frameshift in HCoV-HKU1 genome.

Name	Primer	bp	Tm	GC%
5'_UTR_Fw1	5'-TAATACGACTCACTATAGGGGAGTTTGAG CGATTGACGTTCGTAC-3'	45	65.1	44.4
5'_UTR_Rv1	5'- ATCTGACAAGAGATCGTAAGCTGATAGA CGGTACGAACGTCAATC-3'	45	65.2	44.4
5'_UTR_Fw2	5'- GATCTCTTGTCAGATCTCATTAAATCTAA ACTTTTTAAACAAGAT-3'	45	58.8	26.7
5'_UTR_Rv2	5'- AACCACACTCACAAGCATGGATAACAGG GAATCTTGTTTAAAAAG-3'	45	64.1	37.8
5'_UTR_Fw3	5'- CTTGTGAGTGTGGTTTAATCATAATCTTG TATTTTACTTTCCACA-3'	45	60.9	31.1
5'_UTR_Rv3	5'- CCAACACGTCACTGGCAGAGAGATGAAA AGTGTGGAAAGTAAAAT-3'	45	65.6	42.2
5'_UTR_Fw4	5'-CCAGTGACGTGTTGGTTGTCCTCAGCGTC CCTCCCATAGGTCGCA-3'	45	72.3	60
5'_UTR_Rv4	5'-GATGGTGGTGCTCGAGTGCGACCTATGG GAGGGA-3'	34	72	61.8
3'_UTR_Fw1	5'-TAATACGACTCACTATAGGGTGAGAATGA ATCCTAATTCGACACT-3'	45	62.8	37.8
3'_UTR_Rv1	5'-ATTCCGAATAATAGCGAGGGGTTACCAC CTAGTGTCGAATTAGGA-3'	45	66	44.4
3'_UTR_Fw2	5'-GCTATTATTCGGAATAGGACACTCTCTAT CAGAATGAATTCTTGC-3'	45	62.1	37.8
3'_UTR_Rv2	5'-CTGTAACAACCTACTCTATCTGTTATTACA GCAAGAATTCATTCT-3'	45	61	33.3
3'_UTR_Fw3	5'-AGTAGGTTGTTACAGACTATATATTAATTA GTAGAAATTTTATAT-3'	45	55.6	20
3'_UTR_Rv3	5'-TTATAACTACTCTAACAATCAAATGTCTAA ATATAAAATTTCTAC-3'	45	55.6	20
3'_UTR_Fw4	5'-TTAGAGTAGTTATAAGGTTTAGCTGTAGTA TAAACGCCTCCGGGA-3'	45	63.7	40
3'_UTR_Rv4	5'-TATATATTAAACACTACAATTGATAGCTCT TCCCGGAGGCGTTTA-3'	45	62.5	35.6
3'_UTR_Fw5	5'-AGTGTTTAATATATATATTAGTATATGATT GAAATTAATTATAGC-3'	45	53.5	15.6
3'_UTR_Rv5	5'-GTAATTCCTCCAAAAGGCTATAATTAATT TC-3'	31	53.4	29
FS_Fw1	5'-TAATACGACTCACTATAGGGTTTAAACGG GTTCGGGGTACTAGTG-3'	45	64.9	44.4
FS_Rv1	5'-CACTAGCACAGGGTACTAGCCGGGCATTC ACACTAGTACCCCGAA-3'	45	69.9	55.6
FS_Fw2	5'-TACCCTGTGCTAGTGGTTTATCTACTGATG TTCAATTAAGGGCAT -3'	45	64.7	40
FS_Rv2	5'-GATGGTGGTGCTCGAGAAATGCCCTTAAT TGAACATC-3'	37	108	45.9

Table 6. Oligonucleotide primers used for the infusion cloning of the 5'-UTR, 3'-UTR and Frameshift in HCoV-HKU1 genome.

Name	Primer	bp	Tm	GC%
5'_Inf_Fw	5'- AGGAGATATACCATGGTAATACGACTCACTA TAGGG -3'	36	102	41.7
5'_Inf_Rv	5'- GATGGTGGTGCTCGAGTGCGACCTATGGGAGG GA -3'	34	110	61.8
3'_Inf_Fw	5'- AGGAGATATACCATGGTAATACGACTCACTATA GGG -3'	36	102	41.7
3'_Inf_Rv	5'- GATGGTGGTGCTCGAGGTAATTCCTCCAAAAG GCTA -3'	36	108	50
FS_Inf_Fw	5'- AGGAGATATACCATGGTAATACGACTCACTATA GGG -3'	36	102	41.7
FS_Inf_Rv	5'- GATGGTGGTGCTCGAGAAATGCCCTTAATTGAA CATC-3'	37	108	45.9

Table 7. Oligonucleotide primers used for sequencing the 5'-UTR, 3'-UTR and Frameshift in HCoV-HKU1 genome.

Name	Primer	bp	Tm	GC%
PTriEx_Up	5'- GGTTATTGTGCTGTCTCATCA -3'	21	52.6	42.9
PTriEx_dowm	5'- TCGATCTCAGTGGTATTTGTG -3'	21	52.1	42.9

Table 8. Oligonucleotide primers used for constructing the mCherry + Frameshift + EGFPHIS, 5'-UTR + EGFPHIS and EGFPHIS + 3'-UTR respectively in HCoV-HKU1 genome.

Name	Primer	bp	Tm	GC %
HSVmCherry-Fw1	5'-GAGTTAATCCGGGACCTCAGCCAGA ACTCGCCCC -3'	34	69	61.8
HSVmCherry-Rv1	5'- TTTAAAAACTTGTACAGCTCGTCCAT GCC-3'	29	59.4	41.4
Frameshift-Fw1	5'-GTACAAGTTTTTAAACGGGTTCGGGG TACTAGTGT -3'	35	62.2	42.9
Frameshift-Rv1	5'-CTCACCATAAATGCCCTTAATTGAAC ATCAGTAGATAAACC -3'	41	61	36.6
EGFPHIS-Fw1	5'- GGCATTTATGGTGAGCAAGGGCGA G -3'	25	62.2	56
EGFPHIS-Rv1	5'- GTTACATATGGGCATATGTTGCTCA GTGATGGTGATGGTGATGGTGG -3'	47	67	46.8
T75-5'-UTR-Fw1	5'- GGCTCGGGGCTGTCCGCGGTAATAC GACTCACTATAGGGGAGTTTGAGCG -3'	50	71.9	60
T75-5'-UTR-Rv1	5'-CTCACCATCATTGCGACCTATGGGAG GG-3'	28	63.8	57.1

T75-EGFPHIS-Fw1	5'-CGCAATGATGGTGAGCAAGGGCGAG-3'	25	64.2	60
T75-EGFPHIS-Rv1	5'-GTTACATATGGGCATATGTTGCTCAGTGATGGTGATGGTGATGGTGG-3'	47	67	46.8
EGFPHIS-3'-UTR-Fw1	5'-GAGTTAATCCGGGACCTATGGTGAGCAAGGGCGAG-3'	35	67.2	57.1
EGFPHIS-3'-UTR-Rv1	5'-CATTCTCATCAGTGATGGTGATGGTGATGGTGG-3'	33	63	48.5
3'-UTR-Fw1	5'-TCACTGATGAGAATGAATCCTAATTCGACACTAGGT-3'	36	61	38.9
3'-UTR-Rv1	5'-GGAGAATTTCCTCCGCATGCCAAAAAACCCCTCAAGACCCGT-3'	42	69.5	52.4

Table 9. Oligonucleotide primers used for the side direct mutagenesis of the mCherry + Frameshift + EGFPHIS respectively in HCoV-HKU1 genome.

Name	Primer	bp	Tm	GC%
HSVmCherry.stop-Fw1	5'-GGCGGCATGGACGAGCTGTACTAGTTTTTAAACGGGTT -3'	38	69	50
HSVmCherry.stop-Rv1	5'-AACCCGTTTAAAAACTAGTACAGCTCGTCCATGCCGCC-3'	28	59.4	50
No-TT-Fs.Fw	5'- CGAGCTGTACAAGTTTAAACGGGTTCGGGG-3'	30	63.8	53.3
No-TT-Fs.Rv	5'-CCCCGAACCCGTTTAAACTTGTACAGCTCG-3'	30	63.8	53.3

Table 10. Oligonucleotide primers used for sequencing the mutation in the mCherry + Frameshift + EGFPHIS respectively in HCoV-HKU1 genome.

Name	Primer	bp	Tm	GC%
SacII-Fw	5'- TGGCTGCGTGAAAGCCTTG-3'	19	59	57.9
RFP-Fw	5'-GGCGAGATCAAGCAGAGGCTGAA-3'	23	61.4	50
GFP-Fw	5'-ATGGTGAGCAAGGGCGAGGAG -3'	21	60.8	54.6

Table 11. Oligonucleotide primers used for the overlapping PCR to construct mutations of the mCherry + Frameshift + EGFPHIS respectively in HCoV-HKU1 genome.

Name	Primer	bp	Tm	GC%
Stem1.AMut.fw	5'-AAAAATAATAATGTGAATGCCCGGCTAGTACCCTGTGCTAGTGG -3'	44	70	43
Stem1.AMut.Rv	5'-CACATTATTATTTTTAACCCGTTTAAAAACTTGTACAGCTCGTCCATGCCGCCGGTG-3'	57	72	42

Stem1BMut.Fw	5'-AATAATAAAATATGCTAGTGGTTTATCTACTG ATGTTCAATTAAGGGC-3'	48	62	29
Stem1BMut.Rv	5'-CTAGCATATTTTATTATTCGGGCATTCACAC TAGTACCCCG-3'	41	63	41
Stem2AMut.Fw	5'-TAAATAAAAGGCTAGTACCCTGTGCTAGTG G-3'	31	60	42
Stem2AMut.Rv	5'-CTAGCCTTTTATTTACACTAGTACCCCGAAC CCGTTTAAAAACTT-3'	45	63	38
Stem2BMut.Fw	5'-AAAAATTTATGGTGAGCAAGGGCGAGGAGC TGTTCACCGG-3'	40	71	50
Stem2BMut.Rv	5'-TTGCTCACCATAAATTTTTTTAATTGAACATC AGTAGATAAACCACTAGCACAGG-3'	55	60	33

Table 12. Oligonucleotide primers used for the overlapping PCR to construct mutations of the 5'-UTR + EGFPHIS respectively in HCoV-HKU1 genome. Red nucleotides color indicates mutated sequence.

Name	Primer	bp	Tm	GC%
Mut5-Fw	5'- GGGCTGTCCGCGGTAATACGACTCA -3'	25	64.2	60
Mut5-Rv	5'- CATTGCGACCTATGGGAGGGGAGG -3'	24	63.1	62.5
Mut.5.1A-Fw1	5'-AAAAAATTAAAATTCGTACCGTCT -3'	24	49.7	25
Mut.5.1A-Rv1	5'-AATTTTTAATTTTTTAAACTCCCCTATAGTG-3'	30	51.4	23.3
Mut5.2A-Fw1	5'- AATATATAAAATTACGATCTCTTGTCAGA -3'	29	50.9	24.1
Mut5.2A-Rv1	5'- GTAATTTTATATATTGTACGAACGTCAATC -3'	30	51.5	26.7
Mut5.3A-Fw1	5'-AAATATCTTGTCAGATCTCATTAAATC -3'	27	49.8	25.9
Mut5.3A-Rv1	5'-TCTGACAAGATATTTTAAGCTGATAGAC-3'	28	52.9	32.1
Mut5.4A-Fw1	5'-TTATAAAATCTCATTAAATCTAAACTTTTTAA AC-3'	34	50	14.7
Mut5.4A-Rv1	5'-AATGAGATTTTATAAGAGATCGTAAGCTGA-3'	30	54.4	30
Mut5.5A-Fw1	5'-AAAAATCTAAACTTTTTAAACAAGATTCC-3'	29	50.4	20.7
Mut5.5A-Rv1	5'-AAAGTTTAGATTTTTTGAGATCTGACAAG-3'	29	52.8	27.6
Mut5.6A-Fw1	5'-AAAAAAATTTTTAAACAAGATTCCCT-3'	26	49.3	19.2
Mut5.6A-Rv	5'-TTTAAAAATTTTTTTTTTAATGAGATCTG AC-3'	31	50.1	16.1
Mut5.7A-Fw1	5'-AAAAAACAAGATTCCCTGTTATCC-3'	24	51.8	33.3
Mut5.7A-Rv1	5'-GGAATCTTGTTTTTTAAGTTTAGATTTAATG-3'	31	51.2	22.6
Mut.5.8A-Fw1	5'- AAAAAATTACCTGTTATCCATGCTT-3'	25	51.7	28
Mut.5.8A-Rv1	5'-AACAGGTAATTTTTTTTAAAAAGTTTAGAT-3'	30	50.3	16.7
Mut5.9A-Fw1	5'-ACAATTATTTATAATAGGGAATCTTGTTTA-3'	30	51.3	23.3
Mut5.9A-Rv1	5'-ATTATAAATAATTGTGAGTGTGGTTTAATC-3'	30	50.1	20
Mut5.10A-Fw1	5'-AATATAATTTAATAATAATATTATATTTTACTT TCCACA-3'	39	50.3	10.3
Mut5.10A-Rv1	5'-TGTGGAAAGTAAAATATAATATTATTATTAA ATTATATT-3'	39	50.3	10.3
Mut5.11A-Fw1	5'-AATTTAAAAAATTTTAATCTCTCTGCCAGT-3'	30	53.1	23.3
Mut5.11A-Rv1	5'-ATTAAAATTTTTTAAATTAAAATACAAGATTA TG-3'	34	48.3	8.8
Mut5.12A-Rv1	5'-CATTGCGACCTATGGGAGGTATTTTTATTATA ATTAATATTTTATTGGCAGAGAGATG-3'	58	51.7	18

2.9. SOC medium.

This medium was supplied with the competent cells and was used in the final step of bacterial cell transformation to obtain maximal transformation efficiency of *Escherichia coli* (Clontech). SOC medium is supplied in ten 10 mL bottles of liquid medium with the following composition: 2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose.

2.10. Luria-Bertani (LB) Broth.

This broth was used for culturing E. coli and it was prepared by addition of LB Broth Base (Miller's LB Broth Base (Invitrogen) to 1 L D.W then autoclaved at 115oC for 15 minutes. For most applications ampicillin (Sigma-Aldrich) was added to the broth to a final concentration of 100 µg/ml. This selects the culture of *E. coli* transformed with plasmids encoding β-lactamase for ampicillin resistance.

2.11. Luria-Bertani (LB) Agar.

Agar plates were prepared by using LB-broth containing 1.5 % of Bacteriological agar (Miller) to 1L D.W then autoclaving at 115^{o}C for 15 min.

2.12. Construction of full-length of 5'-UTR, 3'-UTR and Frameshift fragments with and without the reporter gene.

An overlapping polymerase chain reaction was used to build full-length fragments of the HKU1 5'-UTR; 3'-UTR and orf1a/1b frameshift region by following the multistep strategy illustrated in **Figure 18**. CloneAmp HiFi PCR premix (Clontech) was used due to its high sensitivity, specificity, priming efficiency and extension efficiency. Each PCR step was 50µl of reaction mixture and it includes 25µl of CloneAmp HiFi PCR premix (2X), 2.5µl of each primer (0.3µM), 100ng DNA (template), and ultra-pure water to reach the final volume. This procedure was used in the same way for constructing fragments encoding 5'-UTR+ eGFP; eGFP+3'-UTR and mCherry+frameshift+eGFP DNA and their subsequent mutations **Table 5**, **8** and **11** respectively.

The PCR reaction was carried out as follows:

Step	Temperature	Time	Number of cycle
Initial denaturation	98 °C	30 sec	1 cycle
Denaturation	98 °C	30 sec	30 cycles
Annealing	66 °C	30 sec	
Extension	72 °C	30 sec	
Final extension	72 °C	5 min	1 cycle

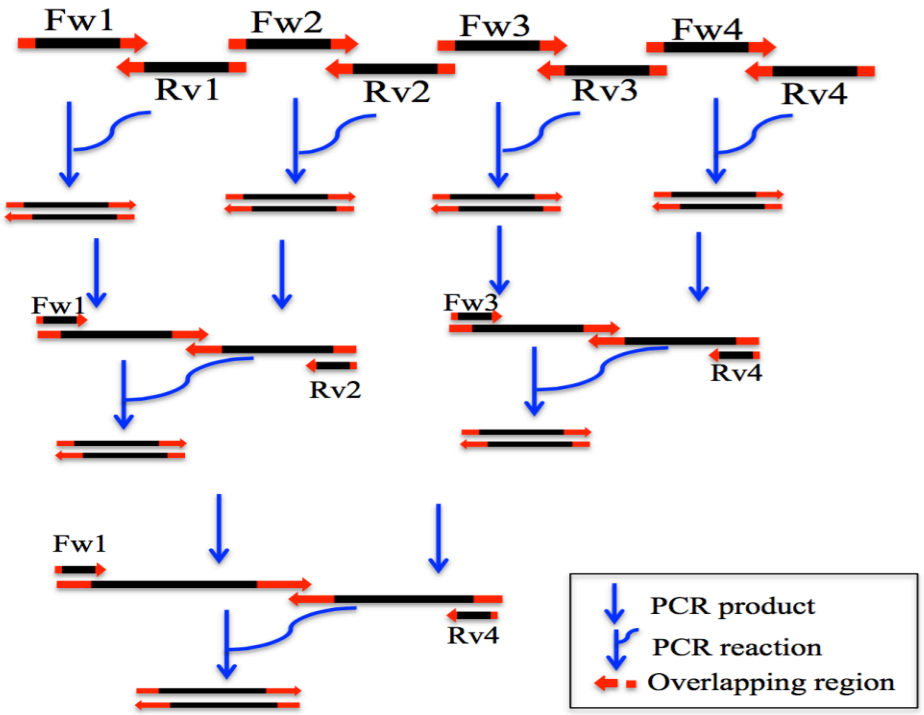

Figure 18. The multistep strategy used to construct the full-length of each 5'-UTR; 3'-UTR and Frameshift DNA fragment in HCoV-HKU1.

2.13. Purification of PCR product.

During each step for 5'-UTR; 3'-UTR and Frameshift DNA fragment construction with and without the gene reporter and their mutations. PCR products were purified by using a NucleoSpin Gel and PCR clean up kit according to the supplied protocol (Clontech). Ultra-pure water was used for PCR product elution and stored at -20°C. The concentration of product was determined using Nanodrop spectrophotometer (ND-1000).

2.14. Agarose gel electrophoresis.

Construction progress for 5'-UTR; 3'-UTR and Frameshift DNA fragment with and without the gene reporter and their mutations were visualised by agarose gel electrophoresis in 1.8 % agarose and 1 x Tris borate EDTA (TBE) buffer (Invitrogen) with 0.5µg gel red per ml of agarose gel. DNA was visualised by exposure of the gel to 260 nm UV illumination provided by a short-wave length UVP GelDoc-it ultraviolet transilluminator Bio imaging system. Low molecular weight DNA ladder (New England Biolabs) and Hyper ladder 1 (Bio line) were used to establish DNA band sizes.

2.15. Infusion cloning reaction for 5'-UTR; 3'-UTR and Frameshift DNA fragment with and without the reporter gene.

The infusion process was used for subcloning of each 5'-UTR; 3'-UTR and Frameshift DNA fragment in to pTriEx 1.1 vector by creating 15 bp of overlap at the terminal site of the cloning insert and linearized cloning vector (Clontech). Then used in the same way for subcloning of each 5'-UTR+eGFP; eGFP+3'-UTR and mCherry+frameshift+eGFP DNA fragment in to pTriEx 1.1 vector **Table 6**.

2.15.1. PCR amplification of cloning inserts.

High fidelity polymerase chain reaction was used for amplification of cloning insert (5'-UTR; 3'-UTR and Frameshift DNA fragment) through the use of infusion primers in **Table 6** with the constructed DNA fragment in section (2.2.12). CloneAmp HiFi PCR premix (Clontech) was used due to high sensitivity, specificity, priming efficiency and extension efficiency of CloneAmp HiFi PCR polymerase. The reaction mixture was included 25µl of CloneAmp HiFi PCR

premix (2X), 3µl of each primer (0.3µM), 250ng DNA (template), and ultra-pure water to reach the final volume (50µl).

PCR was carried out as follows:

Step	Temperature	Time	Number of cycle
Initial denaturation	95 °C	3 min	1 cycle
Denaturation	95 °C	10 sec	30 cycles
Annealing	55 °C	10 sec	
Extension	72 °C	30 sec	
Final extension	72 °C	5 min	1 cycle

2.15.2. Linearization of cloning vector by restriction digestion.

Double digestion of the plasmid DNA was performed according to the manufacturer's instruction protocol (Clontech) with the use of different fast-digest restriction enzymes depending on the preferred restriction location (Thermo Scientific). The reaction mixture comprised 10µl of pTriEx1.1 (1µg), 1µl from each restriction enzyme, 2µl of Green buffer (10X) and ultra-pure water to reach the final volume (20µl). The reaction was carried out for 30 min at 37°C by using PCR T100^{TM} thermal cycler.

2.15.3. Cloning DNA of interest into the linearized vector.

Cloning reactions for each cloning insert (5'-UTR; 3'-UTR and Frameshift DNA fragment) section (2.2.15.1) with the linearized vector (pTriEx1.1) section 2.2.15.2 with and without the gene reporter was performed according to manufacturer's instruction protocol (Clontech) with the use of Clontech online tool to estimate the molar ratio for the insert and vector (1:3). The cloning reaction

included 2µl of 5X In-Fusion HD Enzyme Premix, Xµl of Linearized Vector, Xµl of Insert and Xµl of ultra-pure water to reach the final volume (10µl) depending on the molar ratio result. The reaction mixture was then mixed well and incubated for 15 min at 50°C using a water bath, then placed on ice for 30 min.

2.15.4. Transformation.

Stellar ™ Competent Cells (*E. coli* HTS08 strain) from Clontech were incubated with 2µl plasmid DNA for 30 min on ice then heat-shocked at 42°C for 45 sec. After that cells were immediately returned to ice for 5 min, thereafter 125µl pre-warmed SOC medium was added to these cells and incubated in a 37°C shaking incubator (226 rpm) for 1 hr. After incubation, transformed cells were plated on LB/Ampicillin agar plates and incubated at 37°C for 24 hr. Finally, 30 colonies from the previous culture were selected and sub-cultured in LB/ Ampicillin agar plates at 37°C for 12-16 hrs for further study.

2.16. Colony PCR analysis of transformations.

Colony PCR was performed in order to detect the plasmid DNA that contains the target DNA (5'-UTR, 3'-UTR and Frameshift). A small amount of bacterial colony was taken from LB/Ampicillin agar plate by using a sterile tip and diluted into 30µl ultra-pure water and incubated at 100°C for 2 min. After incubation, the boiled mixture was centrifuge at 13000 rpm for 5 min and 2µl from the supernatant was use as template for the PCR. The reaction mixture was included 5µl of Go Taq Green colour premix (2X), 1µl of plasmid upstream primer (0.3µM), 1µl of target DNA downstream primer (0.3µM), 2µl from the supernatant as DNA template and ultra-pure water to reach the final volume (12.5µl). Colony

PCR was carried out as follows:

Step	Temperature	Time	Number of cycle
Initial denaturation	95 °C	2 min	1 cycle
Denaturation	95 °C	30 sec	30 cycles
Annealing	55 °C	30 sec	
Extension	72 °C	1 min	
Final extension	72 °C	10 min	1 cycle

2.17. Small-scale plasmid DNA purification.

A single colony from a fresh LB agar plate containing ampicillin (100 µg/ml) was transferred into 5 ml of LB broth containing ampicillin (100 µg/ml) and incubated at 37°C shaking incubator (226 rpm) for 16 hrs. The GeneJET plasmid miniprep Kit (Thermo Scientific) was used for the DNA purification following the manufacturer's protocol. DNA was eluted with 30µl of ultra-pure water and the concentration of product was determined by the use of Nanodrop spectrophotometer (ND-1000) and store at - 20°C for further study.

2.18. Sequencing analysis of target DNA.

Sequencing analysis was performed according to the instructions on the Source BioScience website by sending 100ng/µl of pure plasmid DNA section (2.2.17) and 3.2pmol/µl of each primer for this purpose **Table 7** and **10**.

2.19. *In vitro* expression.

According to the manufacture's protocol (Promega) TNT® quick coupled transcription/translation system was used for *in vitro* expression reactions using each of the constructed plasmid DNAs (5'-UTR + eGFPHIS, HSVmCherry + Frameshift + eGFPHIS and eGFPHIS + 3'-UTR). The reaction mixture was including 40 µl TNT® quick master mix, 1 µl Methionine (1mM), X µl template DNA 0.5 µg and X µl nuclease-free water to 50 µl of the total volume. The mixture was incubated at 30°C (PCR thermocycler) for 60-90 minutes and store at - 20°C for further study.

2.20. Protein detection by western blot.

Western blots were carried out to detect protein expression by adding 1µl of the translation reaction to 19 µl of LDS sample buffer (980µl 4x LDS sample buffer (Invitrogen) and 20µl β- mercaptoethanol (Fisher)). The sample was incubated at 99.9°C for 10 min in a PCR thermocycler (Gene flow), centrifuged briefly to collect the contents in the bottom of the tube and 20µl from each sample was loaded onto a precast 4–20% gradient Tris-glycine SDS polyacrylamide gel (Invitrogen). The gels were placed into mini gel tank (Life Technologies) and electrophoresis carried out at 170 V for 32 min in 1X MES (50mM MES, 50mM Tris, 0.1% SDS, 1% EDTA, pH 7.25) running buffer (Invitrogen).

The gel was transferred to a PVDF membrane (Merck) pre-wetted in 100% methanol (Fisher) for 5 min then 15min into 1X transfer buffer from Fisher (10X Tris-Glycine Transfer Buffer (25 mM Tris, 192 mM Glycine pH~8.3) with 20% methanol by use of a semidry western blotting apparatus (ATTO) at 35 V and 150 mA for 1 hr 20 min.

Membranes were blocked overnight at 4°C in 50 ml of 1X TBST buffer [5% skimmed milk powder (MP), 0.05% Tween 20 (Fisher), TBS (Fisher) (50mM Tris-Cl, 150mM NaCl, pH 7.6)]. The membranes were then incubated at room temperature on a rocking platform (25-30 RPM) (Stuart) for 1 hr with a primary antibody, e.g. anti 6X His tag (rabbit polyclonal to His tag) or anti HSV tag (rabbit polyclonal to HSV tag) (Abcam) diluted 1:1000-2000 into 30ml of 1X TBST buffer [5% skimmed milk powder (MP), 0.05% Tween 20, TBS (50mM Tris-Cl, 150mM NaCl, pH 7.6)]. After 3 washes into 50 ml of 1X TBST buffer for a total of 15 min on rocking platform (25-30 RPM) (Stuart).

The membrane was incubated with secondary antibody horseradish peroxidase-conjugated goat anti–rabbit immunoglobulin G antibody (HRP) (Dako) diluted 1:10000 into 30ml of 1X TBST buffer. After 3 washes into 50 ml of 1X TBST buffer for a total of 15 min, bound antibody was visualized using enhanced chemiluminescence (ECL) western blotting detection solutions A (Lumiol) and B (Peroxide) (vol/vol) incubated for 5 minutes in accordance with the manufacturer's protocol (Amersham) then analysis by using gel imaging for fluorescence and chemiluminescence (G: BOX Chemi XX6).

2.21. Site directed mutagenesis.

According to the manufacture's protocol (Promega) Quickchange II site-direct mutagenesis was used to perform the desired mutations in the constructed plasmid DNA (HSVmCherry + Frameshift + eGFPHIS). The reaction mixture included 5µl of 10× reaction buffer, Xµl (5–50 ng) of dsDNA template, 1.25µl (125 ng) of mutagenic oligonucleotide forward primer with 1.25µl (125 ng) of

mutagenic oligonucleotide reverse primer (Table 9), 1µl of dNTP mix, 3µl of quick solution and nuclease-free water to a final volume of 50µl, then addition of 1µl of Pfu*Ultra* HF DNA polymerase (2.5 U/µl). The mixture was incubated in a PCR thermocycler (Gene flow) according to the following thermocycler conditions:

Step	Temperature	Time	Number of cycle
Initial denaturation	95 °C	1 min	1 cycle
Denaturation	95 °C	50 sec	18 cycles
Annealing	60-68 °C	50 sec	
Extension	68 °C	6 min *	
Final extension	68 °C	7 min	1 cycle

*1 minute/kb of plasmid length

The reaction was cooled on ice for 2 min then 1µl of *DpnI* endonuclease restriction enzyme was added and mixed with the reaction by pipetting up and down, followed by incubation at 37 °C for 1hr in order to digest the parental plasmid (DpnI cleaves the methylated DNA in the target sequence 5´-Gm6ATC-3´) then transformation of the mutant plasmid into XL10-Gold Ultra Competent Cells.

2.22. Transformation of XL10-Gold Ultracompetent Cells.

A 45 µl of the pre-thawed XL10-Gold ultracompetent cells were transferred into a 2 ml prechilled Eppendorf round-bottom tube on ice, addition of 2 µl of the β-ME mix provided with the kit to the 45 µl of cells and the cells incubated on ice

for 10 minutes with gentle mixing every 2 minutes. 2 µl of the *Dpn*I treated DNA from each control and sample reactions were added to individual aliquots of the ultracompetent cells. The transformation reactions were gently mixed and incubated on ice for 30 minutes. The cells were heat-shocked at 42°C and plated on LB–ampicillin agar plates as described.

2.23. Baculovirus expression.

To visualize mCherry and GFP expression in cells the baculovirus *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) expression system was used with vectors encoding the full sequence of HCoV-HKU1 mCherry+frameshift+eGFP and mutations thereof.

2.24. *S*f9 cells.

Spodoptera frugiperda, generally known as a fall armyworm, was used as a source for insect cell line *Spodoptera frugiperda* 9. This cell line was used for baculovirus amplification and protein expression. The Sf9 cell line was sustained in suspension culture consisting of Insect Xpress media with L-glutamine (BioWittaker) supplemented with 2% of foetal calf serum (FCS), penicillin (1000 units/L) and streptomycin (1000 µg/L). The cultures were cultivated in 25 and 75 Corning vented flasks at 28°C into an Innova 4430 shaker incubator (New Brunswick Scientific) that shaking at 99 rpm. The cell concentration was maintained in a range from $1.5x10^5$ to $2.5x10^6$ cell/ml through passaging every three to four days. The culture volume maintained per flask was estimated as one-fifth of the total volume of the flask.

2.25. Transfection of *Sf*9 cells and production of recombinant baculovirus.

A 6-well plate (NUNC) was seeded with $1x10^6$ Sf9 cells and incubated at room temperature for 30min to allow the cells to attach to the well. A transfection mix was prepared according to the manufacture protocol by combining of 400ng Bacmid (5µl) (Invitrogen) with 500ng of the transfer vector comprising the desired gene, made up to final volume of 12µl by using sterile water, then gently mixed with 12µl of Lipofectin mix (8µl of Lipofectin (Invitrogen) and 4µl of sterile water) and incubated at room temperature for 30min.

The cell line was washed two times with 1ml of serum free insect cell media (BioWittaker) then replaced with 2ml of serum free medium. 24µl of the transfection mixture (lipid-DNA complex) was added to each well. The plates were incubated overnight at 28°C then the media was replaced with 2% FCS insect cell media. The cell line plates were continue incubated for additional 5 days at 28°C then all the cells and culture media were harvested and centrifuged at 2000rpm for 10min. The supernatant which contains the recombinant baculovirus (P_0) was transferred to sterile falcon tube and store at 4°C for further passage.

2.26. Small scale protein expression using recombinant baculovirus system.

The small-scale protein expression was performed through infection of a 6-well plate (NUNC) which was pre-seeded with $1x10^6$ Sf9 cells and incubated at room temperature for 30 min to allow the cells to attach. The cell line was washed two times with 1ml of insect cells serum free media (BioWittaker) then replaced

with 2ml of Insect Xpress media with L-glutamine (BioWittaker) supplemented with 2% of foetal calf serum (FCS). Then 100µl of a high titre stock of the recombinant baculovirus, typically passage 2 or 3, was added to the well and incubated for 5 days at 28°C. The two-florescence proteins expression was checked during the incubation days by using digital inverted microscope (Fisher). Afterward all the cells and culture media were harvested and centrifuged at 2000 rpm for 10 min.

The cell pellets were used for fluorescence microscopy, western blot or were resuspended with 1ml of Facsflow buffer (Becton Dickinson), then transfer into a FACS tube for flow cytometry. Consistent with the manufactural protocol the fluorescence of 10000 cells was measured by using a FACScan flow cytometer (Becton Dickinson) and the data was analysed by using WinMDI 2.8 (Scripps Research Institute).

2.27. *In vitro* transcription using T7 RNA polymerase.

RNA transcripts of 5'-UTR, 3'-UTR and Frameshift sequences were produced by using the HiScribeT7 High Yield RNA Synthesis Kit (New England Biolabs). *In vitro* transcription was conducted in accordance with the manufacturer's protocol. After double digestion for the plasmid DNA as described in section (2.15.2), an additional 1.5 µl of T7 RNA polymerase mix was added to the reaction mixture that includes 1.5 µl 10X reaction Buffer (0.75X final), 1.5 µl each NTP (7.5 mM each final), X µl template DNA 1 µg and X µl nuclease-free water to 20 µl of the total volume. The mixture was incubated at 37°C (PCR thermocycler) for 16 hrs and stored at -20°C for further study.

2.28. Purification of the transcribed RNA.

RNA resulting from the enzymatic transcription reactions were subject to an RNA clean-up and concentration micro kit (Norgen's) that was used as a rapid method for micro RNA purification and concentration according to the manufacture's protocol without using phenol or chloroform. The volume of the RNA sample was adjusted to 100 μL by adding RNase-free water, 250 μL of binding buffer H and 200 μL of 96 – 100% ethanol were added to the RNA sample and mixed by vortexing for 10 seconds.

A 600 μL of the mixture was applied to the column and centrifuged at 13,000 RPM for 1 minute then the flow through was discarded. 500 μL of Wash Solution K was added to the column and centrifuged for 1 minute, this step was repeated for a second time by adding another 500 μL of wash solution K and centrifuging for 2 minutes, followed by another spin for 1 min in order to ensure that the column is dry.

The column was placed into a fresh 1.7 mL elution tube provided with the kit then 15 μL (higher concentrations of RNA) of RNase free water was added to the column and incubated at room temperature for 2 min. The sample was spun in the centrifuge for 2 minutes at 2,000 RPM, followed by 1 minute at 13,000 RPM. The concentration of RNA was determined by the use of Nanodrop spectrophotometer (ND-1000) and stored at - 80°C for further study.

2.29. SYBR gold gel staining of the transcribed RNA.

The transcribed RNA gel was stained according to the manufacture protocol by placing the gel in a square petri dish with the lid and 50 mL of 1X Syber gold stain (Invitrogen) was added to completely cover the gel. The staining solution was protected from light by covering it with aluminium foil and incubated at room temperature for 40 minutes (depending on the thickness of the gel and the percentage of agarose or polyacrylamide). The stained gel was then visualized by using gel imaging for fluorescence and chemiluminescence (G: BOX Chemi XX6).

2.30. 5' EndTag labeling of the transcribed RNA.

According to vector laboratories instructional manual the RNA 5' end was labelled by combing 1 µl universal reaction buffer nucleic acid (up to 0.6 nmols of 5' ends in ≤ 8 µl), 1 µl alkaline phosphatase and RNase free water to a final reaction volume to 10 µl then mix and incubated for 30 min at 37 °C. The reaction mixture was mixed with 2 µl universal reaction buffer, 1 µl ATPγS, 2 µl T4 polynucleotide kinase and RNase free water to a final reaction volume to 20 µl then mix and incubated for 30 min at 37 °C.

10 µl of fluorescence maleimide as a thiol-reactive label (Vector laboratories) was added, mixed and incubated for 2 hrs at room temperature, added 70 µl of RNase free water and 100 µl of buffered phenol (Fisher) then vortex briefly. The upper aqueous layer (phenol layer) was removed to a clean micro centrifuge tube, add 5 µl of precipitant, 270 µl of 95% ethanol and mix then centrifuged at 13,000 RPM for 30 min. The pellet was washed briefly with 70% ethanol and centrifuge at 13,000 RPM for 3 min, the pellet was dried then

resuspended into 50 µl of RNase free water and stored at - 80°C for further study.

2.31. Electro mobility shift assay for ribosome RNA interaction (EMSA).

The RNA-ribosome samples and control reactions were prepared separately according to Invitrogen manufacturer's protocol by combining of 2 µl of 5X binding buffer (Component E) (750 mM KCl, 0.5 mM dithiothreitol, 0.5 mM EDTA, 50 mM Tris, pH 7.4), 2 µl of RNase free water and X µl of ribosome and X µl of labelled RNA sampled or controlled labelled nucleic acid to the final volume of 10 µl. This was followed by incubation of both reactions for 30 min at 30 °C, then the reaction mixture was mixed with 2 µl of 6X EMSA gel-loading solution (Component D) (Invitrogen) and mixed gently. The RNA–ribosome complexes and control reaction were separated by electrophoresis using a nondenaturing polyacrylamide gel by loading a 15µl of EMSA samples and control to precast 6% Tris-glycine polyacrylamide gel (Invitrogen) with the use of mini gel tank (Life Technologies) and runs at 120 V for 1hr into 1X Tris-glycine running buffer (Invitrogen). The gel was visualized by using gel imaging for fluorescence and chemiluminescence (G: BOX Chemi XX6).

2.32. SYPRO Ruby EMSA of gel staining for ribosome RNA interaction.

The two methods of Sypro Ruby EMSA of gel staining for ribosome RNA interaction were used in which the unlabelled RNA gel was placed into a clean square plastic staining container with lid, a sufficient amount of Sypro Ruby stain

with TCA (Invitrogen) was added to cover the gel; about 50 mL for a 6 cm 9 cm 0.75 mm gel, and the gel incubated with continuous gentle agitation on an orbital shaker at 50 rpm for ~3 hours in the dark. The gel was washed two times into 150 mL of dH_2O for ~10 seconds for a total of 20 seconds then destained with 10% methanol, 7% acetic acid for 60 minutes. The gel was washed two times with 150 mL of dH_2O for ~10 seconds for a total of 20 seconds and visualized by using gel imaging for fluorescence and chemiluminescence (G: BOX Chemi XX6).

Chapter 3 Bioinformatics, predication, cloning and sequencing of 5'- UTR, 3'-UTR and Frameshift of HCoV- HKU1.

3.1 Introduction.

Human coronaviruses, like all other RNA viruses, have specific RNA structures that carry out important roles at both the transcription and translation stages of the virus replication cycle and may influence pathogenicity. Many previous research studies have identified 3 such structures as the 5' untranslated region (5'-UTR), the frameshift (FS) and the 3' untranslated region (3'-UTR) (Araujo *et al.*, 2012, Sawicki *et al.*, 2007). These structures have particular motifs that act at distinct transcription and translation steps during virus replication cycle. More generally, RNA secondary structures have important functions for RNA-RNA interaction, the binding of viral and cellular proteins to the RNA during RNA replication and translation (Brian and Baric, 2005, Liu *et al.*, 2009, Yang and Leibowitz, 2015). For example, the 5'-UTR of mouse hepatitis virus (MHV) includes a higher-order structural sequence that functions as a cis-active element that is vital for viral genome transcription and replication (Liu *et al.*, 2007).

Mouse hepatitis virus (MHV), bovine coronavirus (BCoV) and HCoV-HKU1 have a pseudoknot structure in the 3' untranslated region 63-115 nucleotides downstream of the N gene (nucleotide position 29708-29760) that is essential for viral replication (Woo *et al.*, 2005b). The 3' stem-loop II-like motif in the 3'-UTR of astrovirus, coronavirus and equine rhinovirus genomes has an important role in viral replication in which it might bind to the host's eukaryotic translation initiation factor 1A (eIF-1A) to take-over the host translational machinery for use by the virus, or to bind other translational regulation proteins with similar folds for similar purposes.

The presence of this conserved motif in three different viral families has been suggested to be the result of at least two separate recombination events and they have been important for the design of potential anti-viral therapeutic agents (Hor *et al.*, 2013). The structural elements of pseudoknots in particular are usually used as a helper in order to stabilize a complex 3D structure to enable it to perform an active role as a vital element in the regulation of several biological processes such as binding of ribosomal proteins to RNA (Giedroc *et al.*, 2000), initiation of internal ribosome entry translation (Wang *et al.*, 1995) and controlling a translational frame shift during protein synthesis (Green *et al.*, 2008).

The pseudoknot found at the junction of the 1AB ORF has the latter function, a frameshift which regulates the relative translated levels of polyproteins 1A and 1AB. Frameshifting does not occur on every RNA, most translation therefore stops at, or just after, the frameshift site. However, the occasional frameshift fuses the 3' downstream ORF with the 5' upstream component leading to a defined molar ratio between the upstream and downstream produced proteins, as required during virus replication. It has been found that lowering the frameshifting efficiency results in reduction of virus infectivity for HIV-1 and Murine Maloney Leukaemia Virus (Biswas *et al.*, 2004, Dulude *et al.*, 2006) showing that the evolved ratio of upstream and downstream sequence translation is critical for viral infectivity (Shehu-Xhilaga *et al.*, 2002, Dos Ramos *et al.*, 2004). Similar mechanisms are used in many viruses, notably other retroviruses, to effect the same differential regulation of protein expression.

Coronaviruses have a high level of mutation, which may be beneficial for them to escape from host defences and to rapidly adapt to new ecological niches. Coronavirus mutations are created by inaccurate copying by the virus encoded polymerase with more extensive changes also caused by template switching during replication, leading to a recombinant genome. The polymerase itself lacks

a proofreading 3' to 5' exonuclease domain (Steinhauer *et al.*, 1992) which results in lack of repair of any mismatched bases and, given the extreme size of the genome (*circa*. 30kb), all progeny genomes will contain many mutations.

Template switching happens when one cell is infected by different coronavirus strains and the polymerase starts to replicate one genome but then jumps to another to generate a new recombinant. Despite this extensive variation, the secondary structure elements have been found to be highly conserved, reflecting their importance for virus survival. The essential nature of these signals leads to studies of their action in the virus life cycle; what structures form, what virus and host factors bind to them, how is their formation and function controlled? These studies can identify the shared functions of secondary structure elements, in turn suggesting their candidacy as targets for intervention. For human coronavirus HKU1 there is limited knowledge of the RNA secondary structures present and of their roles in transcription and translation as the majority of studies to date have used mouse hepatitis virus, bovine coronavirus or SARS coronavirus. To counter this deficiency of data, this chapter focusses on the use of HKU1 sequence as the source for studying the structures noted above. The study design will apply the most relevant available techniques starting with a bioinformatics approach using several software packages such as Mfold, Jalview and SnapeGene to identify, align and compare these three untranslated regions of HCOV- HKU1. Predictions are made of how the sequences are most probably structured so that the modelled structures can then be probed by mutagenesis. Once defined, each region is cloned into a plasmid, the vector pTriEx 1.1, and further modified with additional sequences upstream and downstream of the cloned signal in order to provide a reporter plasmid to measure the activity of the fragment bearing the secondary structure.

3.2. Results.

3.2.1. Sequence analysis.

Human coronavirus HKU1, a member of the human coronaviruses, was the subject for this study. The National Centre for Biotechnology Information web site (NCBI) was used to identify and download 30 complete genome sequences of CoV-HKU1 that includes the 5'-UTR, 3'-UTR and the ORF1AB junction including the frameshift (FS). Multiple sequence alignment program Clustal Omega was used in order to identify sequence identity and to eliminate sequences with less than 100% coverage. Subsequently the unique sequences were analysed and viewed using Jalview software.

3.2.1.1. Sequence analysis of CoV-HKU1 5'-UTR.

To examine the features in the 5' UTR of coronaviruses, the first 200 bases of the downloaded sequences were aligned and compared. It was found that four strains had a unique sequence for the 5'-UTR of HCoV-HKU1, these strains are N3, N15, N16 and Ref **Figure 19**. Some variation was evident in these aligned sequences and the base identity and locations of the nucleotide changes among the isolates concerned are summarized in **Table 13**. In the modelled secondary structure of the 5'-UTR (see **Fig. 22** later), the base changes from C to T, T to C or sequences in which the first six bases (GAGTTT) are missing are likely to have an effect on the stability of a folded RNA in the living cell. This is due to the formation of weaker base pairs between the T and G or C and A as it does not follow the canonical rules of Watson-Crick for pairing of the four nucleotides, that is [(A, U), (U, A), (G, C), (C, G), (G, U), (U, G)] (Dirks *et al.*, 2004).

As a result of these changes RNA molecules may be less stable inside living cells and in order to gain more stability a distant part of the single-stranded

RNA may fold back on itself to form a new RNA secondary structure [stem loop of RNA and RNA with pseudoknots], increasing overall structural stability and minimising free energy. In other words, other base pairing interactions may occur within a single molecule or pair of interacting molecules to compensate for mutations picked up by the natural genetic drift in the genome of the virus. When these possibilities are taken together, different kinds of loops can be shown to occur from the possible pairings between nucleotides and these possible structures can be assigned into two large groups termed stem-loops and pseudoknots. The term loop may include hairpin loop, interior loop, multi loop, stack loop, bulge loop and external loop (Borkar and Mahajan, 2014).

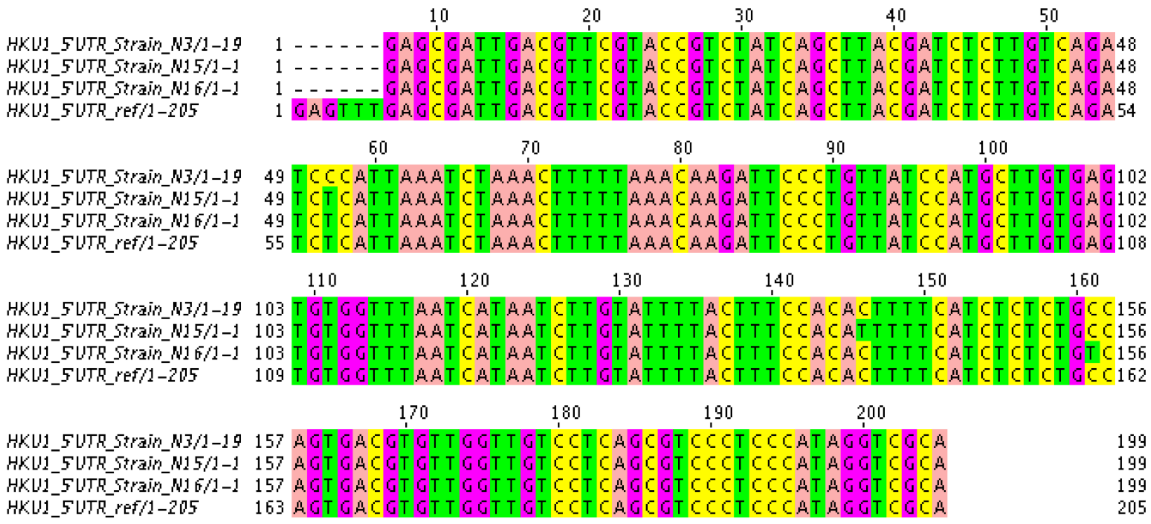

Figure 19. Alignment of the unique CoV-HKU1 5'-UTR sequences. Although largely similar several nucleotide changes, highlighted in different colors, lead to unique sequence identities, (1-205) nucleotide refer to the location of the nucleotide in the complete genomic RNA.

Table 13. Base changes in the 5'-UTR in HCoV-HKU1 strains.

Strain	Base change	Location
N3	Very close to strain Ref. but differed by missing the first six bases (GAGTTT) and a T to C change	1-6 and 57
N 15	Very close to strain Ref. but differed by missing the first six bases (GAGTTT) and a C to T change	1-6 and 146
N 16	Very close to strain Ref. but differed by missing the first six bases (GAGTTT) and a C to T change	1-6 and 161

3.2.1.2. Sequence analysis of HCoV-HKU1 3'-UTR.

To compare the 3'UTR regions, a similar alignment study was done and it was found that of all the genomes analysed nine isolates have a unique sequence for the 3'-UTR of CoV-HKU1, these strains are Ref, N9, N14, N15, N16, N17, N19, N21 and N25 **Figure 20**. The variation evident in these aligned sequences and their nucleotide types and locations are summarized in **Table 14**. In the modelled secondary structure of the 3'-UTR (see **Fig. 23** later) the base changes from C to A, G to T, No-T, T to A, T to C, and A to G are likely to have an effect on RNA stability in to the living cell again due to the formation of weaker base pairs between T-G, C-A and A-A, which do not follow the rules of Watson-Crick for pairing up of the four nucleotides.

Figure 20. Alignment of the unique CoV-HKU1 3'-UTR sequences. Mismatched bases are shown in highlighted colors amid a column of the same color, (1) nucleotide refers to (29646) location of the nucleotide in the complete genomic RNA whereas (281) refers to (29926) location of the nucleotide in the complete genomic RNA.

Table 14. Base changes identified in the 3'-UTR of CoV-HKU1 strains.

Strain	Base change	Location
Ref	T to A, A to G T to C	89 149 151
N 9	C to A G to T No-T T to A C to T G to A C to T A to G T to C G to T A to G	22 48 80 89 91 112 138 139 151 213 219
N 14	C to A G to T No-T T to A C to T G to A C to T A to G T to C G to T A to G A to G	22 48 80 89 91 112 138 139 151 213 21 224
N 15	Very close to N 17 but differed by having T and A	80 and 235
N 16	Very close to N 15 but differed by having C	235
N 17	Very close to N 15 but differed by No- T and C	80 and 235
N 19	T to A C to T G to A C to T A to G T to C G to T A to G A to G	89 91 112 138 139 151 213 21 224
N 21	Very close to N 15 but differed by T to C and A to C	7 and 235
N 25	Very close to N 15 but differed by T to G	46

3.2.1.3. Sequence analysis of CoV-HKU1 1ABFs.

To compare frameshift sequences (FS) at the A/B junction region of the 1AB open reading frame, this region was examined for all available sequences. Six strains were found to have a unique sequence for the 1ABFS of CoV-HKU1, these strains are 4, 5, N15, N16, N25 and Ref **Figure 21**. The variation among the isolates is summarized as changes of nucleotide type and locations in **Table 15**. In the modelled secondary structure of the 1ABFS (see **Fig. 24** later) these base changes from C to T, T to C, G to A and A to G which will again have a plausible effect on RNA stability in the living cell as, again, they do not follow the rules of Watson-Crick base pairing.

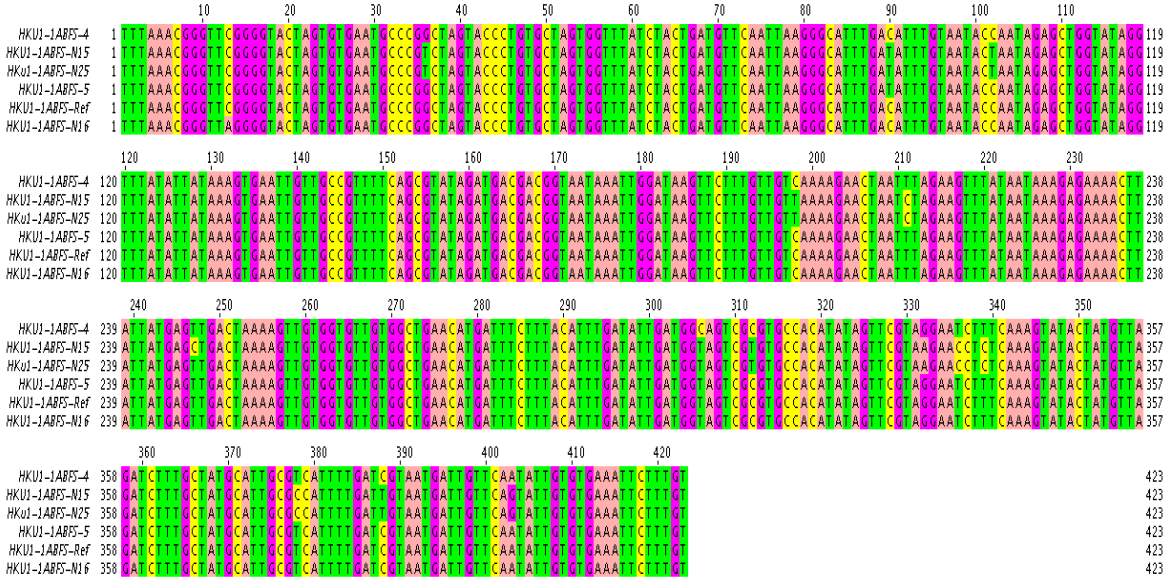

Figure 21. Alignment of the unique CoV-HKU1 1ABFS sequences. Mismatched bases are shown in highlighted colours amid a column of the same colour in the complete genomic RNA sequence.

Table 15. Base changes of the CoV-HKU1-1ABFS strains.

Strain	Base change	Location
4	Very close to Ref. but differed by changing of T to C	306
5	C to T	312
	G to A	332
	T to C	336
	T to C	339
	T to C	378
	C to T	388
	A to G	403
N 15	C to T	90
	C to T	103
	C to T	198
	T to C	211
	T to C	147
	T to C	312
	G to A	332
	T to C	336
	T to C	339
	T to C	378
	C to T	388
	A to G	403
N 16	Very close to Ref but differed by changing C to A	13
N 25	C to T	90
	C to T	102
	C to T	198
	T to C	211
	C to T	312
	G to A	332
	T to C	339
	T to C	378
	C to T	388
	A to G	403
Ref.	Very close to 4 but differed by having T	306

To summarize, alignments of the 5'- and 3'-UTR regions and the 1AB frameshift regions of unique HKU1 coronavirus genomic sequences available in the current databases identified a high degree of similarity but minor variation among them. These changes clearly affect predicted base pairing in stems and loops in the secondary structures models predicted to be formed from these sequences but the biological effect of this variation, if any, is not known.

3.2.2. Predication of secondary structure and TRS in the HCoV-HKU1 5'-UTR sequence.

In order to better understand the RNA secondary structure in the 5'-untranslated region of HCoV-HKU1, a secondary structure prediction for all available HCoV-HKU1 sequences was performed using the default parameters of the program RNA Mfold with 37°C as the folding temperature (http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form). The program predicts optimal and suboptimal secondary structures for an RNA molecule using the most recent energy minimization method of Zuker (Zuker, 2003). In this study, the first predicted folded form was used as the most likely structure as it represents the most optimal secondary structure. It was found that all sequences, including those above that varied slightly in the sequence of the 5'-UTR, were predicted to fold into structures that included three conserved stem loops, which are here numbered as SL-1, SL-3 and SL-5 **Figure 22**. Interestingly, the transcriptional regulatory sequence (TRS) used by the polymerase for discontinuous transcription of the coronavirus genome (UCUAAAC) was found to be located at the apex of SL-2 at a position from 65 to 71 in all available CoV-HKU1 sequences except for the reference sequence (NC_006577.2) where the entire SL-2 was predicted to be single-stranded.

There is some variation in the prediction of Sl-4 and the internal loop in SL-5 in the reference sequence (Ref.) (NC_006577.2) depending on a change from C to U at positions 146 and 161 that means that C-G, U-A or G will make a weaker base pair, compared with C-G. As a result, there is a change in the base pairing and the structure of stem loop. There is a wobble pair at the start or the end of loops which supports the possible presence of the two variants, which could reflect slightly different roles at different stages of the virus replication cycle.

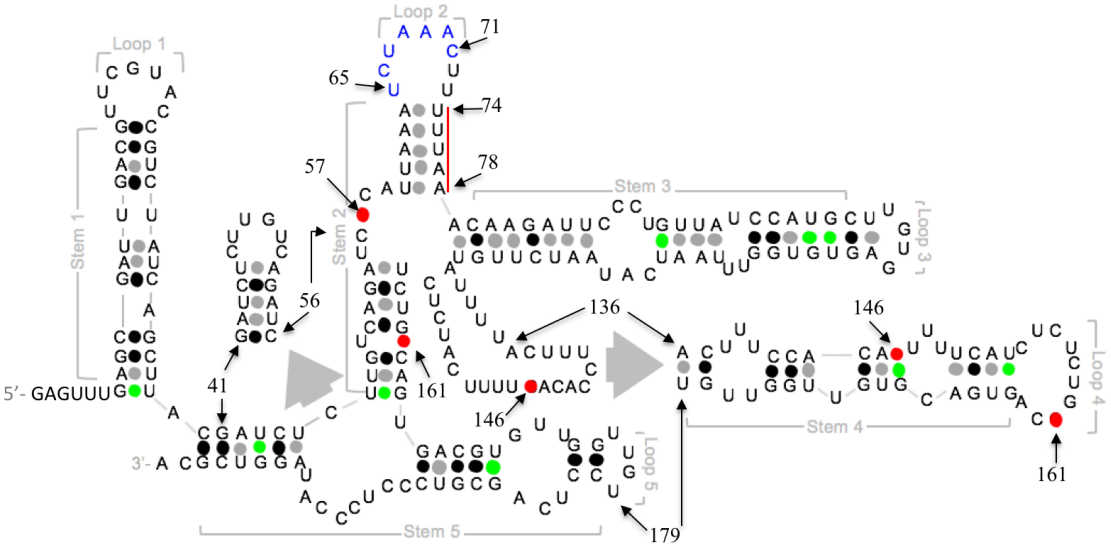

Figure 22. Prediction of secondary structure in all available HCoV-HKU1 5'-UTR sequences. The overall structure consists of SL-1 to SL-5 with each SL consisting of a stem and loop (numbered as shown). The transcriptional regulating sequence TRS (UCUAAAC) is found at the apex of loop-2 at positions 62 to 68. When the sequences used are aligned conserved or variant nucleotides are identified (see Figure 18) and are marked in the structure as (•) strong pair, (•) wobble pair, (•) weak pair, (•) change of C to U, (—) not consistently predicted at positions 74 to 78, arrows indicate the base change and start and end location of variable SL structure.

3.2.3. Predication of secondary structure of 1ABFS and slippery sequence in CoV-HKU1.

In order to better understand the RNA secondary structure of the frameshift element in all available HCoV-HKU1 isolates, a secondary structure prediction for all available HCoV-HKU1 sequences was performed manually according to (Woo *et al.*, 2005b). It was found that all these sequences have the same slippery sequence (UUUAAAC) that includes the first 7 nt (the blue coloured underlined sequence in Fig 23) followed by a pseudoknot structure made up by two stems, two loops, nt variability from G to U at positions 36 and a signature 32 nucleotide stretch in loop 2 **Figure 23**. There are wobble pairs at

the start or the end of loops suggesting the stems could vary in length depending on the cellular environment. The conservation of structure is consistent with the conclusion that the structure shown in all available HCoV-HKU1 has an essential function necessary for virus replication, although this has not yet been formally measured.

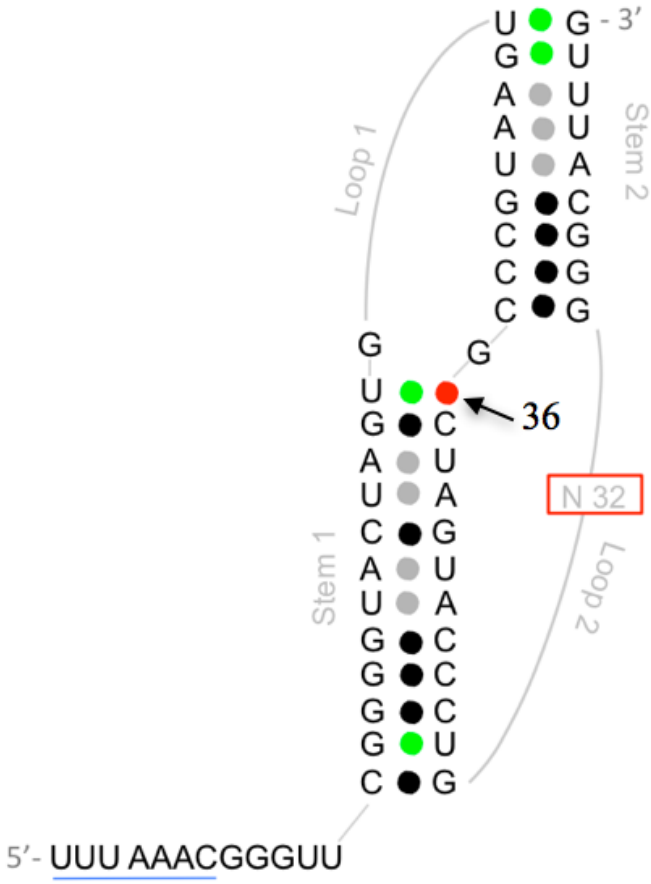

Figure 23. Model of ribosomal frameshift element in CoV-HKU1. The overall structure consists of a slippery sequence followed by two stems and two loops. Highly conserved or variant nucleotides are identified (see Figure 20) and are marked in the structure as (•) strong pair, (•) wobble pair, (•) weak pair and (•) change of G to U at position 36. The slippery sequence is underlined (—) and the constant length of loop 2 indicated (N_{32}).

3.2.4. Presence of secondary structure in the HCoV-HKU1 3'-UTR sequences.

In order to better understand the RNA secondary structure in the 3'-untranslated region of HCoV-HKU1, secondary structure predictions for all available HCoV-HKU1 sequences were performed as before using the default parameters of RNA Mfold with 37°C folding temperature (http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form). As before, the first predicted folded form was used as a base as it represents the most optimal secondary structure. It was found that all sequences including those that varied in the 3'-UTR were predicted to fold into two conserved stem loops, which are here numbered as SL-1 and SL-2 **Figure 24**. However, there is variability in the presence or not of predicted stem loops SL-3 at the positions started from nt 133 to nt 172 and SL-4 at the positions starting from nt 215 to nt 250, the latter of which occurs from the middle of SL-3 in genotypes B and C. Their occurrence depends on the presence of the weak pair U-A or G, which will make a weaker base pair, compared with C-G.

Additionally, there is a change from A to C or G or U at positions 20, 46, 84, 86, 107, 208 and 214 resulting in a change in the structure of stem loops. As for the other folded sequences, there is a significant increase in the number of wobble pairs at the start or the end of predicted stems and loops suggesting a dynamic structure. Nevertheless the conservation of a core folded structure is likely to indicate that it has a fundamental role in virus replication.

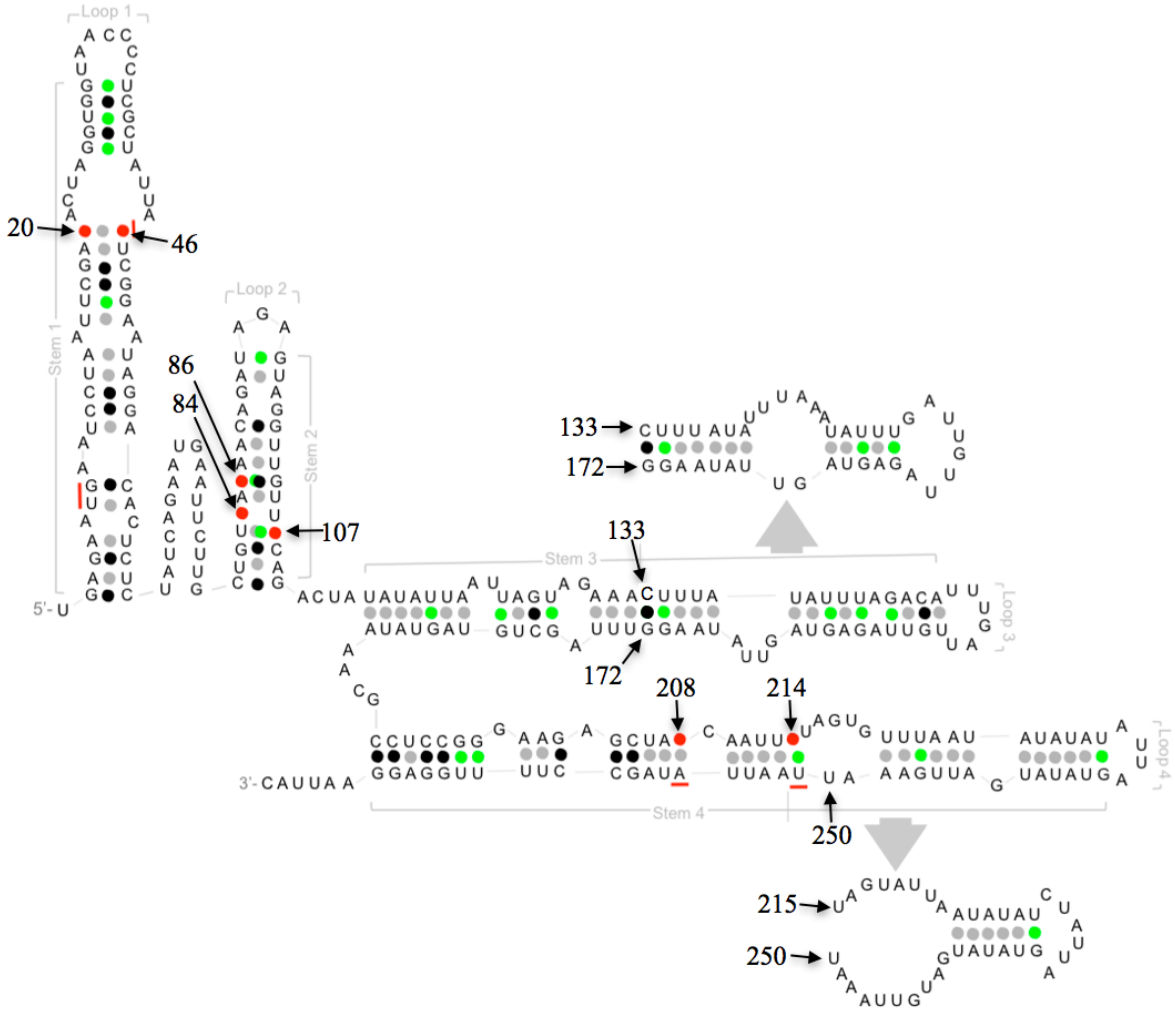

Figure 24. Prediction of secondary structure in all available CoV-HKU1 3'-UTR sequences. The overall structure consists of SL-1 to SL-4 with each one consists of the stem and the loop. However only SL1 and SL2 are consistently predicated, the structures SL3 and SL4 being dependent on the actual sequence used. Variability among the target sequences is shown as: (●) strong pair, (●) wobble pair, (●) weak pair, (●) change of A to C or G or U. Several parts of the fold labelled (—) are not consistently predicted, arrows indicate the base change and start and end location of variable SL structure.

To summarize, MFold predictions of the 5'-UTR, 3'-UTR and 1AB frameshift regions of unique strains of HKU1 coronavirus genomic sequences (N3, N15, N16 and Ref), (Ref, N9, N14, N15, N16, N17, N19, N21 and N25) and (4, 5, N15, N16, N25 and Ref) respectively taken from the entire coronavirus sequence database identified similar overall folds but with minor variation in the length or placement of the stems and loops. It was also noted that biologically significant sequences, such as the TRS, were prominent in the deduced structures. Minor variation in the stems and loops could result in modified RNA folds which alter the availability of such sequences to any *trans*- acting factors. The biological consequence of such changes however has not been reported to date, making their study in an accessible experimental system timely and worthwhile.

3.2.5. Study the RNA secondary structure.

3.2.5.1. Construction of reporter plasmids for the 5'-UTR, 3'-UTR and Frameshift fragments of HCoV-HKU1.

Bioinformatics alignments and secondary structure predication of the selected HKU1 CoV sequences revealed minor variation on a conserved background that folded into structures not dissimilar to those published for other coronaviruses (Plant *et al.*, 2005, Raman and Brian, 2005). However, as the role of this variation might be important and as there is limited knowledge about the RNA secondary structure of HCoV-HKU1 specifically, genetic constructs were designed in order to get a better idea of how these 5'-UTR, 3'-UTR and frameshift RNA fragments function. These constructs were assembled *in vitro* by the process of overlapping PCR as discussed in chapter 2 section (2.12) and cloning

maps for the designed fragments in the expression vector pTriEx1.1 were derived by using SnapeGene software as shown in **Figures 25, 26** and **27**.

The assembled PCR fragments were ultimately flanked by sequences homologous to the vector and final assembly was achieved through use of the Infusion cloning reaction according to the manufacturer's instruction protocol (Clontech) as discussed in chapter 2 sections (2.12-16). The linearization of the pTriEx1.1 vector, necessary to reduce the background of transformants that would otherwise swamp any Infusion recombinants, was achieved by the use of *NcoI* and *XhoI* as fast-digest restriction enzymes (Thermo Scientific) **Figure 28** following which an infusion reaction was completed followed by transformation into Stellar TM Competent Cells (*E. coli* HTS08 strain). Ampicillin resistant colonies were grown for plasmid extraction and the DNA present analysed by double digestion in order to confirm the presence of the desired DNA fragments in the pTriEx1.1 vector **Figure 29**. Final examples of the positive clones for use in further experiment are shown in **Figures 30**.

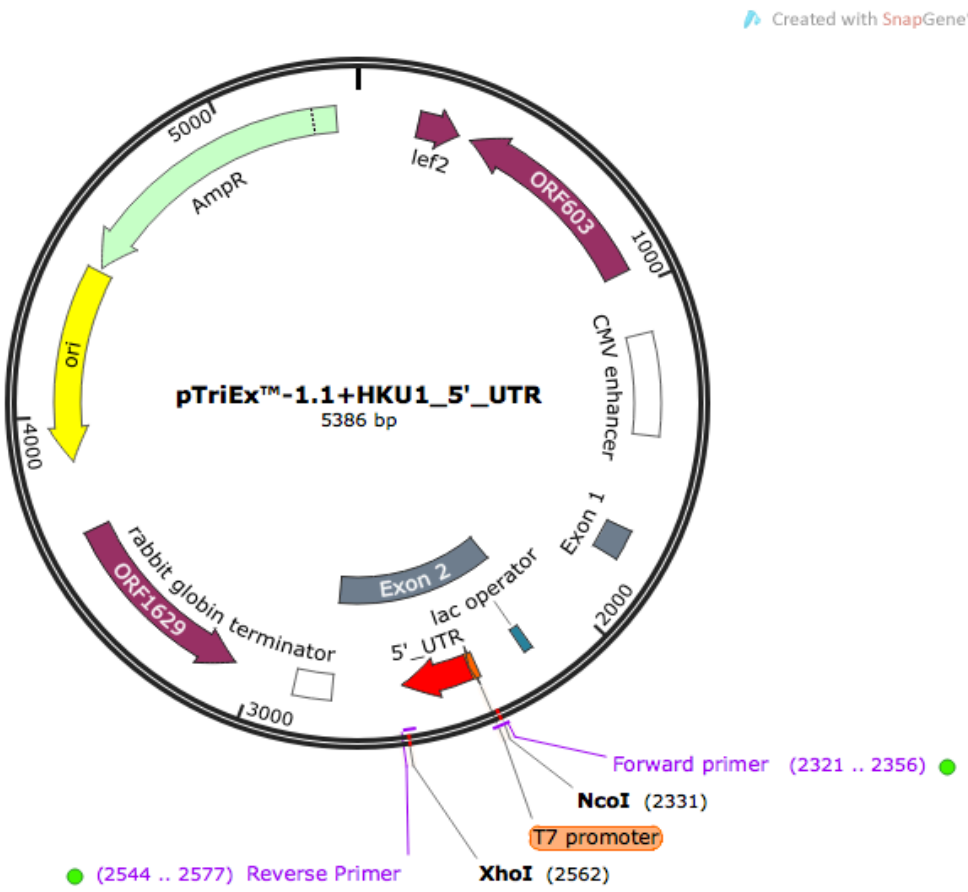

Figure 25. The cloning map for pTriEx1.1 with HKU1-5'-UTR. Key features of the plasmid are marked.

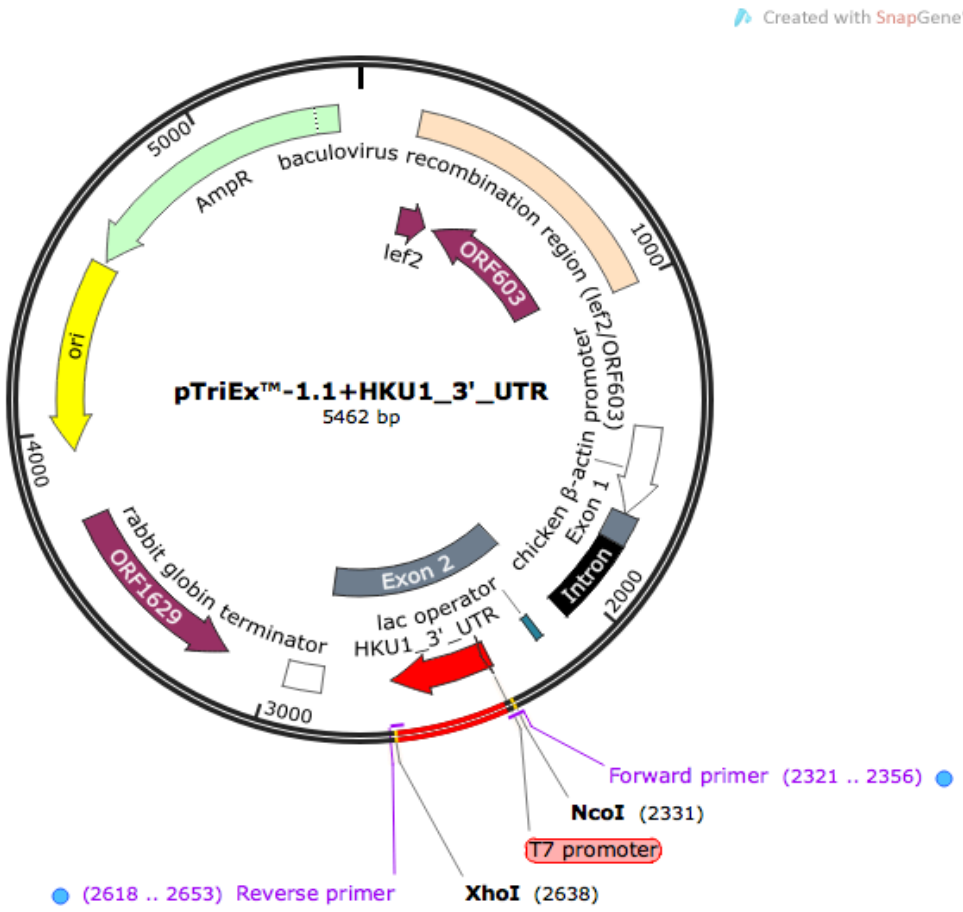

Figure 26. The cloning map for pTriEx1.1 with HKU1-3'-UTR. Key features of the plasmid are marked.

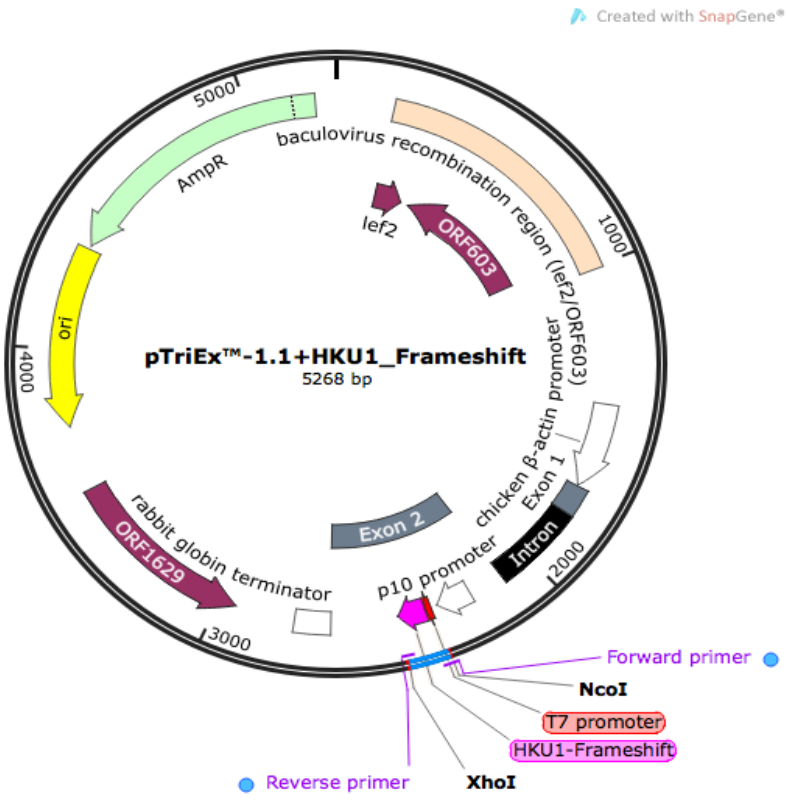

Figure 27. The cloning map for pTriEx1.1 with HKU1-Frameshift. Key features of the plasmid are marked.

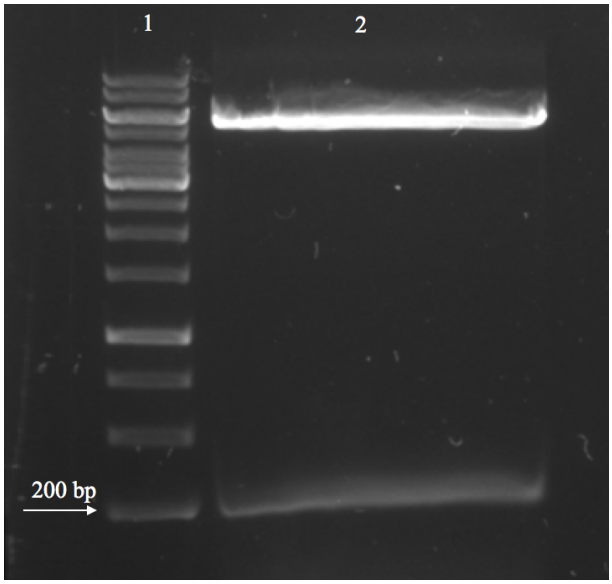

Figure 28. Gel electrophoresis of the double digest of pTriEx1.1. Lane 1: Hyperladder 1kb, lane 2: pTriEx 1.1 vector digested with *NcoI and XhoI* restriction enzymes. The size of the excised *NcoI-XhoI* band representing multicloning sites that are not used is shown. The residual vector lacking this fragment was used for the infusion reactions.

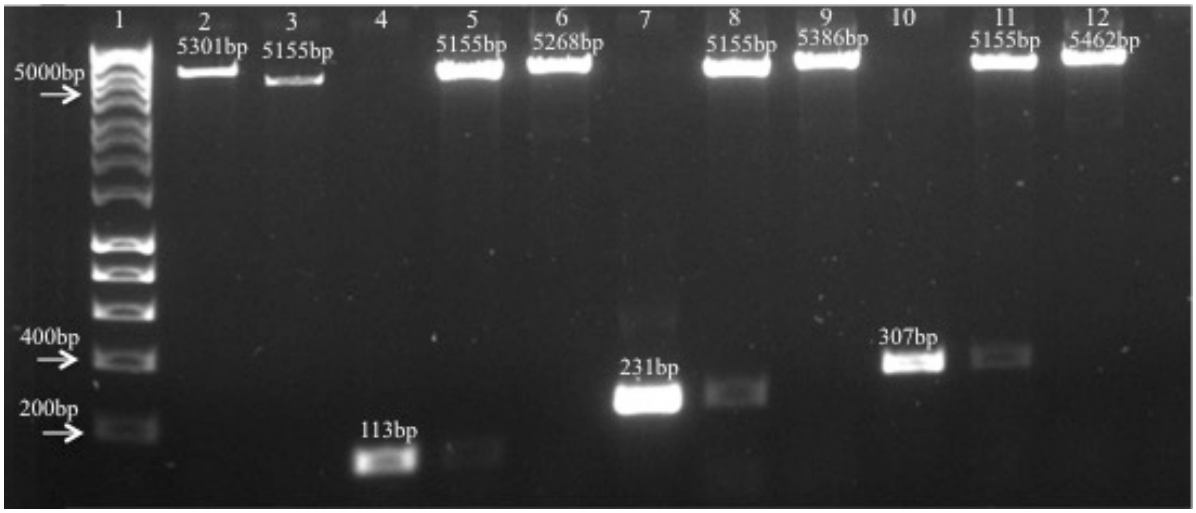

Figure 29. Gel electrophoresis of the stages of the infusion cloning for frameshift, 5-'UTR and 3'-UTR using pTriEx1.1 vector. Lane1: hyperladder 1kb, lane 2: pTriEx 1.1 vector linearized with *NcoI* restriction enzyme, lane 3: pTriEx 1.1 double digested with *NcoI* and *XhoI*, lane 4: Frameshift (Insert), lane 5: double digestion for the plasmid DNA in order to confirm existing of the Frameshift, lane 6: Frameshift infused plasmid DNA linearized with *NcoI* restriction enzyme, lane 7: 5'-UTR (Insert), lane 8: double digestion for the plasmid DNA in order to confirm existing of the 5'-UTR, lane 9: 5'-UTR infused plasmid DNA linearized with *NcoI* restriction enzyme, lane 10: 3'-UTR (Insert), lane 11: double digestion for the plasmid DNA in order to confirm existing of the 3'-UTR, lane 12: 3'-UTR infused plasmid DNA linearized with *NcoI* restriction enzyme.

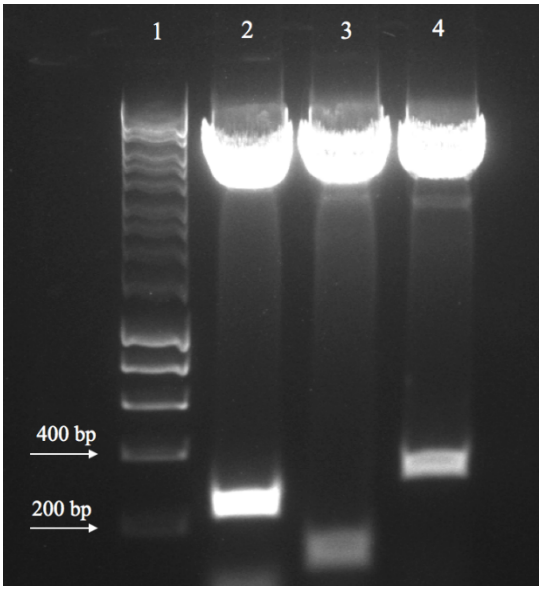

Figure 30. Gel electrophoresis of double digestion products showing the desired DNA fragments. Lane1: hyperladder 1kb, lane 2: pTriEx 1.1 + 5'-UTR digested with *NcoI* and *XhoI*, lane 3: pTriEx 1.1 + Frameshift digested with *NcoI* and *XhoI*, lane 4: pTriEx 1.1 + 3'-UTR digested with *NcoI* and *XhoI*. All the released band sizes agree with the fragments as designed.

3.2.6. Sequencing alignment of 5'-UTR, 3'-UTR and Frameshift of HCoV-HKU1.

The final positive clones for the 5'- and 3'- UTR and the frameshift constructs were confirmed by sequencing analysis according to the Source BioScience website as described in chapter 2 section (2.18), to ensure that there was no mutation in the 5'-UTR, 3'-UTR and frameshift fragments generated by overlapping PCR or in the upstream T7 promoter sequence. The derived sequence (upper sequence and the trace file in figure) was compared to each of the designed sequences (lower sequence in figure). No untoward mutations were found in any of the assembled fragments or in the flanking vector sequence as far as was evident from these sequences **Figures 31**, **32** and **33**.

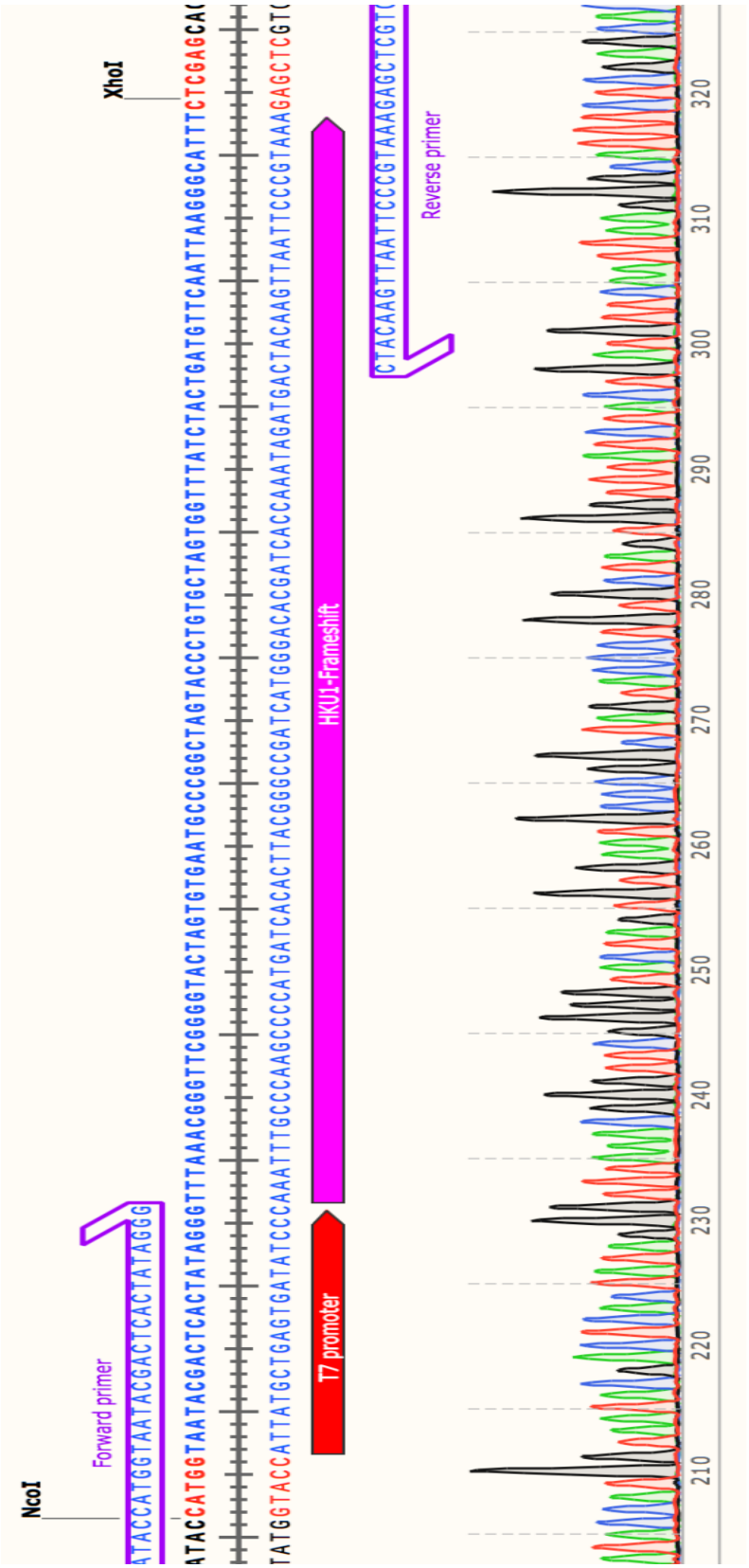

Figure 31. Sequence alignment of the pTriEx1.1+HKU1-Frameshift with the synthesized sequence.

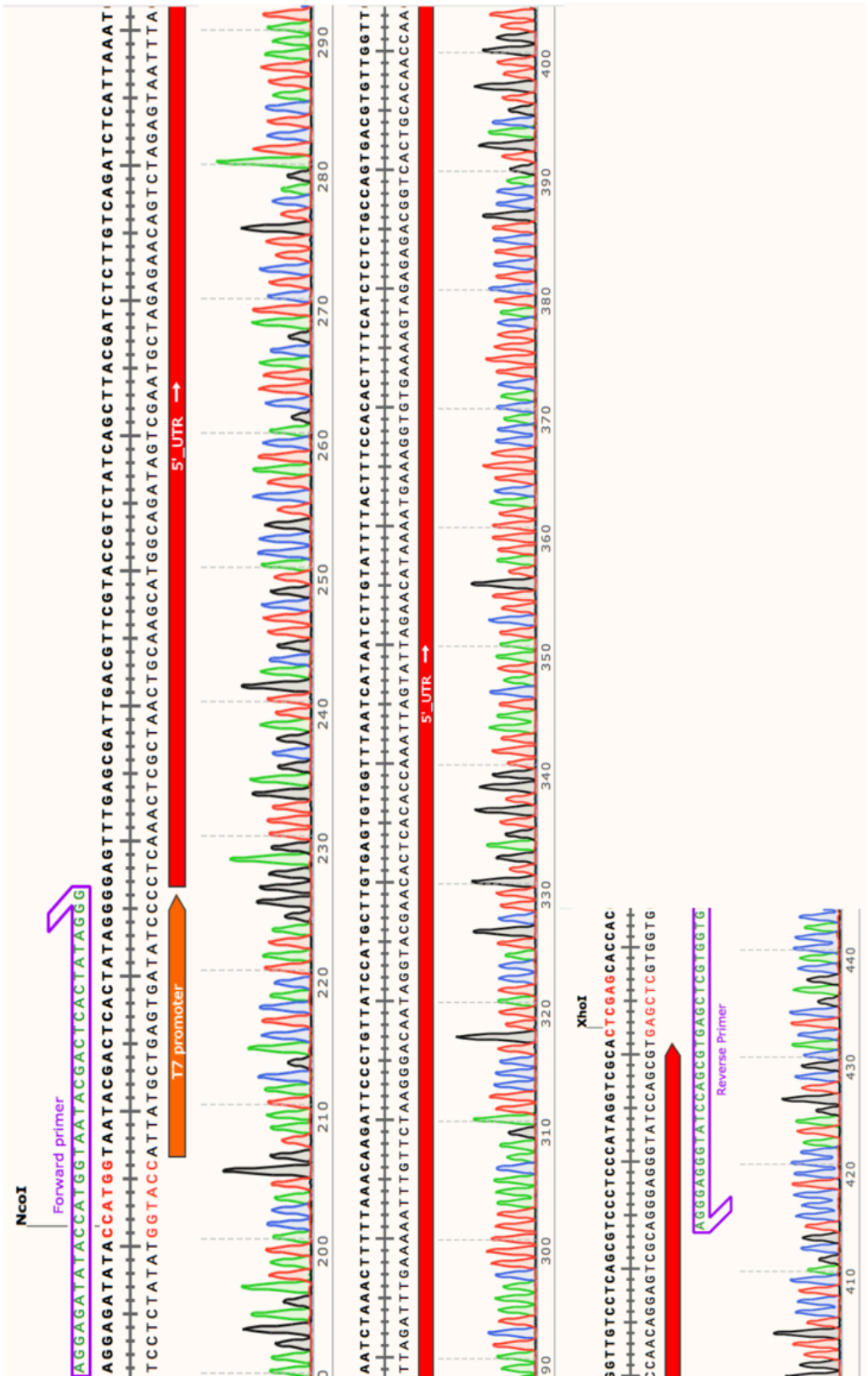

Figure 32. Sequence alignment of the pTriEx1.1+HKU1-5' -UTR with the synthesized sequence.

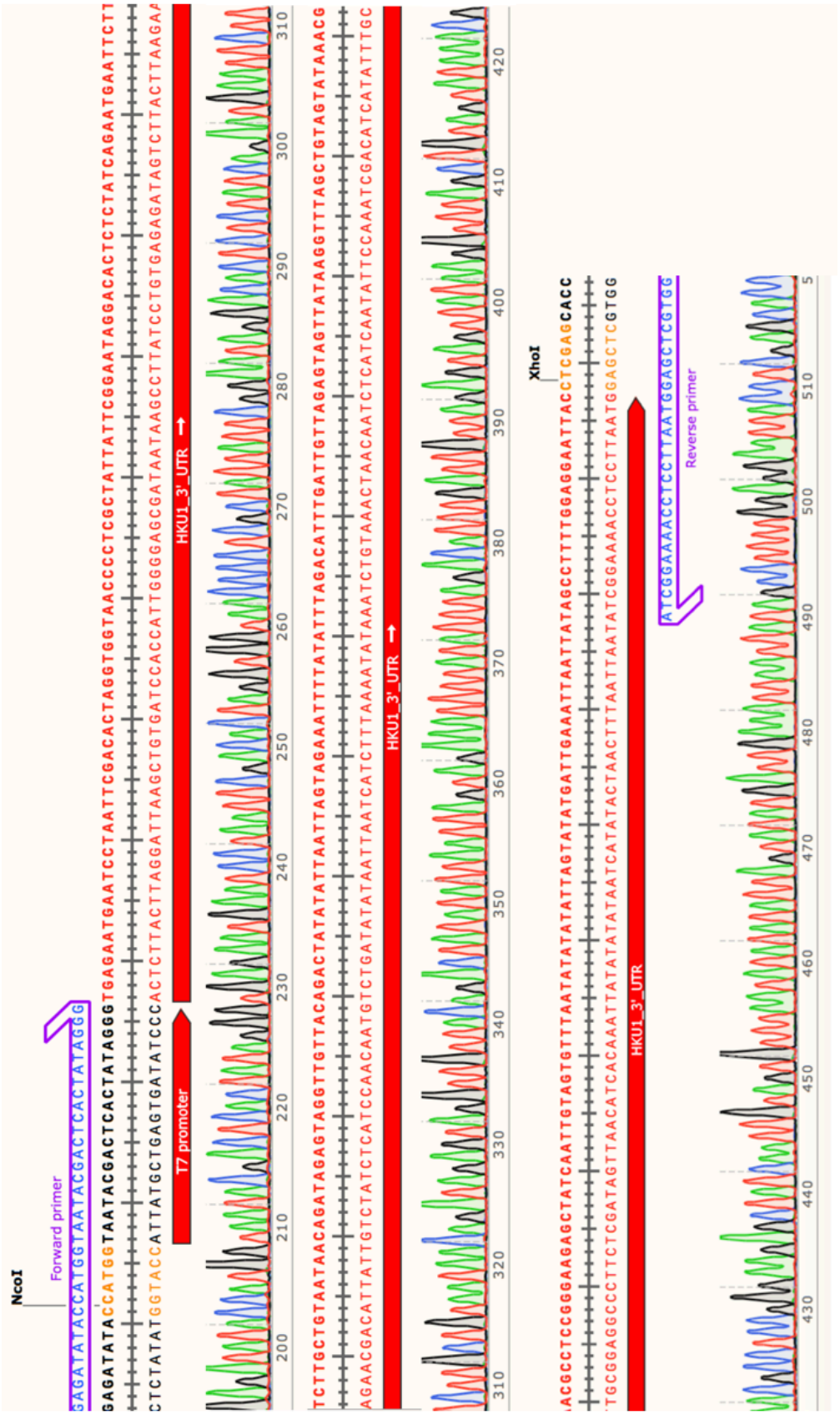

Figure 33. Sequence alignment of the pTriEx1.1+HKU1-3' -UTR with the synthesized sequence.

3.2.7. Functional analysis of RNA secondary structure.

3.2.7.1. Construction of vectors with reporter gene with 5'-UTR, 3'-UTR and Frameshift fragments of HCoV-HKU1.

The constructions of the key UTR RNA control elements described above do not allow a measurement of their function as no reporter gene was present on the vector used. Therefore, in a second set of cloning reactions, two reporter genes, the enhanced GFP sequence (eGFP) and the monomeric red fluorescent protein (mCherry), were added as markers to allow the functional detection of the 5'-UTR, 3'-UTR and frameshift RNA secondary structures derived from HCoV-HKU1. The cloning of these constructed fragments was as detailed in the previous section (3.2.5) and cloning maps for these fragments with the addition of these reporter genes were assembled using SnapeGene software and are presented in **Figures 34 A** and **B**, **35 A** and **B** and **36 A** and **B**.

The infusion cloning kit was used according to the manufacture's instruction protocol (Clontech) as mentioned in chapter 2 sections (2.12-16) with linearization of the pTriEx1.1 vector by the use of *SacII*, *BstAPI*, *PpuMI* or *SphI* as restriction enzymes (Thermo Scientific) depending on which enzymes were shown as non-cutters of the chosen inserts according to the SnapeGene maps **Figure 37**. High fidelity PCR was used to amplify the two reporter genes eGFP and mCherry to ensure no error was generated **Figure 38**. The infusion reactions were transformed into Stellar TM Competent Cells (*E. coli* HTS08 strain) and ampicillin resistant colonies picked for plasmid extraction. Extracted DNA was double digested in order to confirm the presence of the desired DNA fragments in the pTriEx1.1 vector plasmid DNA **Figure 39**.

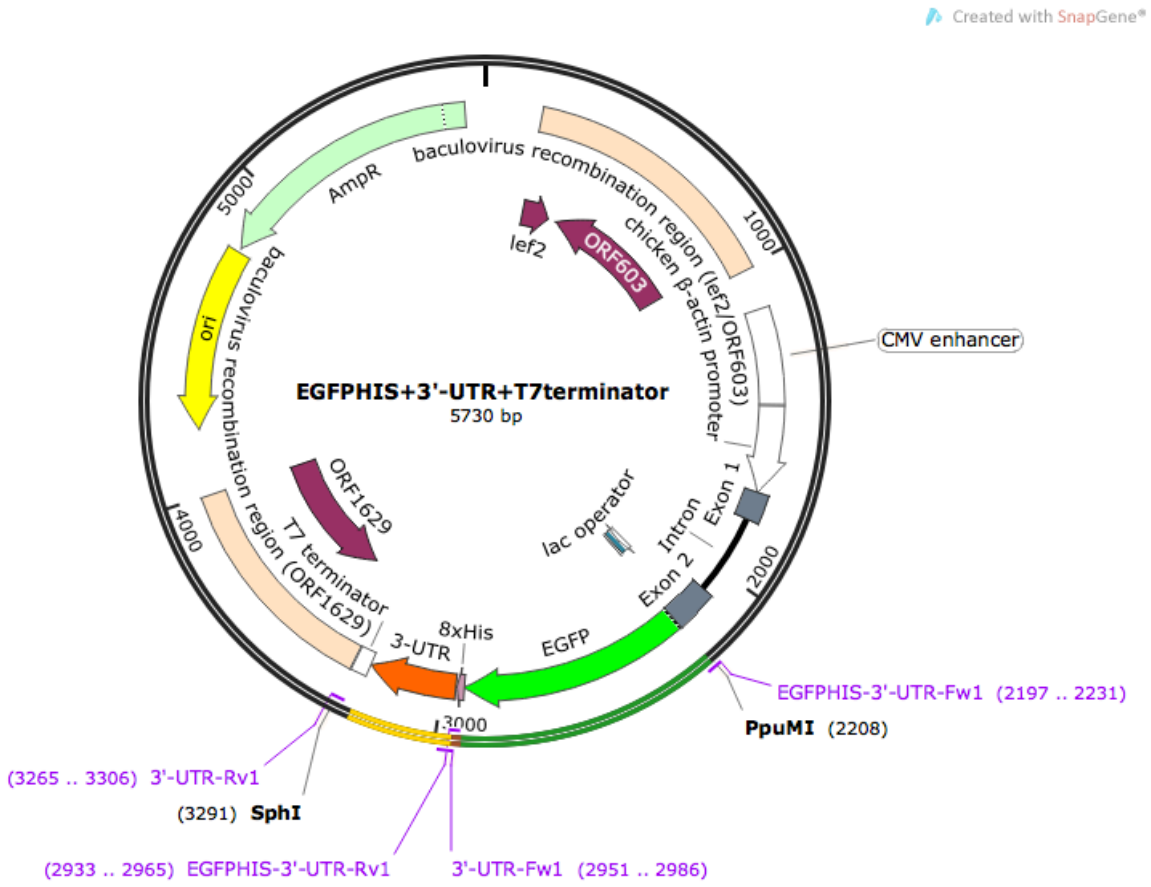

Figure 34 A. The cloning map for pTriEx1.1 with eGFP-His+3'-UTR+T7 terminator. Key features of the plasmid are marked.

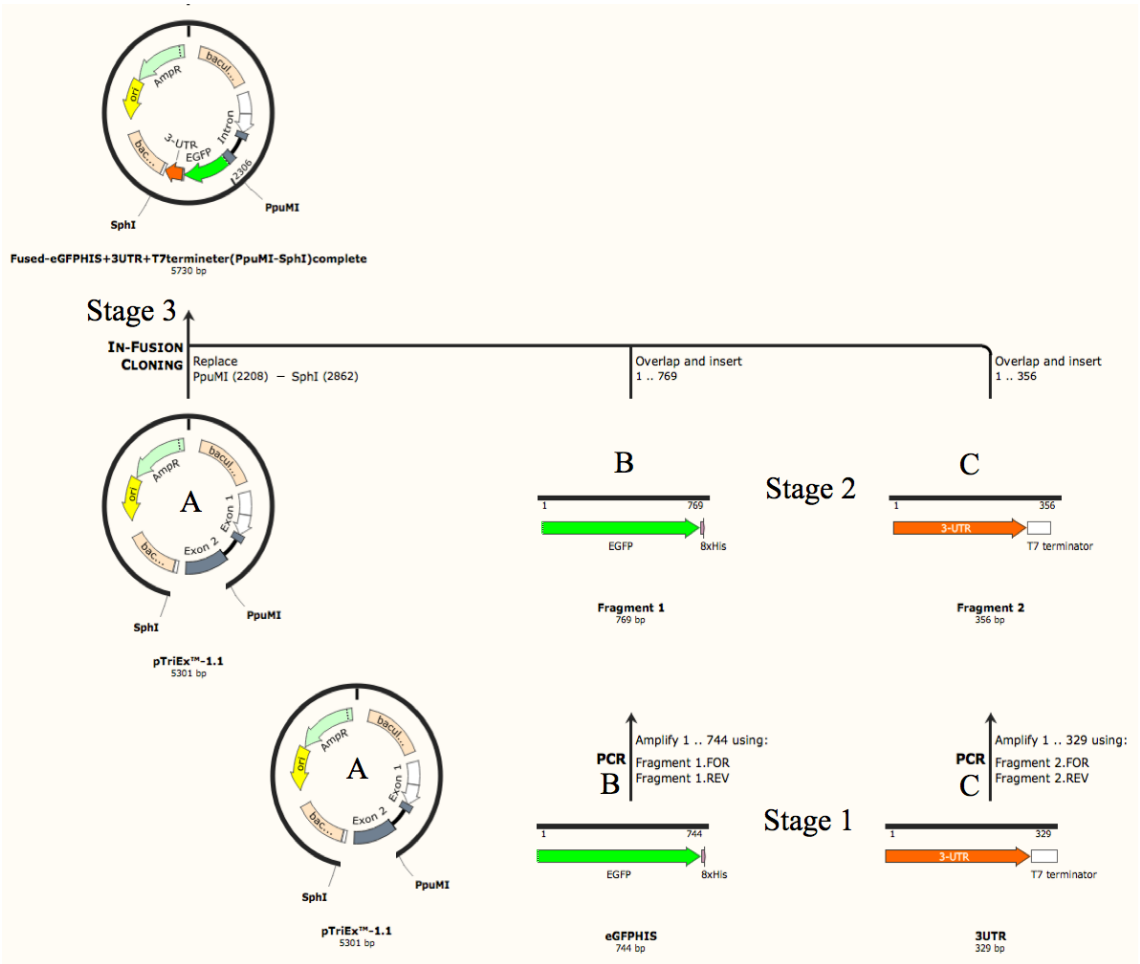

Figure 34 B. An overview of all stages used for construction of pTriEx1.1 with eGFP-His+3'-UTR+T7 terminator throughout the entire infusion method. Stage 1: A- pTriEx1.1 vector double digested with *PpuMI* and *SphI* restriction enzymes, B- amplification for eGFP-His sequence from plasmid DNA containing eGFP-His (a gift from B. Abdulsattar), C- amplification for the 3'-UTR-T7 terminator sequences from plasmid DNA containing 3'-UTR- T7 terminator (constructed in section 3.2.5), Stage 2: A- gel extraction and purification for pTriEx1.1 residual vector as used for the infusion reactions, B- addition of overlapping 15 bp to eGFP-His (insert), C- addition of overlapping 15 bp to 3'-UTR-T7 terminator (insert), Stage 3: incubation of the infusion mix in different conditions, transformation of Stellar ^{TM} Competent Cells (*E. coli* HTS08 strain) using LB/Ampicillin plate agar. Finally, the confirmed colony containing the desired plasmid DNA were selected and sub-cultured in LB/ Ampicillin agar plates at 37°C for 12-16 hrs, then stored in 50% glycerol at - 80°C for further study.

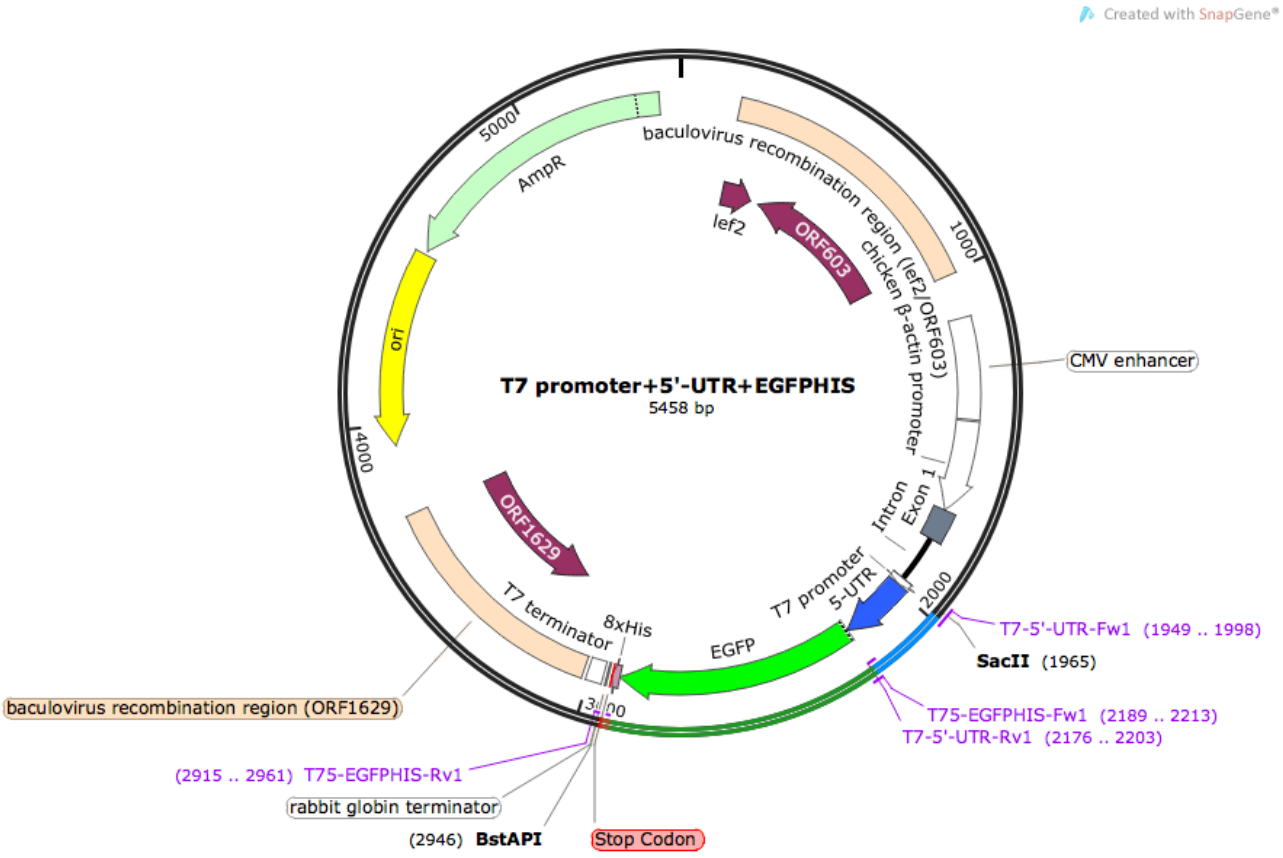

Figure 35 A. The cloning map for pTriEx1.1 with T7 promoter+5'-UTR+eGFPHis +T7 terminator. Key features of the plasmid are marked.

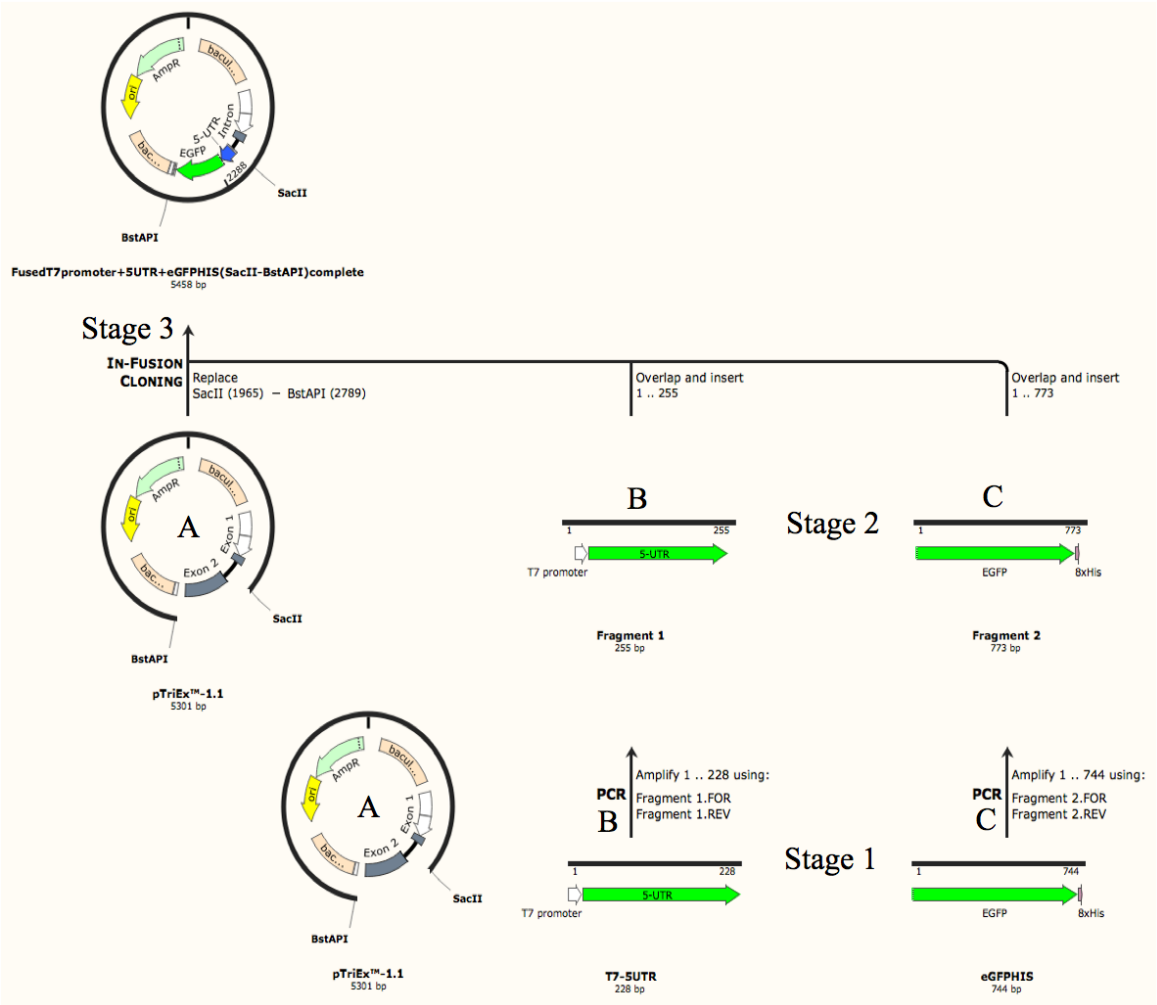

Figure 35 B. An overview of all stages used for construction of pTriEx1.1 with T7 promoter+5'-UTR+eGFP-His+T7 terminator throughout the entire infusion method. Stage 1: A- pTriEx1.1 vector double digested with *SacII* and *BstAPI* restriction enzymes, B- amplification for T7 promoter+5'-UTR sequence from plasmid DNA containing T7 promoter+5'-UTR (constructed in section 3.2.5), C- amplification for eGFP-His sequences from plasmid DNA containing eGFP-His (a gift from B. Abdulsattar), Stage 2: A- gel extraction and purification for the pTriEx1.1 residual vector used for the infusion reactions, B- addition of overlapping 15 bp to T7 promoter+5'-UTR (insert), C- addition of overlapping 15 bp to eGFP-His (insert), Stage 3: includes incubation of the infusion mix, transformation of Stellar ^{TM} Competent Cells (*E. coli* HTS08 strain) using LB/Ampicillin plate agar. Finally, the confirmed colony containing the desired plasmid DNA were selected and sub-cultured in LB/ Ampicillin agar plates at 37°C for 12-16 hrs, then stored in 50% glycerol at - 80°C for further study.

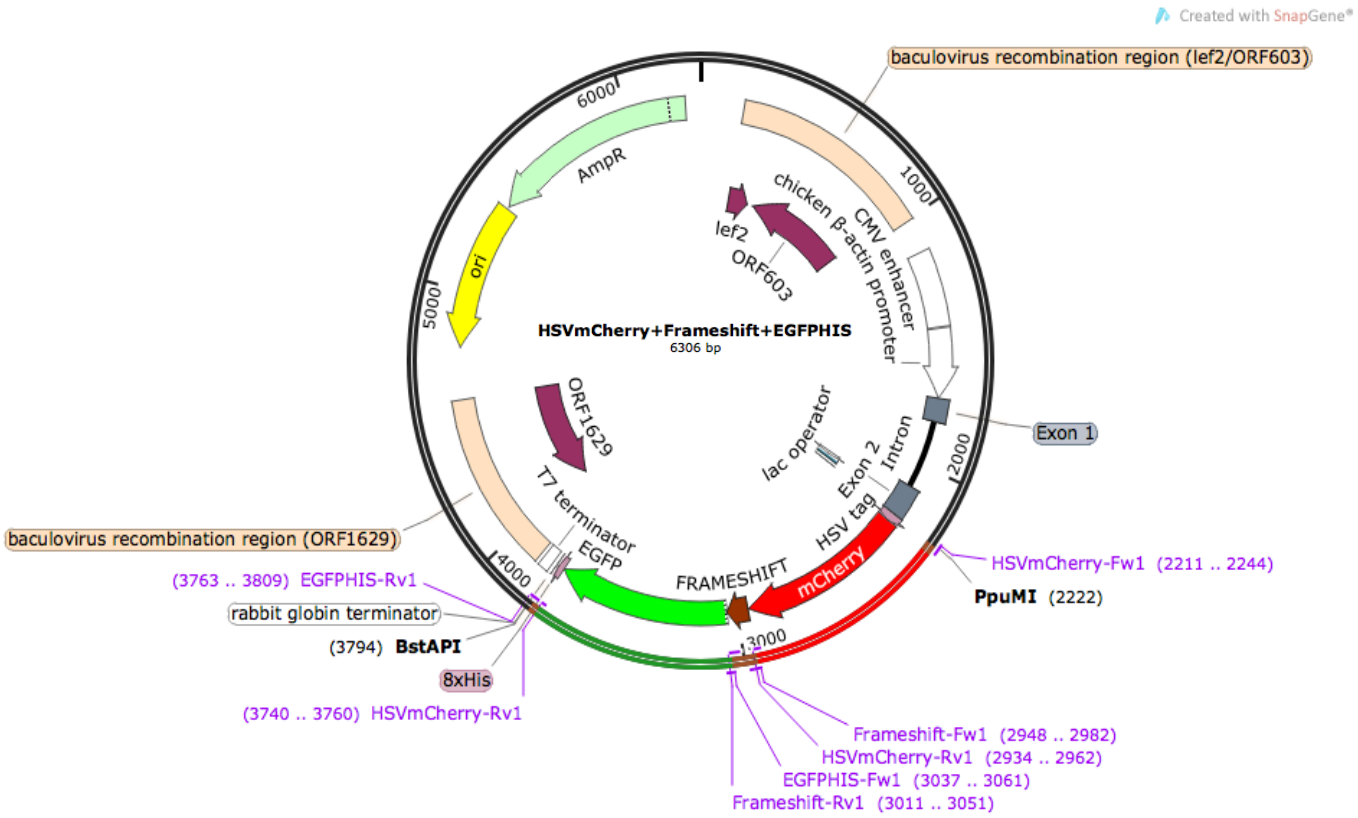

Figure 36 A. The cloning map for pTriEx1.1 with HSV-mCherry + Frameshift + eGFP-His. Key features of the plasmid are marked.

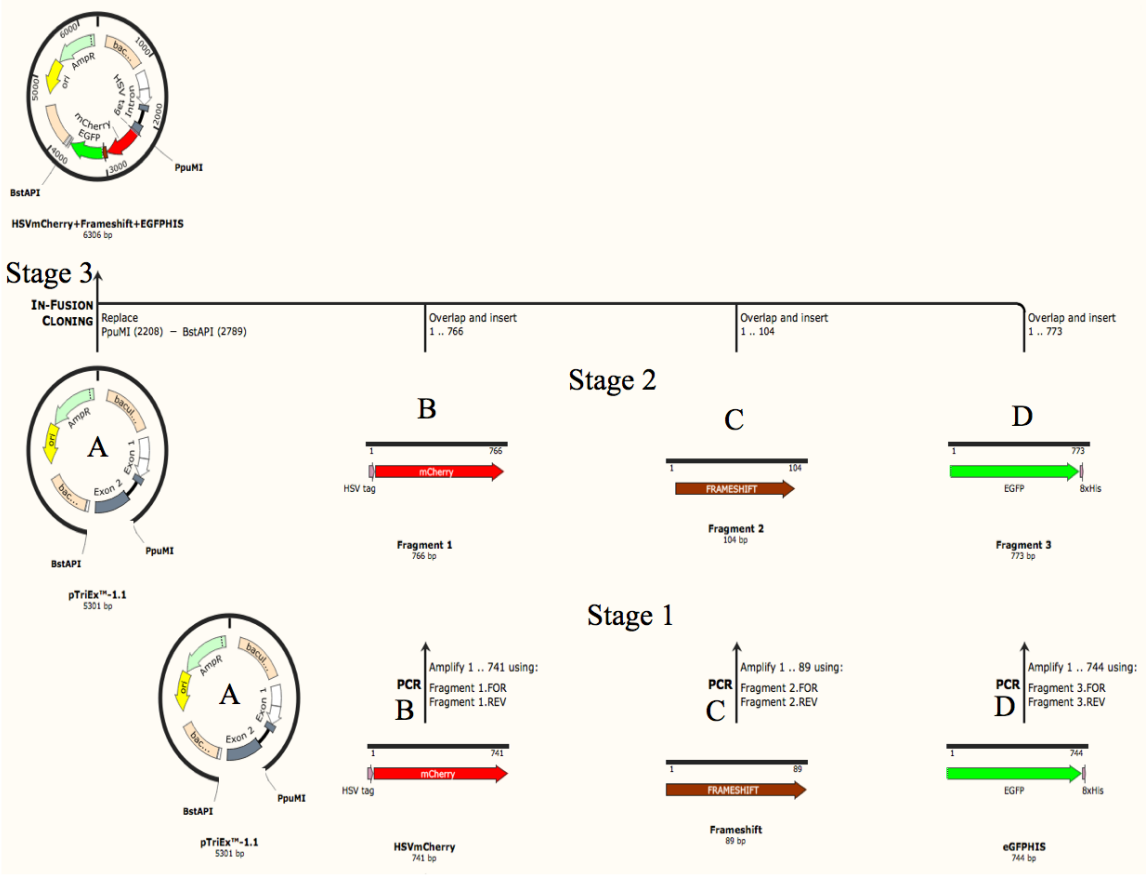

Figure 36 B. An overview of all stages used for construction of pTriEx1.1 with HSV-mCherry+Frameshift+eGFP-His throughout the entire infusion cloning methods. Stage 1: A- pTriEx1.1 vector double digested with *PpuMI* and *BstAPI* restriction enzymes, B- amplification for HSV-mCherry sequence from plasmid DNA containing HSV-mCherry (a gift from I. Jones), C- amplification for 1ab-frameshift sequences from plasmid DNA containing 1ab-frameshift (constructed in section 3.2.5), D- amplification for eGFP-His sequences from plasmid DNA containing eGFP-His (a gift from B. Abdulsattar), Stage 2: A- gel DNA extraction and purification of the pTriEx1.1 residual vector used for the infusion reactions, B- addition of overlapping 15 bp to HSV-mCherry (insert), C- addition of overlapping 15 bp to 1ab-frameshift (insert), D- addition of overlapping 15 bp to eGFP-His (insert), Stage 3: includes incubation for the infusion mix and transformation of Stellar ™ Competent Cells (*E. coli* HTS08 strain) using LB/Ampicillin plate agar. Finally, the confirmed colony containing the desired plasmid DNA were selected and sub-cultured in LB/ Ampicillin agar plates at 37°C for 12-16 hrs, then stored in 50% glycerol at - 80°C for further study.

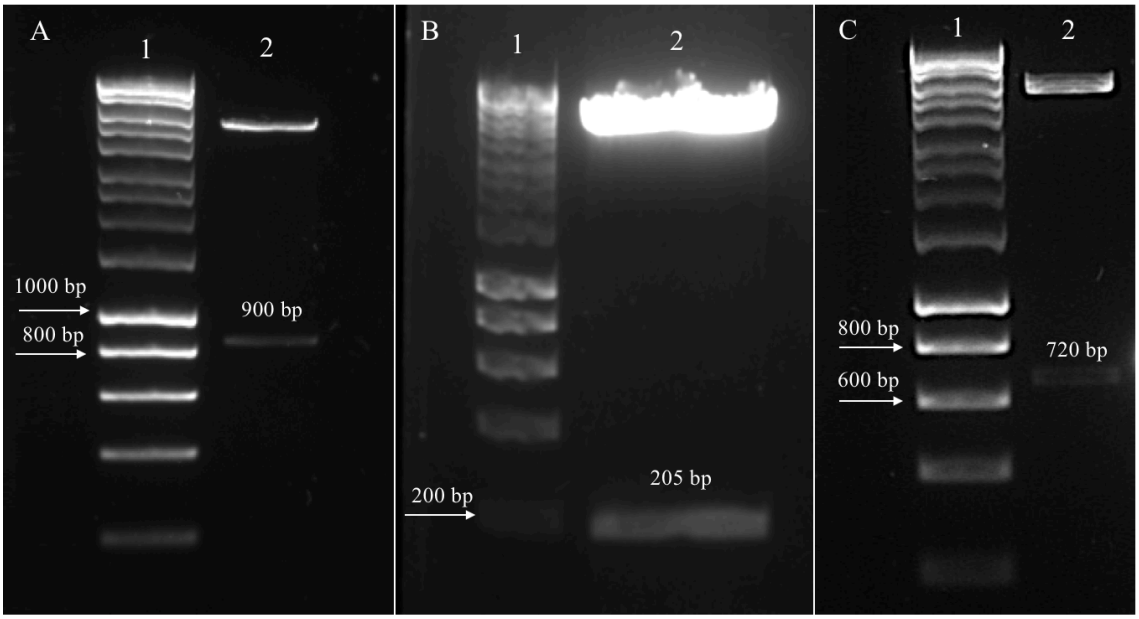

Figure 37. Gel electrophoresis of pTriEx 1.1 double digested with different restriction enzymes. (A) Lane1: Hyperladder 1kb, lane 2: pTriEx 1.1 double digested with *SacII* and *BstAPI* for 5'-UTR+eGFP, (B) Lane1: Hyperladder 1kb, pTriEx 1.1 double digested with *PpuMI* and *BstAPI* then *NcoI* and *Bsu36I* for RFP+Fs+eGFP construction, (C) Lane1: Hyperladder 1kb, pTriEx 1.1 double digested with *PpuMI* and *SphI* for eGFP+3'-UTR construction.

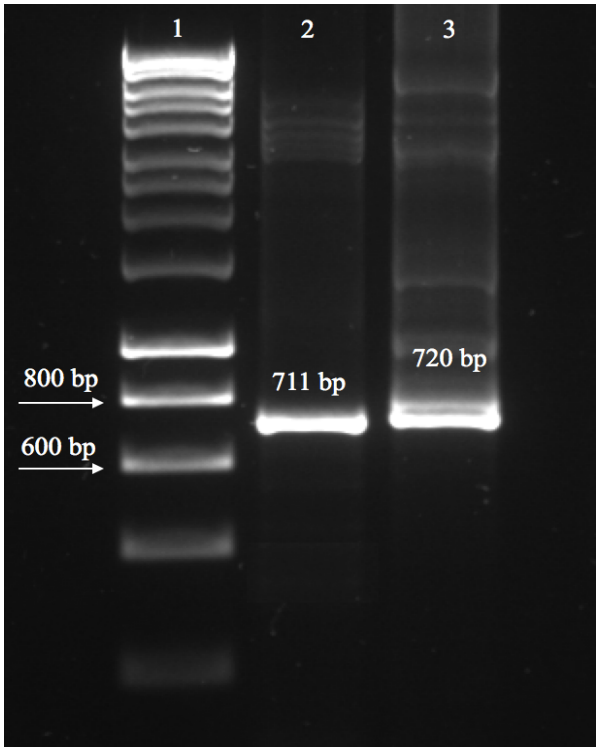

Figure 38. Gel electrophoresis of mCherry and eGFP high fidelity PCR products. Lane1: Hyperladder 1kb, lane 2: the mCherry amplified product, 3: the eGFP amplified product.

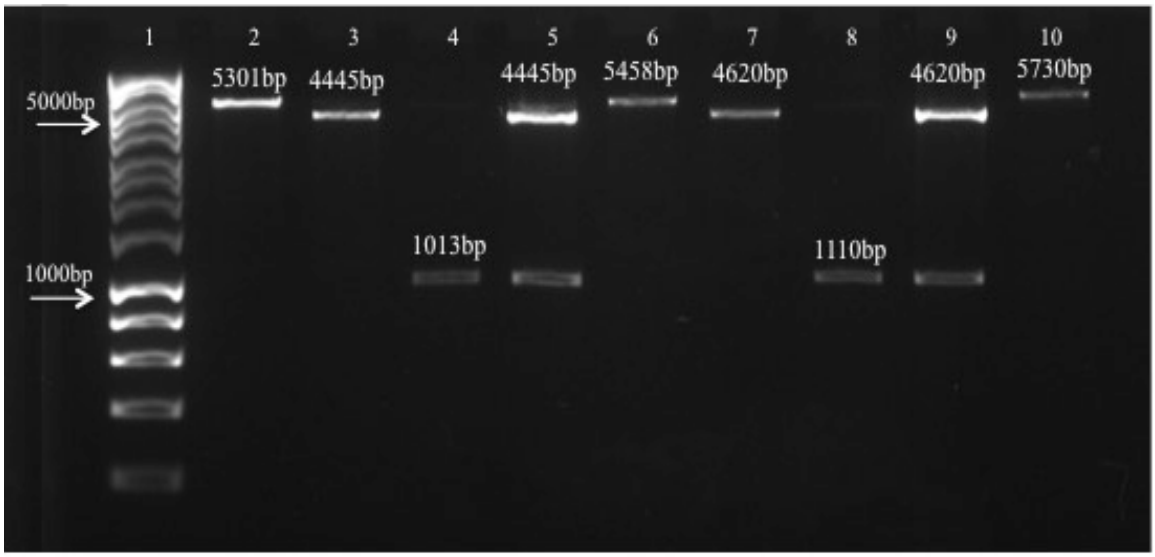

Figure 39. Gel electrophoresis of infusion stages for 5-'UTR+eGFP-His and eGFP-His+3'-UTR using pTriEx1.1 vector. Lane1: Hyperladder 1kb, lane 2: pTriEx 1.1 vector linearized with *SacII* restriction enzyme, lane 3: pTriEx 1.1 double digested with *SacII* and *BstAPI*, lane 4: 5'-UTR+eGFP-His (Insert), lane 5: double digestion for the plasmid DNA in order to confirm existing of the 5'-UTR+eGFP-His, lane 6: plasmid DNA linearized with *SacII* restriction enzyme, lane 7: pTriEx 1.1 double digested with *PpuMI* and *SphI*, lane 8: eGFP-His+3'-UTR (Insert), lane 9: double digestion for the plasmid DNA in order to confirm the existence of the eGFP-His+3'-UTR, lane 10: plasmid DNA linearized with *PpuMI* restriction enzyme.

3.2.7.2. Sequencing alignment of 5'-UTR+eGFP-His, eGFP-His+3'-UTR and HSV-mCherry + frameshift+eGFP-His of HCoV-HKU1.

Since it is known that any non-designed mutation could lead to aberrant results, sequencing analysis was performed according to Source BioScience website as mentioned in chapter 2 section (2.18). This ensured that there was no mutation or change in the upstream T7 promoter plus the 5'-UTR+eGFP-His, eGFP-His+3'-UTR and HSV-mCherry+ 1ABframeshift+eGFP-His sequences. This was performed by doing a comparison between each of the designed upstream T7 promoter plus the 5'-UTR+eGFP-His, eGFP-His+3'-UTR and HSV-mCherry+1AB-frameshift+eGFP-His sequences (the upper sequence) and the received data for the constructed plasmid DNA contained the desired fragment (the four coloured peaks). These red, green, blue and black coloured peaks represent the four DNA nucleotide T, A, C and G respectively. As is clear from the figure, there was no mutation or change in the orientation of each of the mentioned fragments and plasmids were found to be as designed **Figures 40**, **41**, **42A** and **42B**.

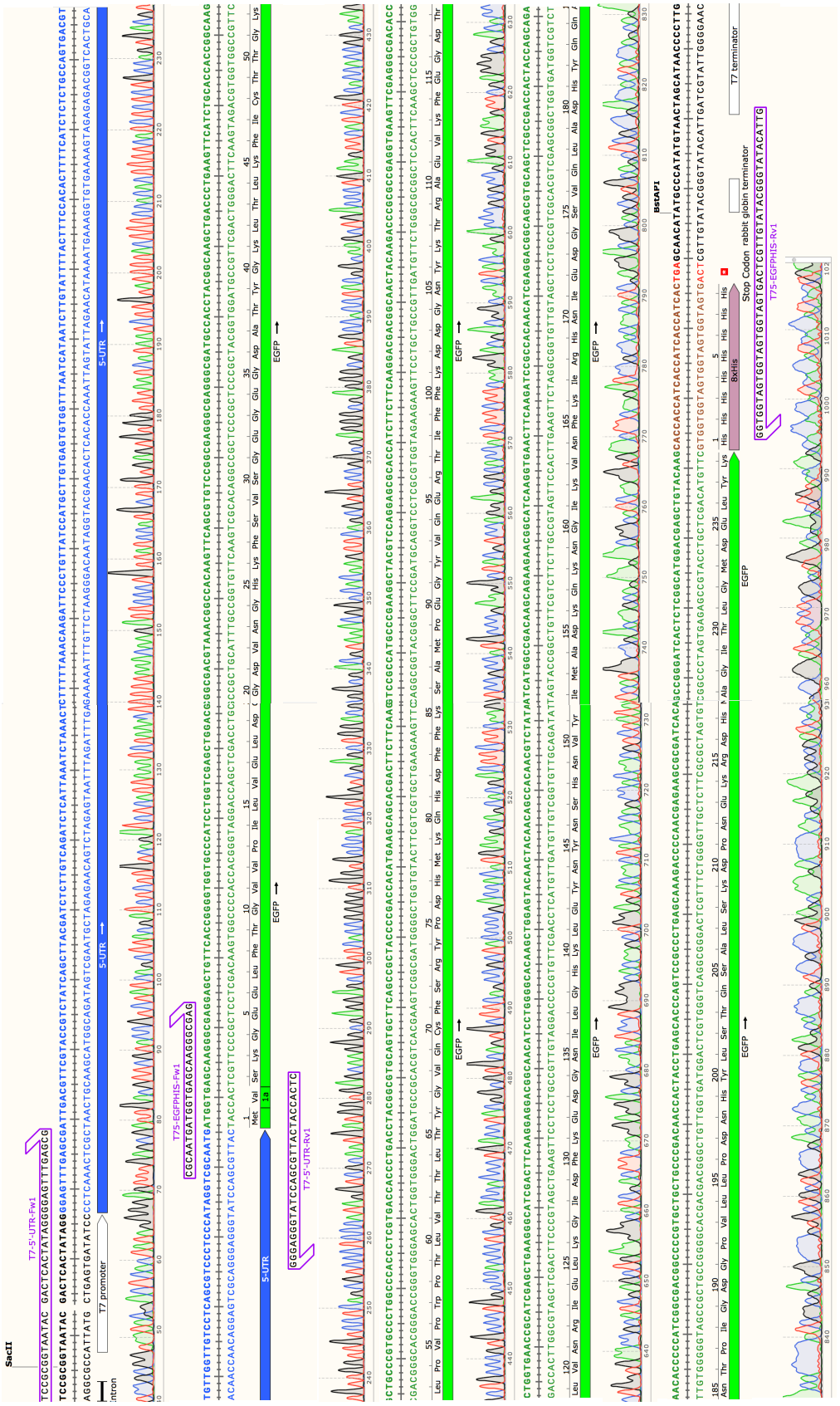

Figure 40. Sequence alignment of the pTriEx1.1+5' -UTR-eGFP-His with the synthesized sequence.

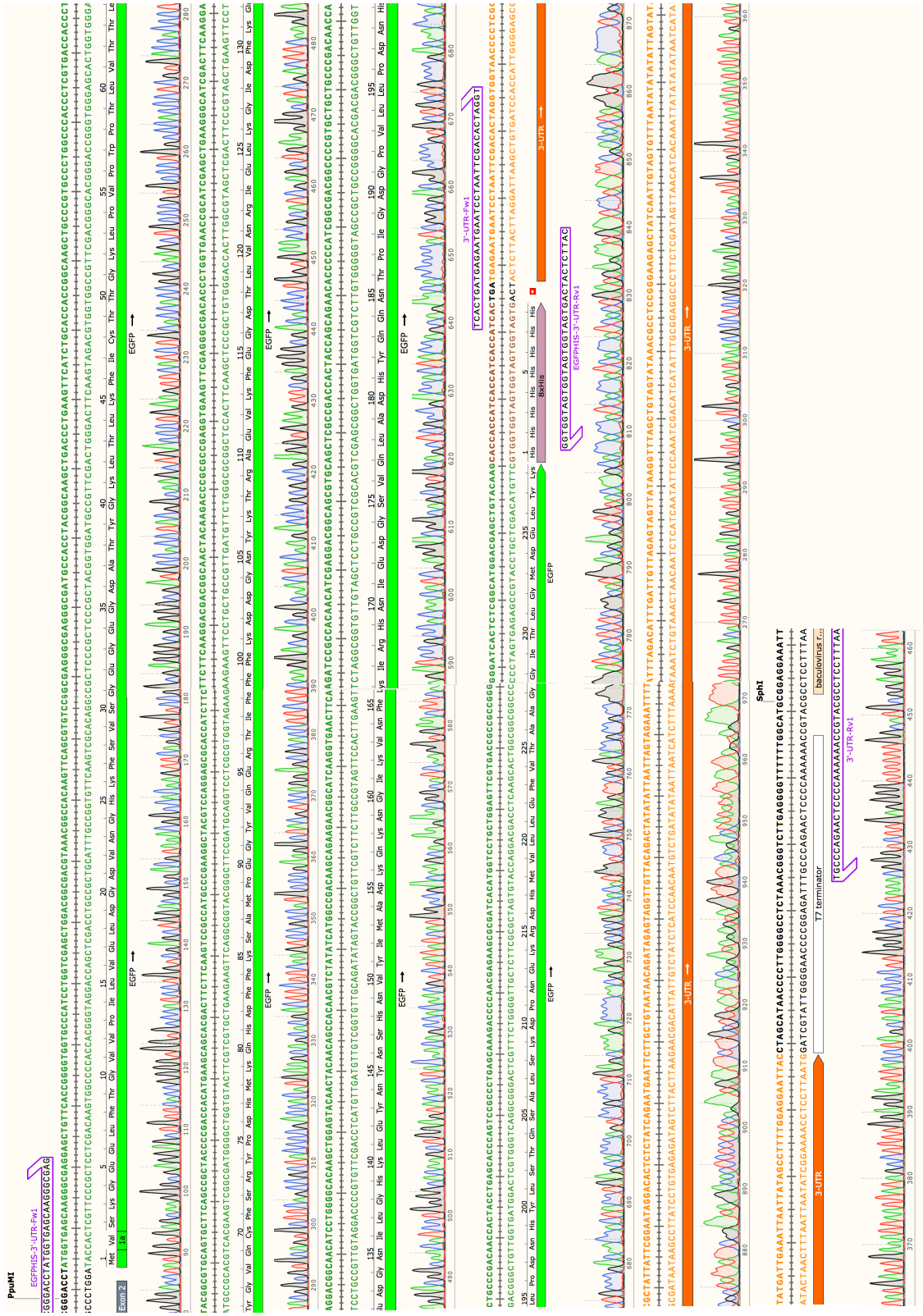

Figure 41. Sequence alignment of the pTriEx1.1+eGFP-His +3' -UTR-with the synthesized sequence.

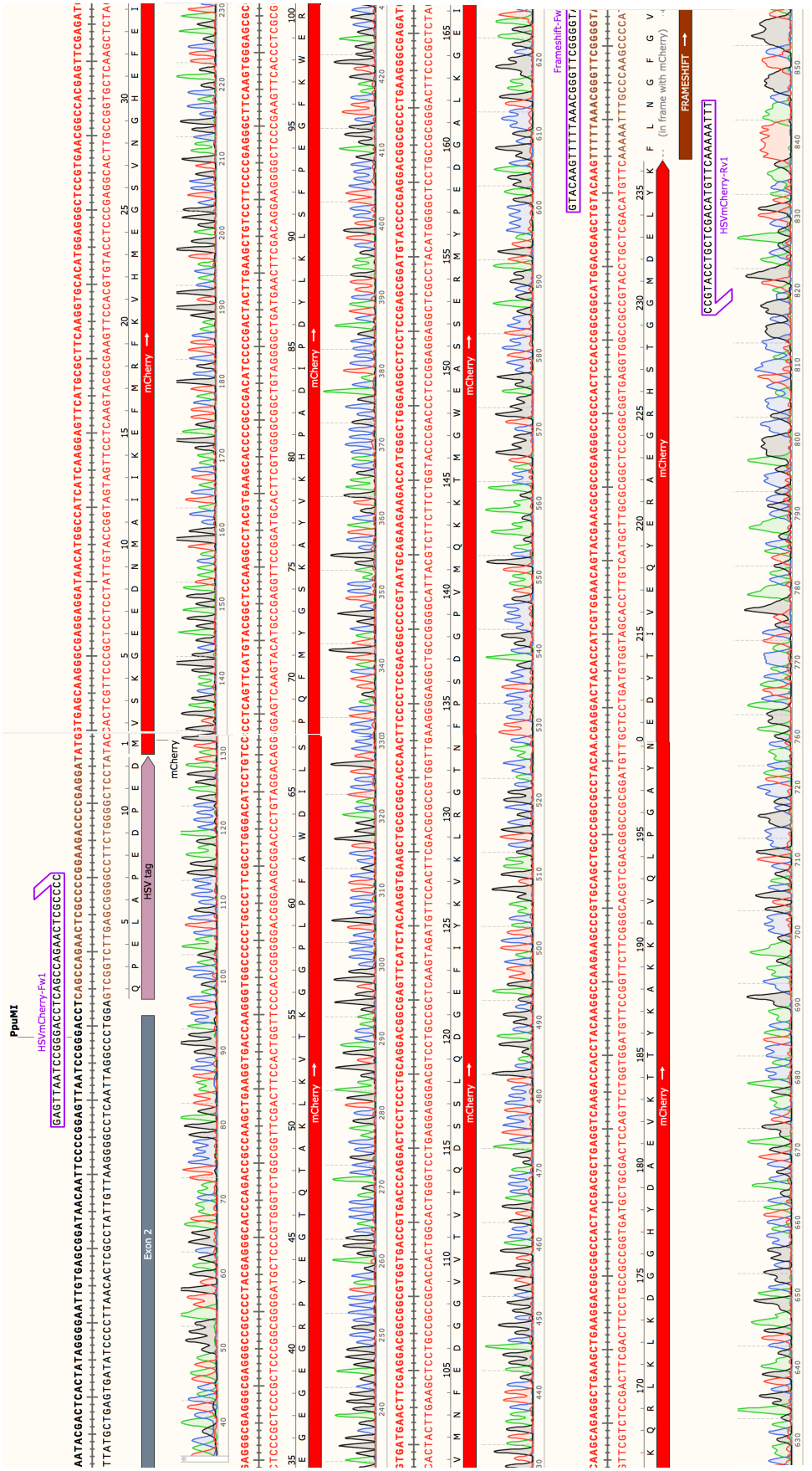

Figure 42A. Sequence alignment of the pTriEx1.1+HSV-mCherry+Frameshift+eGFP-His with the synthesized sequence.

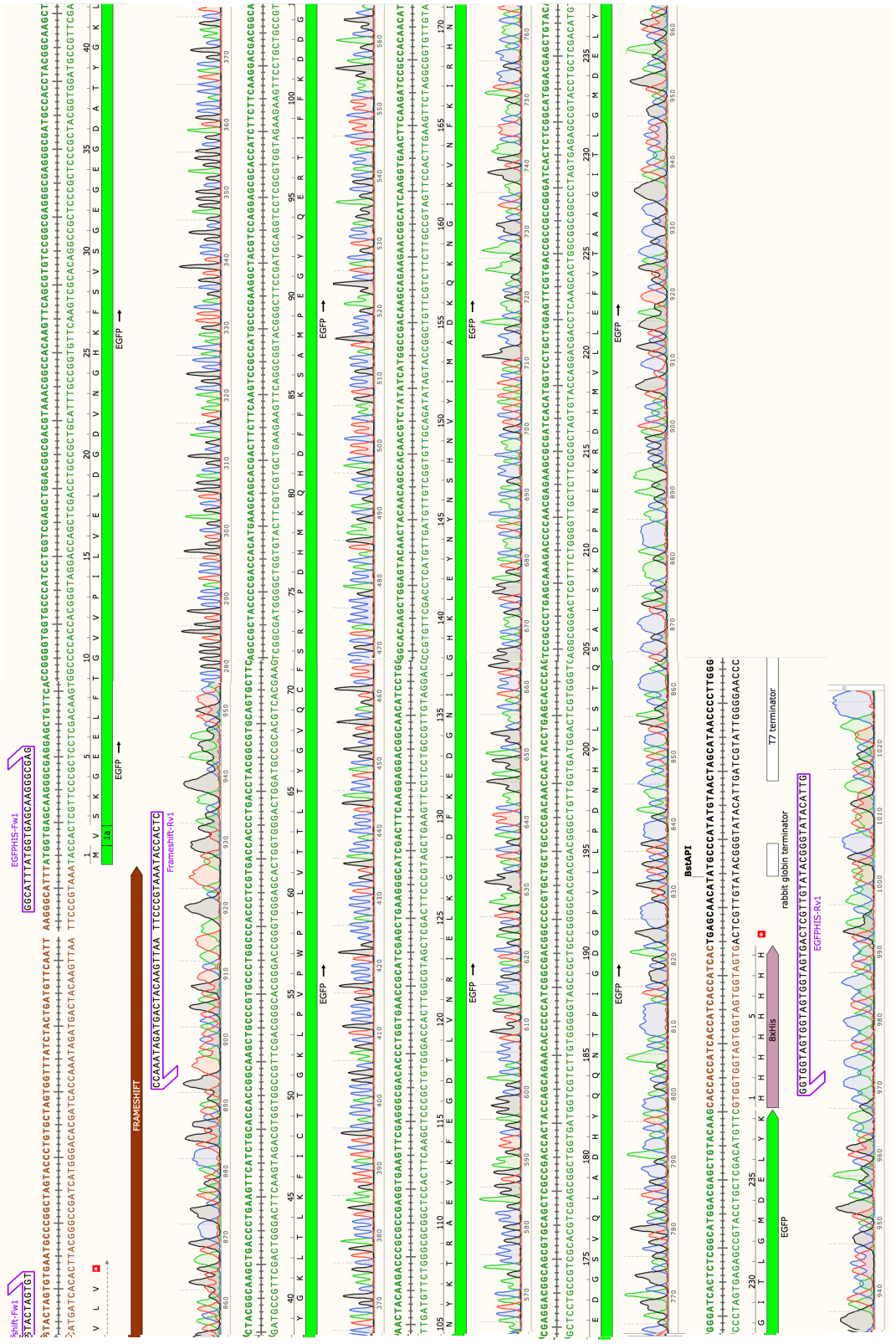

Figure 42B. Sequence alignment of the pTriEx1.1+HSV-mCherry+Frameshift+eGFP-His with the synthesized sequence.

3.3. Discussion.

There is limited knowledge of the RNA secondary structures of human coronavirus HKU1 despite there being 30 strains sequenced. This is because most studies to date of their role in transcription and translation of the virus have used mouse hepatitis virus, bovine coronavirus or SARS coronavirus. These studies can suggest functions for the analogous sequences in other coronaviruses but these then need to be confirmed by dedicated study. This chapter establishes the framework for such study, which will focus on HKU1 as a model for studying the structures noted above, by isolating and constructing genetic configurations suitable for later manipulation and read-out.

From the HKU1 dataset available in the databases it was observed that four, nine and six strains have a unique sequences the HCoV-HKU1 5'-UTR, 3'-UTR and frameshift respectively **Figures 19**, **20** and **21**. The variation results from nucleotide changes that occur during replication, frequently the change of C to T, T to C and G to A as shown in **Tables 13, 14** and **15** and which has an effect on base pairing interactions predicated for the region (Dirks *et al.*, 2004). Variation is to be expected as coronaviruses have a high level of mutations, which may be useful for them to escape from host defences and to rapidly adapt to new ecological niches. The mutations occur originally by inaccurate copying by the virus encoded polymerase with more extensive changes also caused by template switching during replication, leading to recombinant genomes. The polymerase itself lacks a proofreading 3' to 5' exonuclease domain (Steinhauer *et al.*, 1992), which results in lack of repair of any mismatched bases and, given the extreme size of the genome (*circa*. 30kb), all progeny genomes will contain many mutations.

The extent of mutation is such that it has recently been shown that a 3'-to-5' exoribonuclease activity, nsp14-ExoN, may act as a proofreading apparatus and that this activity is necessary to avoid error catastrophe in MHV and SARS (Denison *et al.*, 2011). Template switching happens when one cell is infected by different coronavirus strains and the polymerase starts to replicate one genome but then jumps to another to generate a new recombinant (Zuniga *et al.*, 2010). Despite this extensive variation at both nucleotide and genome level, the predicted secondary structure elements are found to be highly conserved, reflecting their importance for virus survival.

Despite relative conservation, RNA secondary structure prediction showed different structural folding for the unique isolates identified and indicated several examples of base pairing which did not follow the canonical rules of Watson-Crick for pairing up of the four nucleotides, that is [(A, U), (U, A), (G, C), (C, G), (G, U), (U, G)] (Dirks *et al.*, 2004). These differences include the predicted transcriptional regulatory sequence (UCUAAAC) sequence that located in SL-2 of the 5'-UTR which was predicted to be present in the SL-2 loop of all available HCoV-HKU1 except the reference sequence (NC_006577.2) where the entire SL-2 was predicted to be single-stranded **Figure 22**. Similarly, the frameshift sequence was shown in some strain to exhibit variation in a stem structure on loop 2 **Figure 23** (Woo *et al.*, 2005b). Likewise, the sequence of HCoV-HKU1 3'-UTR were predicted to vary in SL-3 and SL-4 in the Genotypes B and C respectively **Figure 24**.

The modelled variation is due to the increase in weaker base pairs such as G-U, A-A and U-U compared with C-G and A-U. As a result, these changed RNA molecules may not be as stable inside living cells and in order to gain stability, other parts of the single-stranded RNA chain may fold back on itself to

form a new RNA secondary structure. Such structures are highly varied and include stem loops of RNA and RNA with Pseudoknots, all likely to increase structural stability with minimum free energy. Many kinds of loops occur within two large groups of stem-loops and pseudo knots in which sub-loops may include hairpin loop, interior loop, multi loop, stack loop, bulge loop and external loop (Borkar and Mahajan, 2014).

For this complexity to be adequately explored suitable genetic constructs based on key folded RNA structures are required and this chapter described their creation. Successful cloning and sequencing was completed for overlapping PCR and infusion cloning products designed as described for the full length of HCoV-HKU1 5'-UTR, 3'-UTR and frameshift DNA fragments, all cloned into a plasmid, the vector pTriEx 1.1, which will act as the base vector as it is capable of multiple modes of transcription **Figures 25**, **26** and **27** respectively. To these additional sequences were added upstream and downstream of the cloned signal in order to provide a reporter plasmid to measure the activity of the fragment bearing the secondary structure **Figures 34 A**, **35 A** and **36 A** respectively.

These recombinant plasmids will be used in the subsequent chapters to study the role of sequence variation within the secondary structures on their function in the virus life cycle. This data should identify point which they share or differ with already reported functions and in turn suggest their candidacy as targets for intervention. The outline strategy for these constructs is to incorporate several mutations on each side of the 5'-UTR, 1ABFS or 3'-UTR sequences and to measure their activity in relation to the unmodified sequence in order to deduce the effect of the applied mutation on patterns of transcription and protein expression as will be explained in more detail in the following chapters.

Chapter 4. Functional analysis of HCoV-HKU1 5'-UTR and 3'-UTR.

4.1. Introduction.

As outlined in chapter 1, coronaviruses replication and transcription requires RNA motifs that are commonly located within the non-coding regions at the 5' and 3' ends of the genome (Lin *et al.*, 1994, Friebe *et al.*, 2005, Sawicki *et al.*, 2007). To study the function of these sequences, reporter genes (a DNA sequence that is used to 'tag' another DNA sequence of interest) have been constructed with the intention of using them to monitor expression level and, when combined with points mutations representing the diversity highlighted in chapter 3, the regulation effect of these secondary RNA structures. Reporter genes configured in a way can be easily monitored permitting the function of the 'target' sequence to be measured. For instance, eGFP can be used for such purposes. It is an enhanced green fluorescent protein (eGFP) gene, is a modified version of the green fluorescent protein gene of the jellyfish *Aequorea victoria*. It is a protein composed of 238 amino acid residues (26.9 kDa) that exhibits bright green fluorescence when exposed to light in the blue to ultraviolet range (excitation maximum = 488 nm; emission maximum = 507 nm).

Sequences flanking eGFP have been converted to a Kozak consensus translation initiation site to further increase the translation efficiency in eukaryotic cells (Prendergat and Mann, 1978) and a His-tag which is a poly histidine amino acid motif in the protein that consists of at least six histidine (His) residues, usually at the N- or C-terminus of the protein. It is also known as hexa histidine-tag, 6xHis-tag, His6 tag and by the trademarked name His-tag (Hengen, 1995).

eGFP and associated tags were used in this study as described in chapter 3 (3.2.5 and 3.2.7) (Wan. *et al.*, 2002) in order to measure the effect of sequences appended upstream. The designed plasmids were used with a cell free transcription and translation system for protein expression (Goren and Fox, 2008).

The TNT® quick coupled transcription/translation systems is an appropriate single-tube, coupled transcription/ translation reaction for these studies as it uses eukaryotic *in vitro* translation components. When compared with standard rabbit reticulocyte lysate systems the time and process for achieving *in vitro* translation is improved (Thompson and Pelham, 1979, Hemmer *et al.*, 1989). Standard rabbit reticulocyte systems make use of RNA synthesized in vitro from a choice of the SP6, T3 or T7 RNA promoters and requisite RNA polymerase. The TNT® quick coupled transcription/translation system shortens the process by combining the RNA polymerase, nucleotides, salts and ribonuclease inhibitor with the reticulocyte lysate solution in a single master mix.

Protein production can be improved by two- to six-fold in a 60- to 90-minute reaction by comparison with a standard in vitro rabbit reticulocyte lysate reaction using RNA templates.

Two configurations for transcription and translation of genes cloned downstream of either the T7 or SP6 RNA polymerase promoters are available in the TNT® quick coupled transcription/translation systems. They use 0.2–2.0µg of circular plasmid DNA comprising a T7 or SP6 promoter, or a PCR-generated fragment containing a T7 promoter, which is added to an aliquot of the TNT® mix and incubated in a total reaction of 50µl for 60–90 minutes at 30°C. The translated proteins are then resolved by SDS-polyacrylamide gel electrophoresis (SDS-

PAGE) and detected by western blot. Using this system, typical assays can be completed and results obtained in 5–6 hours (Taylor *et al.*, 2013).

Accordingly, coupled TNT reactions were assembled and programmed with purified plasmids encoding the 5' and 3' HKU1 UTRs coupled to eGFP-His. That is, 5'UTR-eGFP-His, and eGFP-His+3'UTR both as described in the preceding chapter, and their expression activity assessed by measuring the level of eGFP-His expression compared to a plasmid encoding eGFP-His alone.

4.2. Results.

4.2.1.1. Protein expression of HCoV-HKU1 5'-UTR and 3'-UTR.

In order to measure the effect of HCoV-HKU1 5'-UTR and 3'-UTR on protein expression, a comparison was made next to two controls, eGFP-His which is eGFP-His cloned into the pTriEx vector without any HKU1 sequence addition (positive control) and the other without any plasmid addition (negative control). An equal concentration of about 500 ng (1 µl) from 5'-UTR (474.6 ng/µl), 3'-UTR (452.7 ng/µl), eGFP-His (444.5 ng/µl) or D.W. control were added to each reaction mixture which includes 40 µl TNT® quick master mix, 1 µl Methionine (1mM) and X µl nuclease-free water to 50 µl of the total volume, then the mixture was incubated at 30°C (PCR thermocycler) for 90 minutes. The quick coupled T7 TNT system used was unmodified and the resultant proteins were detected by western blot with an anti His antibody as detailed in chapter 2 sections (2.19 and 2.20).

The western blot analysis showed a single band of the molecular weight predicted for eGFP in all test lanes which was absent from the negative control lane. The intensity of the band was greatest following the addition of 5'-UTR to the eGFP-His coding region, whereas the expression level of the construct with eGFP-His flanked by the 3'-UTR showed no effect on protein expression compared to the eGFP-His alone control - **Figure 43**. This data suggests augmented transcription or translation of eGFP-His dependent on the 5' UTR sequence.

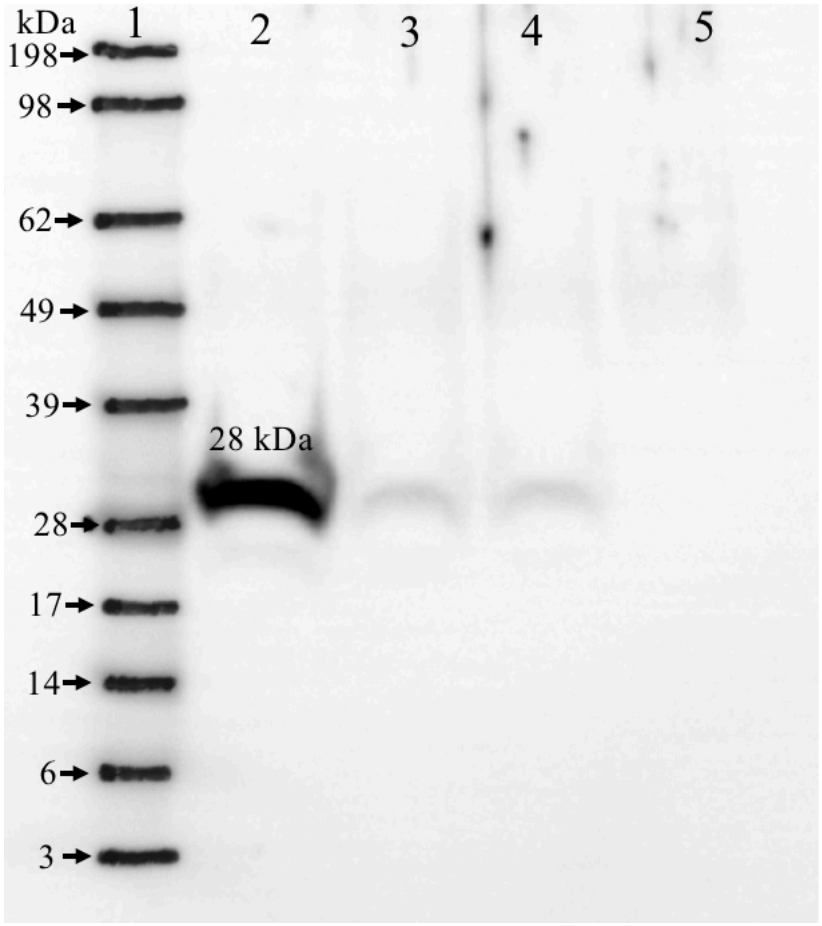

Figure 43. Western blotting analysis of HCoV-HKU1 5'-UTR+eGFP-His, eGFP-His+3'-UTR, eGFP-His and No-plasmid expression. Lane 1: prestained see blue plus standard ladder (Invitrogen), lane 2: 5'-UTR+eGFP-His, lane 3: eGFP-His+3'-UTR, lane 4: eGFP-His and lane 5: No-plasmid.

4.2.1.2. Relative expression of HCoV-HKU1 5'-UTR+GFP and eGFP+3'-UTR, GFP and No plasmid.

To confirm the protein expression levels observed in the previous section and to investigate the reproducibility of the reactions, three independents replicates were performed and image studio 4.0 software was used to analyse the subsequent western blots through calculation of area of intensity for each band for each group. The average of the replicates was plotted. In addition, a

statistical significance was measured by calculating the t test for each group and a comparison among each of the resulted p-value of 5'-UTR+eGFP-His and eGFP-His+3'-UTR with the p-value of eGFP-His only. The graph shows a significant increase in expression of the 5'-UTR+eGFP-His (*) when compared with eGFP-His only group (asterisk indicate p-value < 0.05) whereas no significant change of expression was observed for the eGFP+3'-UTR construct compared with eGFP alone **Figure 44**. These data confirm the stimulatory effect of the 5'UTR sequence and confirm that the construct design can be used for an investigation of the role of sequence variation in the 5'UTR on its activity.

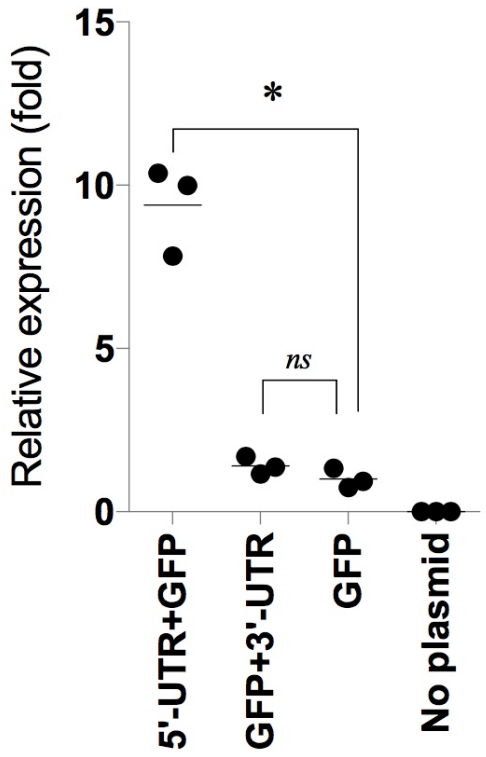

Figure 44. Relative expression of 5'-UTR+eGFP-His and eGFP-His+3'-UTR, eGFP-His only and No plasmid. The graph shows relative expression of the 5'-UTR+eGFP-His and eGFP-His+3'-UTR, eGFP-His only and no plasmid using quick TNT system as detected by western blot. Data are expressed as an average of three independent experiments. (*) was significant when compared with eGFP-His group, asterisk indicate p-value < 0.05 with the 5UTR+eGFP-His in t test (0.006), (ns) was not significant when compared with eGFP-His group, asterisk indicate p-value >0.05 with the eGFP-His+3UTR in t test (0.1587).

4.2.1.3. Construction of stems and loops mutation in the HCoV-HKU1 5'-UTR+eGFP- His plasmid.

A total of 26 oligonucleotide primers were designed for the introduction of 12 mutations into both sides of the stem loop structures of the 5'-UTR, in each of which C or G in the desired mutation side was changed to A in order to make a wobble pair, resulting in a predicated change in folded structure – see **Table 12** and **Figure 45,** based on the bioinformatics analysis described in chapter 3. The assembly of the mutants was performed by using overlapping high-fidelity PCR (as described in section 2.12) using one fragment amplified with a non-mutant forward primer with a mutant reverse primer **Figure 46A** and a second corresponding fragment amplified with a mutant forward primer and a non-mutant reverse primer **Figure 46B**. Following gel extraction and clean-up for each product, an overlapping high-fidelity PCR was used to amplify the final products **Figure 46C**. As before, gel extraction, clean-up and subcloning into the requisite pTriEx1.1 based plasmid, effectively replacing the wild type sequence with the mutant sequence, was done as described (2.15-17).

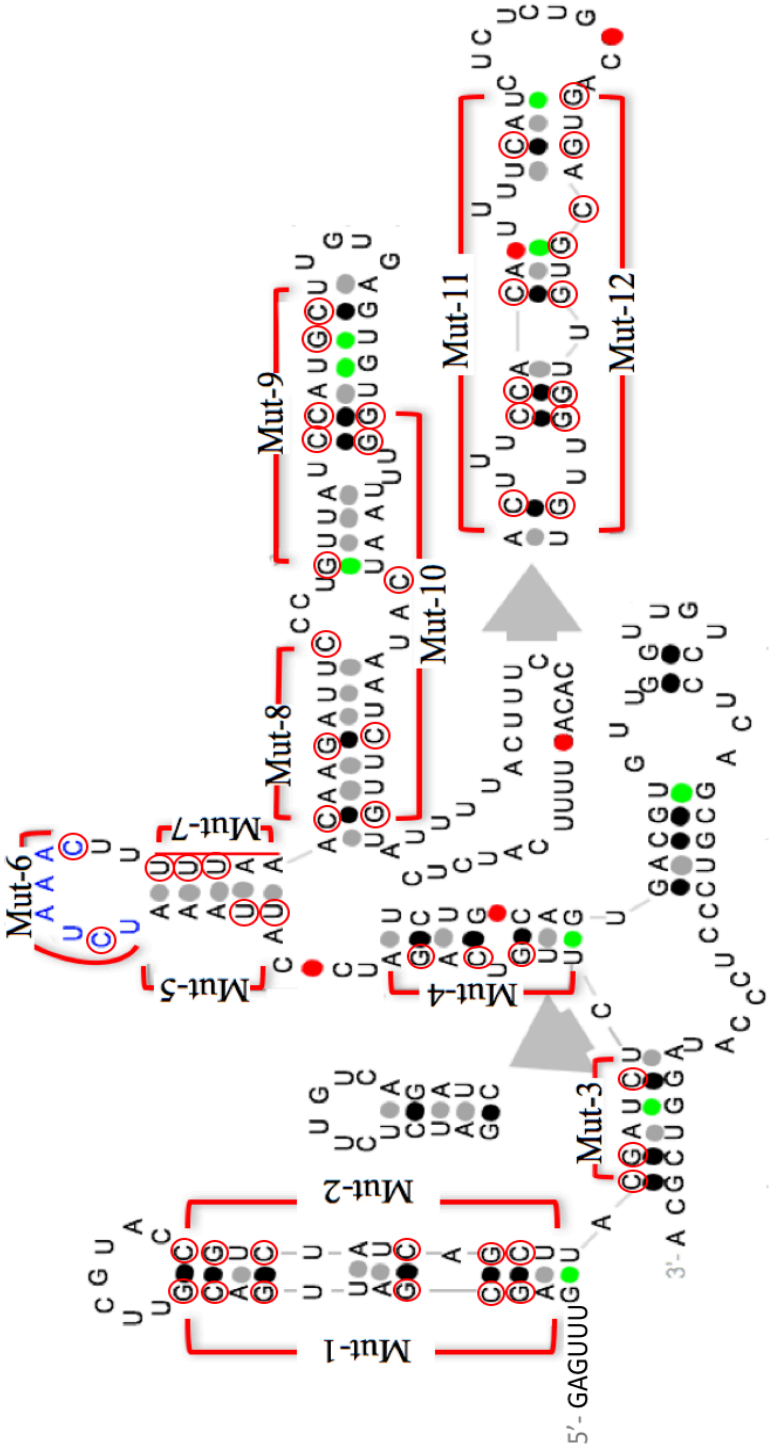

Figure 45. Construction of 12 desired mutation into the HCoV-HKU1 5'-UTR+eGFP-His in pTriEx 1.1 plasmid. Red coloured bracket includes the location of different desired mutations, red coloured circle includes nucleotide change from G or C to A in to HCoV-HKU1 5'-UTR sequence (WT). When the sequences used are aligned conserved or variant nucleotides are identified (see Figure 18) and are marked in the structure as (•) strong pair, (•) wobble pair, (•) weak pair, (•) change of C to U, (—) not consistently predicted.

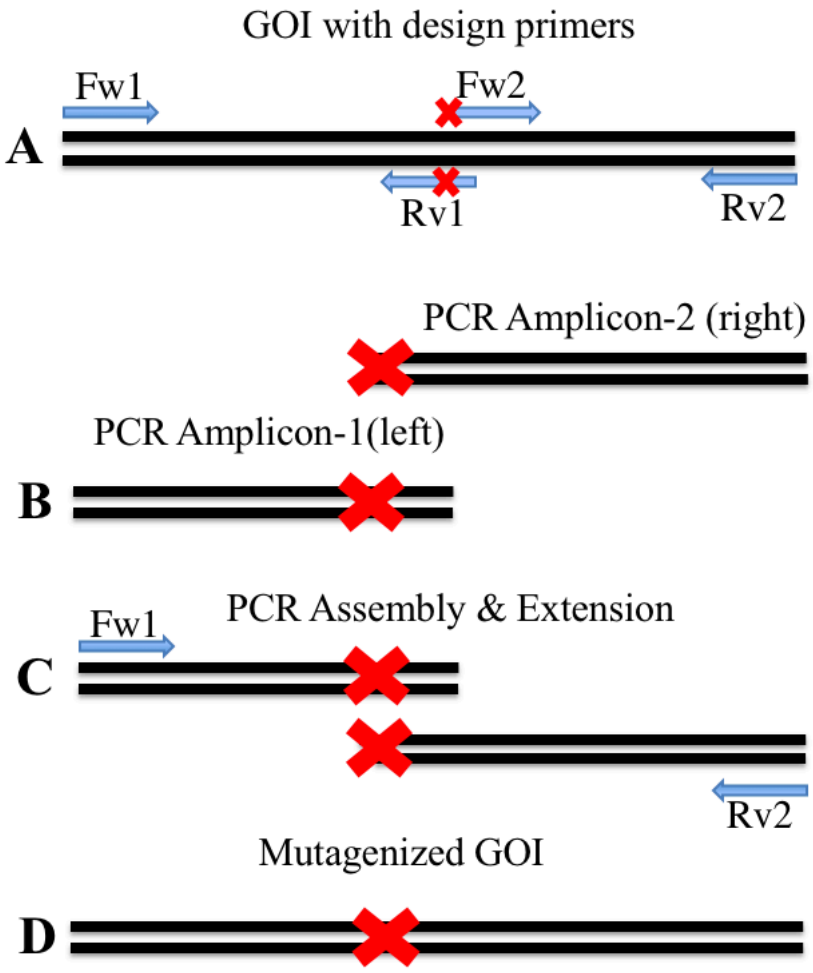

Figure 46. The multisteps of the mutagenic PCR strategy used to construct each mutation in the full-length 5'-UTR+eGFP-His mutations in pTriEx 1.1 plasmid. (A) Two high fidelity PCR amplifications of the gene of interest with the design primers, (B) DNA gel extract for the PCR products the left and right amplicons, (C) high fidelity PCR assembly and extension, (D) DNA gel extract for the PCR product result in the mutagenized gene of interest.

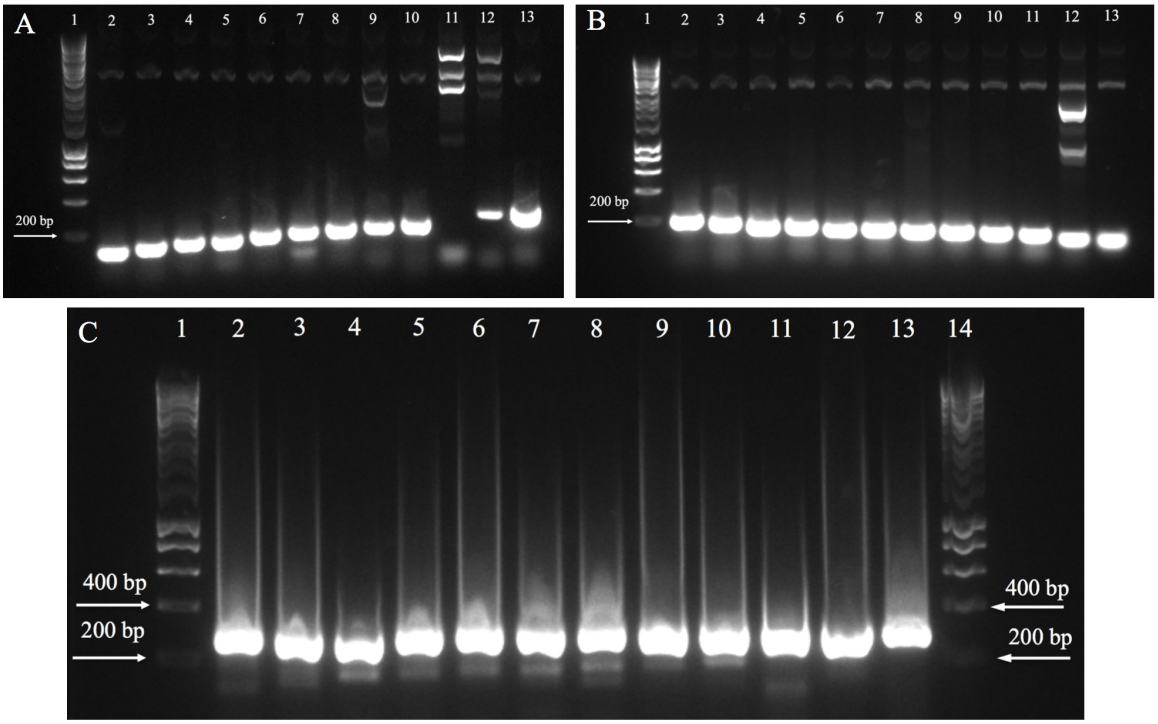

Figure 47. Gel electrophoresis of an overlapping PCR reaction for construction of HCoV-HKU1 5'-UTR+eGFP-His mutations in pTriEx 1.1 plasmid. (A) lane 1: Hyper ladder 1kb, lane 2-13: mutation reactions between non-mutant forward primer with each mutant reverse primer. (B) lane 1: hyper ladder 1kb, lane 2-13: mutations reactions between each mutant forward primer with non-mutant reverse primer. (C) lane 1 and 14: Hyper ladder 1kb, lane 2-13: mutations reactions between each two products from A and B.

4.2.1.4. Sequence alignment for constructed stems and loops mutation in the HCoV-HKU1 5'-UTR+eGFP-His plasmid.

DNA sequence analysis of each of the derived mutants was performed according to Source BioScience website as described in chapter 2 section (2.18) to ensure the presence of the desired mutation, the design of which was to probe the relevance of the highlighted region on function. For example, a nucleotide change from G and C in the target sequence to A will increase the number of wobble pairs in a predicated stem-loop and change the predicted stem or loop folding pattern on one side of the 5'-UTR structures. The sequences obtained for the desired mutations in the 5'-UTR sequence (the lower sequence in the figure) were aligned with the base vector 5'-UTR+eGFP-His (the upper sequence) in each case **Figure 48**. The colours of the trace files shown are standard; red, green, blue and black coloured peaks represent nucleotides T, A, C and G respectively. The data showed the presence of the desired mutations as designed in all of the clones tested.

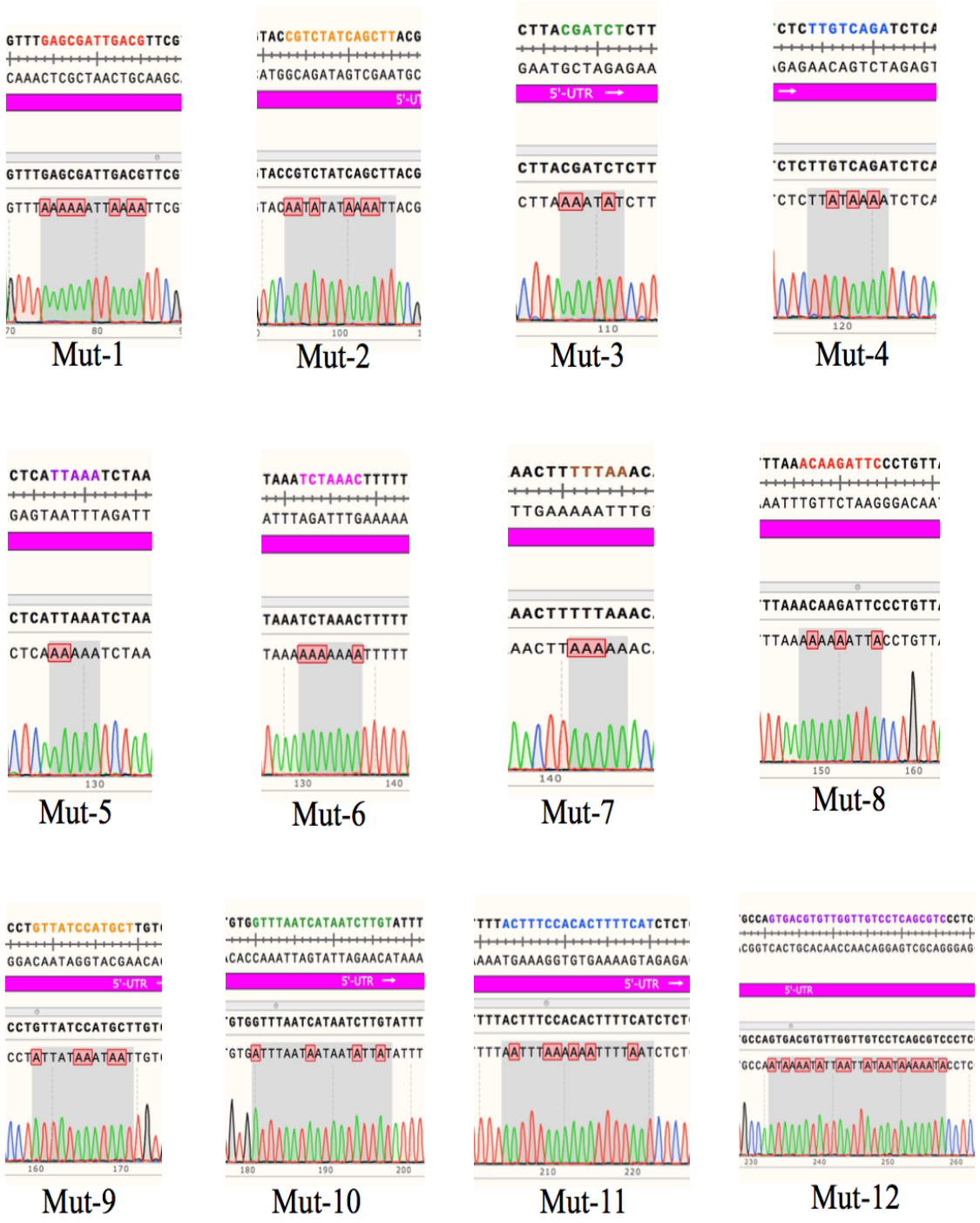

Figure 48. Confirmation of desired mutation into the constructed HCoV-HKU1 5'-UTR+eGFP-His in pTriEx 1.1 plasmid. Red coloured square represents nucleotide change from G or C to A compared to upper coloured sequence.

4.2.1.5. Effect of stems and lops mutation on protein expression of HCoV-HKU1 5'-UTR.

Following the data shown in the previous sections and after confirmation of the 12 desired mutations in both sides of the stems in the predicted stem-loops of HCoV-HKU1 5'-UTR, further eGFP-His protein expression comparisons were made among the 12 mutations. The un-mutated 5'-UTR+eGFP-His (wild type) was used as the baseline expression, as well as eGFP-His alone, acting as the positive expression control, and no plasmid addition acting as the negative control. In addition, a second plasmid was introduced into all the reactions, a plasmid encoding a His-tagged N13 protein of CoV-MHV (a gift from B. Abdulsattar), not linked a UTR sequence, to provide a loading control to be used to normalise the GFP expression data. This addition controls for the possibility that individual 5'UTR mutant plasmid preparations may contain inhibitors of the TNT reaction and so reduce overall translation. All reactions were performed by using the quick coupled T7 TNT system as before and the expressed eGFP-His product was detected by His tag western blot as described (2.19 and 2.20).

The western blot analysis demonstrated a range of protein expression levels among the 5'-UTR+eGFP-His mutations when compared to eGFP-His only and the parental sequence, whereas there was a constant level of protein expression by the sequence encoding the loading control (the 50 KDa N13 protein of CoV-MHV). The relative efficiency of the 5'UTR is all cases can be calculated by comparing the intensity of the 29kDa GFP band with the 50kDa N13 band in **Figure 49**.

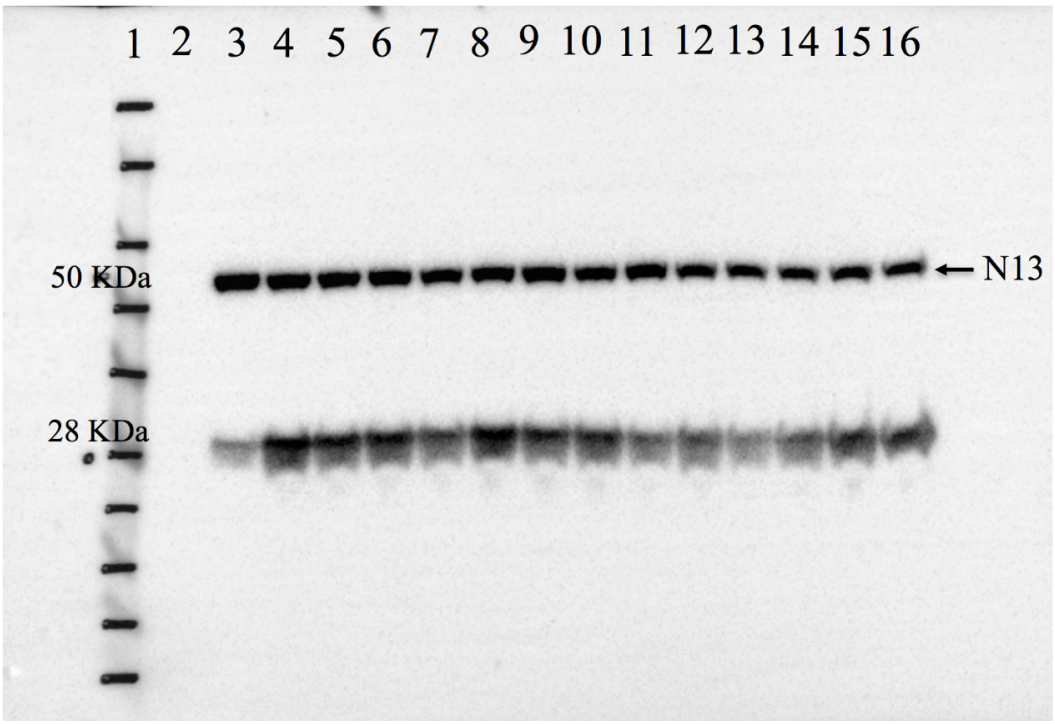

Figure 49. Western blotting analysis of HCoV-HKU1 5'-UTR+eGFP, 12 stems and loops mutations on 5'-UTR structures, eGFP and No-plasmid expression. Lane 1: prestained see blue plus standard ladder (Invitrogen), lane 2: No-plasmid, lane 3: eGFP, lanes 4-15: mutations on stems and loops of HCoV-HKU1 5'-UTR structure, lane 16: 5'-UTR+eGFP (wild type).

4.2.1.6. Normalized effect of stems and lops mutation on protein expression of HCoV-HKU1 5'-UTR.

Concerning the results above that showed a variation in protein expression downstream of HCoV-HKU1 5'-UTR when structural mutations were added, and in order to provide a level of reproducibility to them, three more independent experiments were preformed, and mean and standard deviation for each mutation was calculated using the GFP level normalised to the level of N13. The graph shows a notably skewing in the normalized data dependent on the sidedness of the mutations. Changes in nucleotide sequence in the left side of stem1, upper part of left side stem2, right side stem and the sequence between stem1 and 2 (mutations 1,5,2 and 3) showed a conspicuous increase in protein expression compared to the wild type with the actual stimulation being 45%, 37%, 33% and 31% respectively **Figure 50**.

In addition, mutations located in the upper part of the right side stem2, loop of stem2 and the lower part of left side stem2 (mutations 7,6 and 4) increased expression by 20%, 19% and 11% respectively. There is a slight change of protein expression on the right side stem4 (mutation 12) compared to the wild type - 7 % **Figure 50**. On other hand, there was generally a slight decrease in protein expression in mutation 9,11 and 8 which are located at the second part of the left side stem3, left side stem4 and the first part of left side stem3 compared to the wild type, by 9%, 13% and 16% of WT respectively **Figure 50**. Furthermore, there was a notable decrease in protein expression in mutation 10 which is located on the right side of stem3 compared to the wild type by 33 % **Figure 50**.

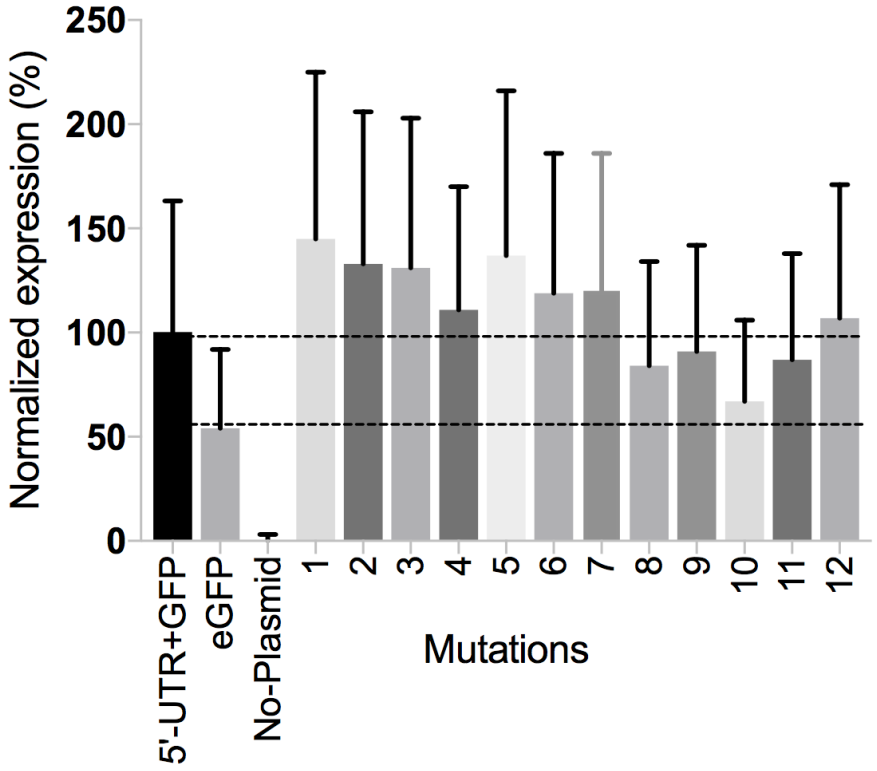

Figure 50. Normalized effect of stems and lops mutation on protein expression of HCoV-HKU1 5'-UTR. The average and SD of the expression level of the proteins are shown, mutations 1-12 referred to the location of mutated stem and loop in the HCoV-HKU1 5'-UTR.

4.3. Discussion.

It was a challenging task to study the effect of HCoV-HKU1 5'-UTR and 3'-UTR on protein expression due to the limited previous work about these structures in beta coronaviruses in general and HCoV-HKU1 specifically. However, what data there is gives a framework for the hypothesis that similar structures in HKU1 may exist and that their predicted folding is important for function. The successful construction of HCoV-HKU1 5'-UTR+eGFP-His and eGFP-His+3'-UTR was then completed as explained previously in chapter 3 to formally investigate these functions. This made use of the eGFP-His as a reporter gene as explained in chapter 3 (sections 3.2.5 and 3.2.7 respectively) (Wan. *et al.,* 2002). A cell free translation system was used for the measurement of protein expression (Goren and Fox, 2008) after considering the benefits of time, process saving and the increase in protein production by (two- to six-fold) by using of TNT® quick coupled transcription/translation systems (Hemmer *et al.*, 1989, Taylor *et al.*, 2013).

The results of the western blot analysis and the statistical analysis shows a significant variation of protein expression in which the addition of the 5'-UTR to eGFP lead to an increase in protein expression compared to eGFP alone (Elfakess and Dikstein, 2008), whereas the addition of 3'-UTR shows no effect on protein expression compared to eGFP **Figure 43** and **44** respectively. Considering the experimental set-up this variation could be explained by: 1) an increase of transcription, 2) an increased rate of translation or 3) an improved initiation of translation brought about by an internal ribosome binding site with the 5'UTR sequence. Addressing each possibility in turn, the experimental conditions and the constructed maps of plasmids HCoV-HKU1 5'-UTR+eGFP-His and

eGFP-His+3'-UTR indicates that all of them have the same T7 promoter and the same T7 RNA polymerase (provided by the TNT components) which leads to transcription changes being unlikely to be the cause of the effect observed.

Similarly, the ribosomes and other factors required for translation are all identical, which excludes this as a source of significant variation for the rate of translation. The most likely explanation is the existence of an internal ribosome entry site on the HCoV-HKU1 5'-UTR which increases the initiation rate of translation leading to more translated protein in a given time period. Ribosomes can be recruited to structured RNA elements, known as IRES elements, within the 5′ UTR of the mRNA in a process known as IRES-mediated translation. It has been found that the distance from the 3′-end of the IRES to the initiating codon differs depending on the origin of the IRES element. In IRES-mediated translation, the ribosomal subunit attaches to the 3′ end of the IRES structure and scans in a 5' to 3' direction along the mRNA until it reaches the initiating codon.

Once the 40S ribosomal subunit reaches the initiating codon, which is mostly (AUG), the 60S ribosomal subunit attaches to form an 80S ribosome that can decode the RNA into protein. The 40S subunit carries the initiator methionine-tRNA and certain eukaryotic initiation factors (eIFs). As a result of their function IRES elements have been termed "ribosome landing pads" (RLP) and have been found in several viral and cellular RNA elements. In the case of virus infection they are specially used to take over the translational apparatus of the host during viral infection when host specific translation is faltering as a result of host cell shutdown. Elsewhere they can initiate translation of specific mRNAs during cellular stress or other periods when general global translation is repressed (Thiel and Siddell, 1994, Hellen and Sarnow, 2001, Spahn *et al.*, 2004,

Moon *et al.*, 2016).

On examination of the HKU1 sequence it was found that a section of 13 nucleotides (AUUUAUUGUUUGG) had an element similar to the IRES related sequence described for MHV (UUUUAUUCUUUUU) (Woo *et al.*, 2005b). Positive strand RNA viruses such as those of the family Coronaviridae, contain several intra-molecular structures that are important for many viral processes and bear some similarity to more well characterised examples such as the 5'-UTR of hepatitis C virus (HCV) which also contains an internal ribosomal entry site (IRES), required for efficient translation of the viral polyprotein (Colussi *et al.*, 2015, Masante *et al.*, 2015). Similarly, picornavirus RNAs have unusually long 5' non-translated regions (5'NTRs) which contains a potential internal ribosomal entry site, e.g. the 5'NTR of encephalomyocarditis virus RNA has an IRES located between nucleotides 260 and 484 which plays a critical role in the efficient translation in both mono- and dicistronic mRNAs contain many non-initiating AUG triplets (Jang *et al*., 1988, Gale *et al.*, 2000).

In order to improve the accuracy of the analysis of HCoV-HKU1 5'-UTR gene expression analysis, normalization of gene expression data for the WT and the mutated sequences against a "housekeeping" gene (reference or internal control genes) was required. In cases of cellular translation this can be genes that are known to have constant levels of expression, such as the whole cell/cytoplasmic Alpha actin 43 and Alpha tubulin 55 KDa binding protein (TBP 38 KDa), ACTB, GAPDH or HPRT1. In the *in vitro* analysis performed here expression of the N 13 protein of MHV was used as a positive control for protein expression to ensure that any change in protein expression level observed for the mutant UTR constructs was solely related to control of eGFP translation

within the TNT system reaction. As with housekeeping gene expression the N13 band acted as a tool for normalizing gel loading differences and for Western blot quantification, which were measured by ImageJ and Image studio analysis of the western blot data (Thellin *et al.*, 1999, de Jonge *et al.*, 2007).

It was observed from the result of the western blot analysis and the statistical analysis of the replicates for the 12 constructed mutations in the 5'-UTR of HCoV-HKU1 that significant variation of protein expression occurred. Mutations generally clustered in the upstream half of the UTR were found to be stimulatory whilst those in the downstream half were generally inhibitory. As the introduced mutations were intended to destabilise stem-loop structures this would be consistent with the ribosome accessing the messenger in the upstream half and being stimulated by local secondary structure melting. However, as no direct RNA binding was measured this remains uncertain.

Mutation No. 10 in the downstream half of the sequence noticeably decreased protein expression when compared to the parental UTR construct almost to the level of GFP expression alone. The target sequence for mutation number 10 is located on the right side of stem3 (GTTTAATCATAATCTTGT) to (ATTTAATAATAATATTAT) and it reduced the number of C:G base pairs. This will undoubtedly affect the local stem-loop structure but could also have a more general effect on the secondary RNA structure of the UTR as a whole, explaining its particular effect. The localisation of points of significant downregulation may indicate targets for future antisense antiviral therapy (Moon *et al.*, 2016).

In general, antisense oligonucleotides (ASOs) are small synthetic pieces of single-stranded DNA that have the ability to selectively inhibit gene expression.

In principle, an oligonucleotide with a specified sequence of 17 nucleotides or more can be designed to hybridize specifically to any single gene within the human genome. Achieving inhibition at the gene or mRNA level is believed to be a much more efficient intervention in a disease process than inhibition at the protein level (Agrawal, 1992). ASOs act as gene-silencers, interfering with the virological processes to inhibit production of infectious virus particles in infected cells, potentially reducing cytopathic effects. This inhibition occurs through hybridization of the oligonucleotide to target gene sequences followed by DNA disruption or target the messenger RNA (mRNA) resulting in RNA degradation or blockage of translation (Stephenson and Zamecnik, 1978, Bitko and Barik, 2001)

In contrast to the 5' UTR, addition of the 3'UTR had no effect of the translation of eGFP. This would be consistent with a role in replication rather than translation as has been shown by others (Gale *et al.*, 2000, Yang and Leibowitz, 2015).

Overall the variation in protein expression between the 5'-UTR and 3'-UTR of HCoV-HKU1 compared to an unmodified eGFP standard demonstrates that useful functional data can be obtained by study in this way. As a result, further analysis of reporter protein expression via similar constructs will be explained in more detail in the next chapter.

Chapter 5. Kinetics of protein expression and ribosome RNA interaction analysis of HCoV-HKU1 5'- UTR and 3'-UTR.

5.1. Introduction.

Following what was observed in the previous chapter concerning the enhancement effect of the HKU1 5'-UTR on protein expression, a further two methods were used to visualise and study this variation. The first technique includes using the TNT® quick coupled transcription/translation systems as outlined in the previous two chapters but with modification to include sample collection at specific time points (Wan *et al.*, 2002, Goren and Fox, 2008). This provides a time course for the synthesised protein as measured by visualising the percentage of protein accumulation during a 90 minutes period. The question to be addressed by this simple variation is if the stimulation of translation observed to be dependent on the 5' UTR sequence is constant throughout the course of synthesis or if there is a particular phase of synthesis, for example initiation, where stimulation is most evident.

The second method was an electro mobility shift assay (EMSA) or gel retardation assay (Ryder *et al.*, 2008, Alves and Cuha, 2012). EMSA is a simple, fast, sensitive technique which depends on the observation of changes in the mobility of nucleic acid from free to bonded form as a protein-nucleic acid complex (Hellman and Fried, 2007, Holden and Tacon, 2011). In the case of protein binding to RNA, EMSA usually uses RNA synthesized *in vitro* from the T7 RNA polymerase (i.e. the 5'-UTR and 3'-UTR of HCoV-HKU1 in the circular plasmid DNA comprising a T7 promoter upstream of the cloned gene) and intact ribosomes from *E. coli* as the source of protein for the interaction (Balbas and Bolivar, 1990, Studier *et al.*, 1990).

EMSA involves migration of nucleic acids (RNA in this study) through a polyacrylamide gel in TBE buffer towards the anode. This electrophoretic migration through the gel depends on the molecular weight of the RNA, its three-dimensional shape and the physical properties of the gel substrate. Two of these factors, the molecular weight of the RNA and its shape in solution, change with interaction with a protein or protein complex, which modifies the RNA conformation and significantly increases the molecular weight to that of a ribonucleoprotein particle (RNA–protein complexes). Such changes lead to variance in the mobility observed in the gel (Stead and McDowall, 2007).

In addition, EMSA is a robust technique which is able to adapt to a wide range of conditions including labelled nucleic acid, very low concentration (0.1nM or less) and sample volume (20 µL or less) (Hellman and Fried, 2007). EMSA is often used to define critical interactions for regulation of transcription, translation, nuclear transport, control of DNA replication, DNA damage repair, RNA processing and maturation (Alves and Cuha, 2012). The use of EMSA in this chapter is to provide a direct measure of the ability of the modified 5' UTR to access the ribosome, potentially providing a mechanism for the stimulation observed.

5.2. Results

5.2.1. Kinetic Protein expression of HCoV-HKU1 5'-UTR and 3'-UTR.

With regard to obtaining a better understanding of the variation and accumulation of the translated products with time for both HCoV-HKU1 5'-UTR and 3'-UTR constructs during a standard protein expression reaction a comparison was made next to two controls, eGFP-His which is eGFP-His cloned into the pTriEx vector without any HKU1 sequence addition (positive control) and the other a no plasmid addition (negative control). An equal concentration of DNA from the 5'-UTR (474.6 ng/µl), 3'-UTR (452.7 ng/µl), eGFP-His (444.5 ng/µl) was added to each reaction mixture which included 40 µl TNT® quick master mix, 1 µl Methionine (1mM) and 8 µl nuclease-free water to 50 µl of total volume. The mixture was then incubated at 30°C for 90 minutes.

A sample of 10µl was withdrawn from the reaction mixture into 10 PCR tubes for collection of the synthesised protein at regular time points throughout the incubation and the samples placed directly on ice to arrest translation. Samples were either analysed directly or stored at -20°C ensuring no further reaction after the collection time. The samples were resolved by 10% SDS-PAGE and the resultant proteins were detected by western blot using an anti-His antibody as detailed in chapter 2 sections (2.19 and 2.20). The western blot analysis showed a single band of the predicted molecular weight for eGFP-His after 90 min in all test lanes which was absent from the negative control **Figure 51, 52** and **53**.

The intensity of the bands increased with time in all samples but the overall intensity of the bands varied with the construct in use. As before, eGFP-His flanked by the 5'-UTR gave the greatest signal whereas eGFP-His flanked by the 3'-UTR showed only a slight effect on protein expression compared to the eGFP-His alone control - **Figure 51, 52** and **53**.

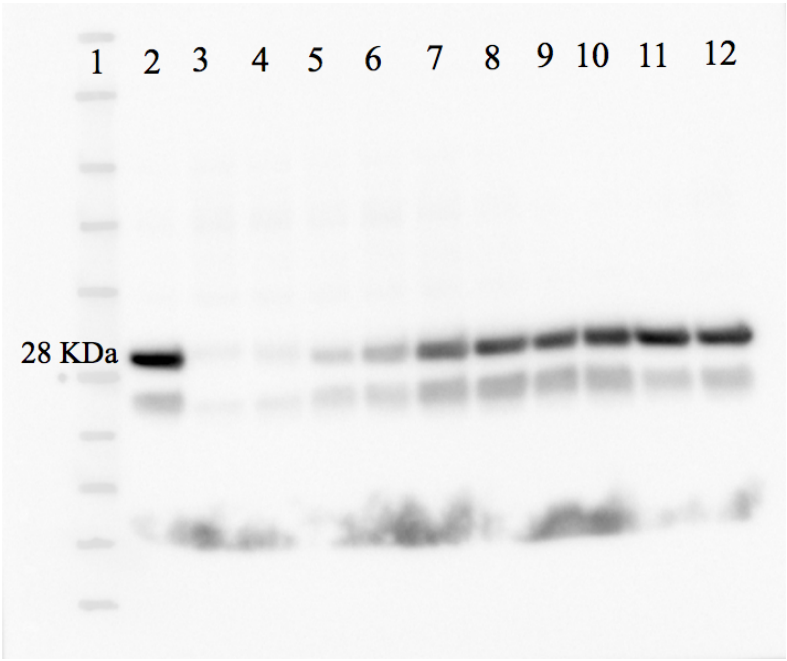

Figure 51. Kinetic western blotting analysis of eGFP-His and No-plasmid expression. Lane 1: prestained see blue plus standard ladder (Invitrogen), lane 2: eGFP-His expression after 90min, lane 3: No-plasmid expression after 90min, lane 4-12: eGFP-His expression at 10 minute intervals during 10-90 min respectively.

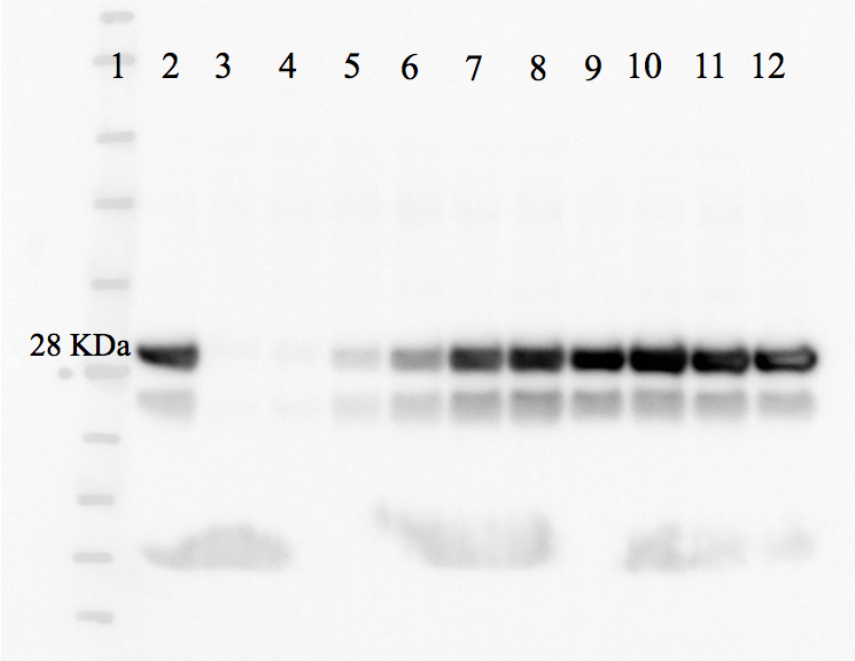

Figure 52. Kinetic western blotting analysis of HCoV-HKU1 5'-UTR+eGFP-His and No-plasmid expression. Lane 1: prestained see blue plus standard ladder (Invitrogen), lane 2: eGFP-His expression after 90min, lane 3: No-plasmid expression after 90min, lane 4-12: 5'-UTR+eGFP-His expression at 10 minute intervals during 10-90 min respectively.

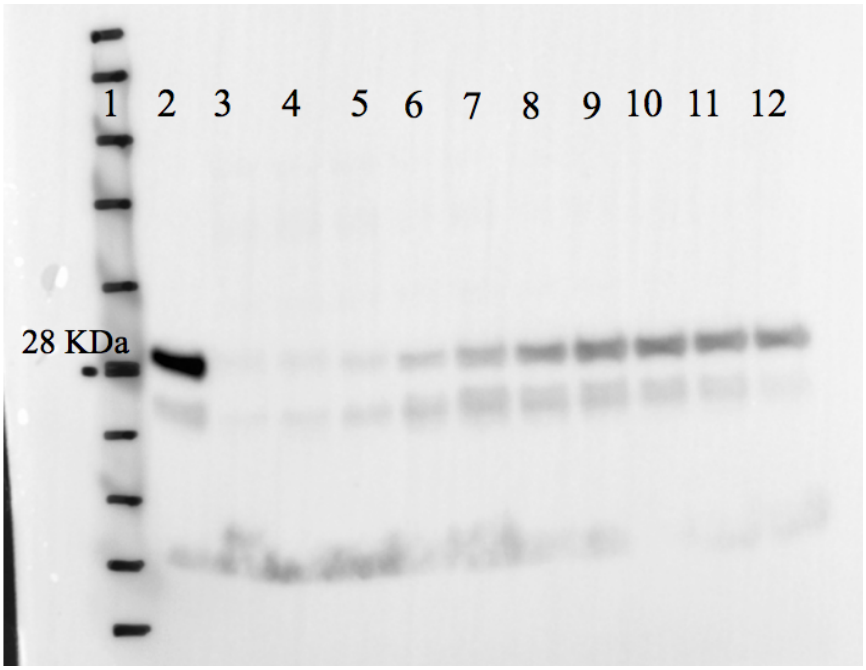

Figure 53. Kinetic western blotting analysis of HCoV-HKU1 eGFP-His+3'-UTR and No-plasmid expression. Lane 1: prestained see blue plus standard ladder (Invitrogen), lane 2: eGFP-His expression after 90min, lane 3: No-plasmid expression after 90min, lane 4-12: eGFP-His+3'-UTR expression at 10 minute intervals during 10-90 min respectively.

5.2.2. Normalisation for kinetic protein expression effects of HCoV-HKU1 5'-UTR and 3'-UTR.

The results of the previous section showed a gradual increase in protein expression for all constructs with time but a generally higher level of HCoV-HKU1 5'-UTR and 3'-UTR constructs when compared to eGFP-His as control. In order to confirm these results three independent experiments were preformed and the mean and standard error for the intensity of the bands for HCoV-HKU1 5'-UTR, 3'-UTR and eGFP-His were plotted. The graph confirms the improved kinetic profile of protein expression between the 5'-UTR+eGFP-His and the eGFP-His control. The increase was apparent after 20min of protein accumulation with the highest protein accumulation at 60min. This was followed by a slight decrease until the last time point **Figure 54**.

There was also a less significant increase in the levels of protein expression for the eGFP-His+3'-UTR and eGFP-His with similar kinetics overall. As for the 5'UTR construct there was a gradual decrease at later time points for both the eGFP-His+3'-UTR and eGFP-His **Figure 54** which may represent protein instability once the synthetic phase is exhausted. The maximum levels of translated eGFP were similar among the different constructs and probably represent the depletion of critical components such as amino acids. However, the kinetics were clearly different, plausibly reflecting an increased initiation rate for the ribosome consistent with entry to the initiator ATG being facilitated by the presence of a 5'UTR sequence. Accordingly, experiments were designed to address the direct interaction of the ribosome with the RNA transcript.

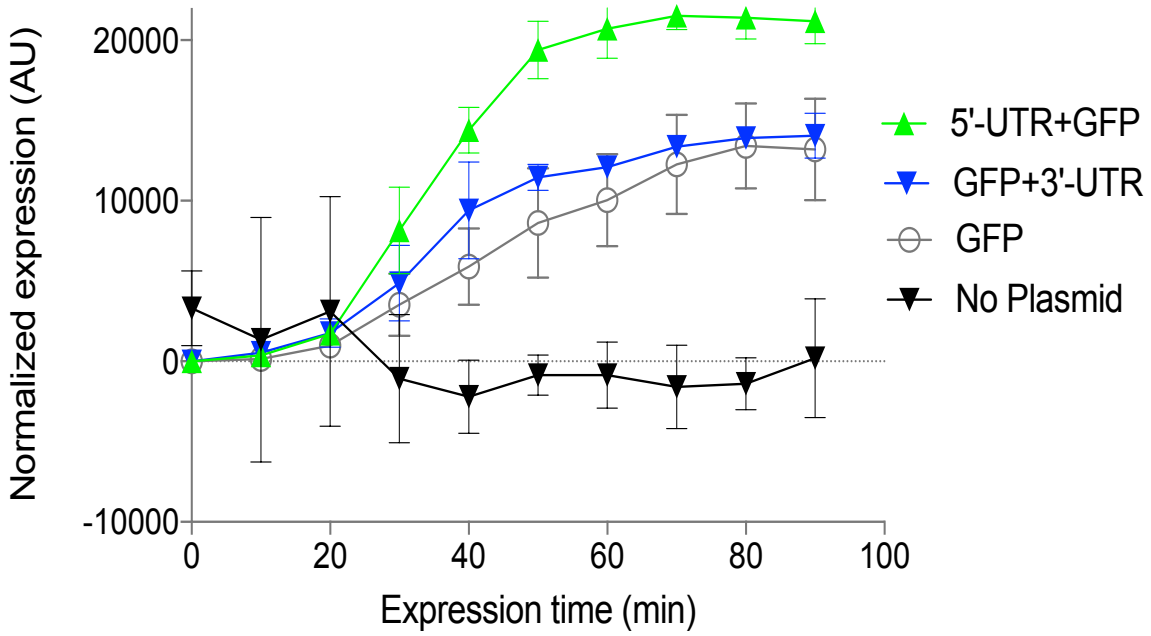

Figure 54. Normalized Kinetic western blotting analysis of HCoV-HKU1 5'-UTR+eGFP-His, eGFP-His+3'-UTR, eGFP-His and No-plasmid. Coloured key features of the expressed proteins are marked.

5.2.3. Transcription and purification for RNA of HCoV-HKU1 5'-UTR and 3'-UTR.

To provide RNA transcripts with which to carry out EMSA *in vitro* transcription reactions were conducted in accordance with the New England Biolabs manufacturer's protocol. This was done after construction of a set of plasmid DNAs containing the T7 promotor directly upstream of the HCoV-HKU1 5'-UTR or 3'-UTR and sited between the *Nco*I and *Xho*I restriction enzyme sites in the vector. Double digestion for each of the described plasmids with NcoI and XhoI restriction enzymes as mentioned in chapter 3 section (3.2.5) released the requisite UTR sequence appended with a T7 promoter. This was followed by gel electrophoresis to confirm the size of digested fragments and then purification using a DNA gel extraction and purification procedure as noted in chapter 2 section (2.13). An *in vitro* transcription reaction was then done for the purified digested fragment using the HiScribeT7 high yield RNA synthesis kit as mentioned in chapter 2 section (2.27) for 16 hours (overnight) in a PCR thermocycler at 37°C. The anticipated product in all cases is a RNA transcript (< 0.3 kb) encoding the HCoV-HKU1 5'-UTR or 3'-UTR.

RNA clean-up and concentration was done using the Norgen micro kit as a rapid method for micro RNA purification and concentration and depended on the manufacture's protocol without using phenol or chloroform as mentioned in chapter 2 section (2.28). The concentration of the resultant RNA of HCoV-HKU1 5'-UTR and 3'-UTR was determined by the use of Nanodrop spectrophotometer (ND-1000) in the RNA-40 channel. Concentrations for the two preparations were 459.8 and 226 ng/ µl respectively **Figure 55**.

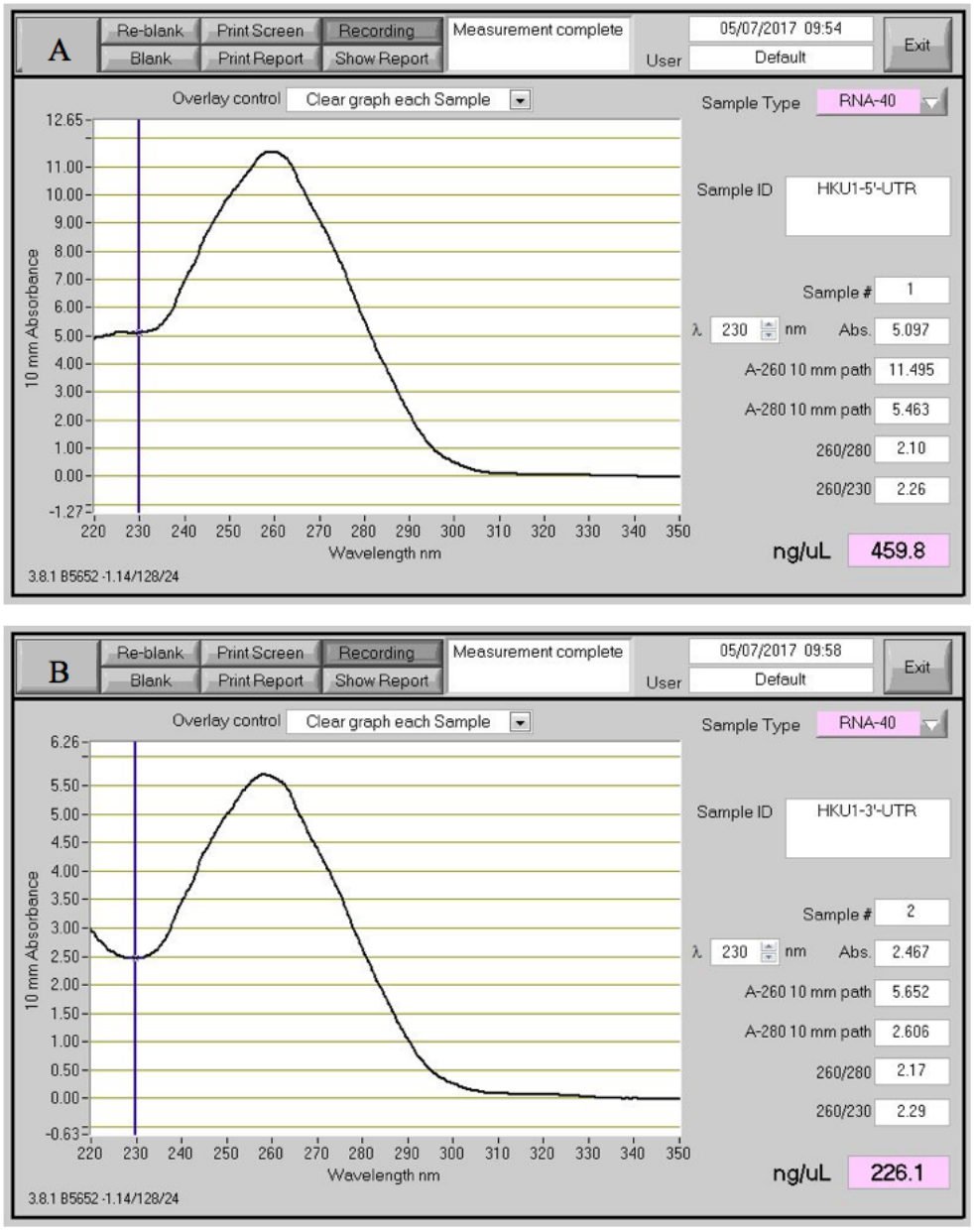

Figure 55. Determination the concentration and purity of HCoV-HKU1 5'-UTR, 3'-UTR. A and B: shows the concentration and high purity peak at 260/280 nm for the RNA of each HCoV-HKU1 5'-UTR and 3'-UTR using Nanodrop spectrophotometer (ND-1000) at RNA-40.

Following synthesis and clean-up the RNA was analysed by gel electrophoresis for each HCoV-HKU1 5'-UTR and 3'-UTR in a non-denaturing polyacrylamide gel which was performed by loading 1, 2 and 3µl from each of the concentrated 5'-UTR (459.8 ng/ µl) and 3'-UTR (226 ng/ µl) respectively onto a precast 6% Tris-glycine polyacrylamide gel and electrophoresis done at 120 V for 1hr in 1X Tris-glycine running buffer. The gel was stained with SYBR green or gold nucleic acid stain as previously mentioned in chapter 2 section (2.29) **Figures 56** in order to visualize the appropriate concentration of RNA for EMSA.

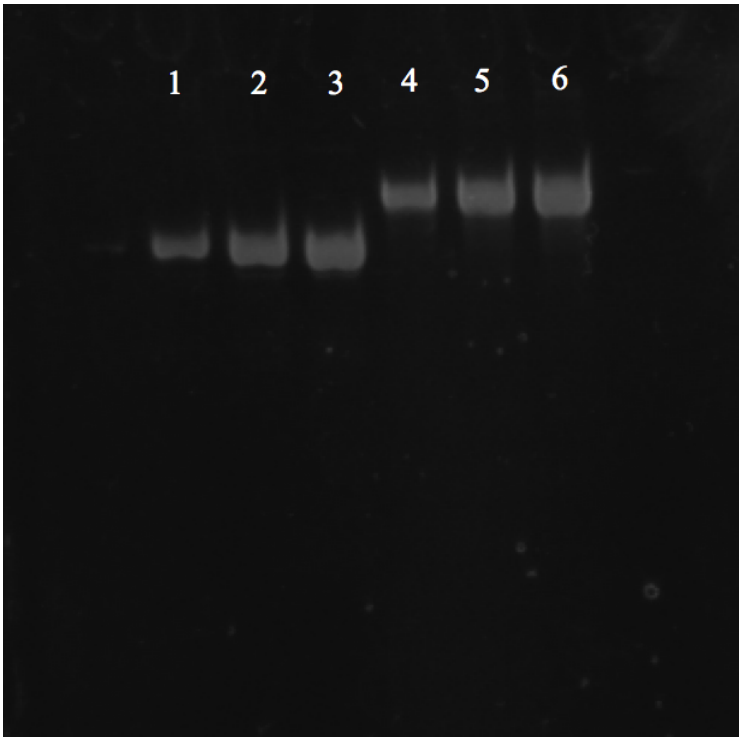

Figure 56. SYBR Green staining of RNA of HCoV-HKU1 5'-UTR. A nondenaturing polyacrylamide gel showing different loadings of RNA of HCoV-HKU1 5'-UTR and 3'-UTR lane 1-3 and 4-6 respectively.

5.2.4. Analysis of Ribosome RNA interaction of HCoV-HKU1 5'-UTR and 3'-UTR.

As cooperative activity including short and long-distance RNA-RNA interactions and conformational changes of RNA secondary structures could control viral replication and transcription, EMSA was performed with each of the HCoV-HKU1 5'-UTR and 3'-UTR purified RNA with intact ribosomes from *E coli* in order to investigate the possibility of a direct interaction for each of the above-mentioned fragments with the protein synthesis machinery. The reaction was performed as previously described in chapter 2 section (2.31) with modification in the use of an equal concentration about 500 ng from the 5'-UTR and the 3'-UTR (459.8 and 226 ng/ µl respectively) with 0.5-10 µg/ µl of intact ribosomes. The reaction was incubated for 30 min at 30 °C and mixed gently with 2 µl of 6X EMSA gel-loading solution. The RNA–ribosome complexes were separated by electrophoresis using a non-denaturing polyacrylamide gel which was run at 120 V for 1hr then stained with SYBR green (RNA) or SYPRO Ruby (protein) EMSA gel stains respectively.

The free RNA fragments are observed in all wells migrating at the position expected from the no protein addition control (lanes 7 and 14) **Figure 57**. In lanes where RNA and ribosomes had been mixed however additional bands appeared between the slot and the free RNA marker representing shifted products. That these bands were a mix of RNA and protein was confirmed by straining a parallel gel with a protein strain **Figure 57**. Despite these bands not resolving well on this gel system there was a quantitative change in the effective concentrations for positive band shifts with each RNA.

The results showed a better interaction between the 5'-UTR with the intact ribosome then the interaction with the 3'-UTR, in which the 5'-UTR (459.8 ng/µl) interact at 1.5µg/µl from the intact ribosome. Whereas, the 3'-UTR (452 ng/µl) interact at 5µg/µl from the intact ribosome (Red rectangle) **Figure 57**.

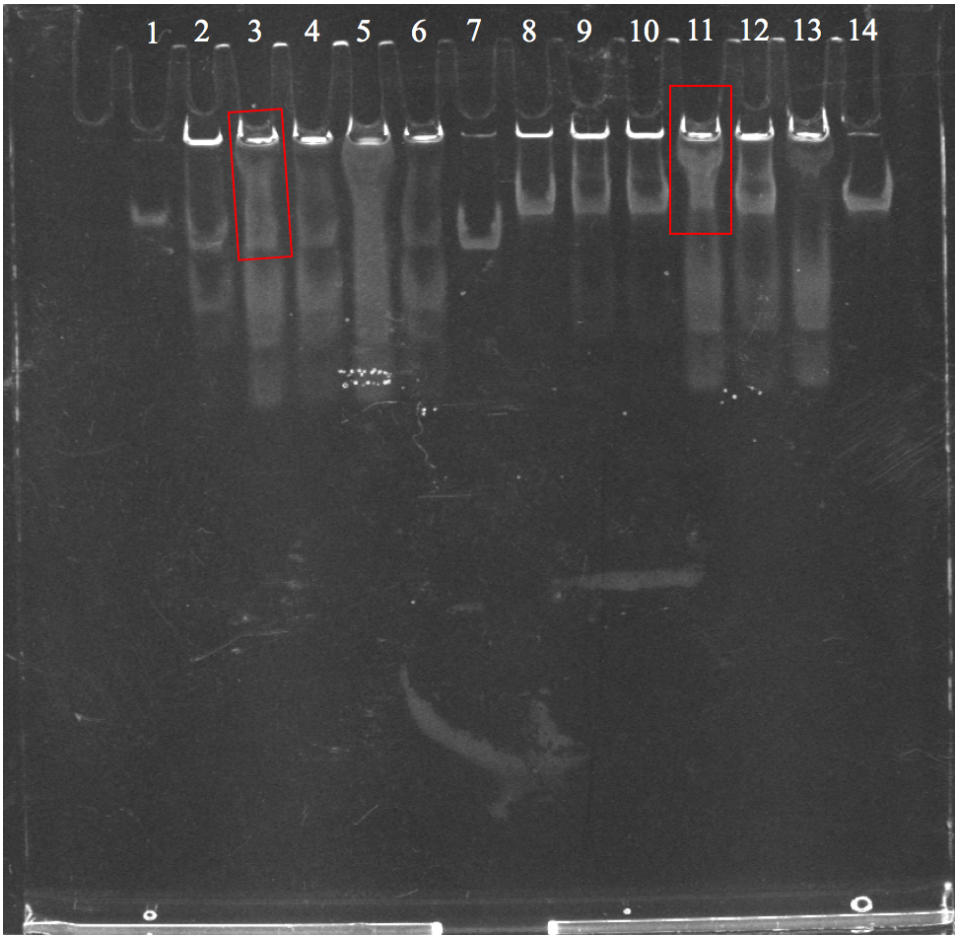

Figure 57. SYBR Green staining of RNA of HCoV-HKU1 5'-UTR and 3'UTR interaction with the intact ribosome. A nondenaturing polyacrylamide gel showing the interaction between the RNA of HCoV-HKU1 5'-UTR and 3'-UTR with the intact ribosome. lane 1: Unstained Protein Ladder (10- 200kDa Thermo Scientific), lane 2-6: 5'-UTR (459.8 ng/µl) interact with the 0.5-10µg/µl from the intact ribosome respectively, lane 7: 5'-UTR (459.8 ng/µl) only, lane 8-13: 3'-UTR (452 ng/µl) interact with the 0.5-10µg/µl from the intact ribosome respectively, lane 14: 3'-UTR only, the red rectangle shows the different interaction points between the 5'-UTR with the intact ribosome then the interaction with the 3'-UTR.

5.3. Discussion.

A course time was used in this part of the study to visualise the effect of HCoV-HKU1 5'-UTR and 3'-UTR on protein expression (eGFP-His) and to examine the optimal accumulation rate and time for protein expression. This is based on the benefits of using the eGFP-His as a reporter gene (Wan *et al.*, 2002) and TNT® quick coupled transcription/translation systems (Hemmer *et al.*, 1989, Taylor *et al.*, 2013). The result for normalized kinetic western blotting analysis of HCoV-HKU1 5'-UTR+eGFP-His and eGFP-His+3'-UTR supports the previous single time point results of protein expression which showed that the addition of 5'-UTR to the eGFP-His results in an increase in protein expression compared to eGFP-His (Elfakess and Dikstein, 2008).

By contrast the addition of the 3'-UTR showed only a slight effect, if any, on protein expression compared to eGFP-His **Figure 54**. The kinetic profile of the translation reactions showed an increased rate of synthesis when compared to plasmids lacking the 5'UTR consistent with a role in ribosome entry. This possibility was further supported by direct ribosome binding experiments which showed a clear difference in the concentration as which a band shift was observed for the 5'- versus 3'- UTR. The results of the EMSA, which is used to identify protein complexes with test RNA (Hellman and Fried, 2007, Alves and Cuha, 2012), found that the 5'-UTR of HCoV-HKU1 showed binding at 1.5µg/µl of intact ribosomes whereas the 3'-UTR (452 ng/µl) interacted at 5µg/µl of intact ribosomes or greater **Figure 57**. The bacterial ribosome is very large, 1340 KDa, which prevented sharp migration and an estimation of the molar ratios of the complexes formed.

However, if the IRES function of the 5'UTR is active then a 1:1 UTR RNA:ribosome complex would be expected. At higher concentrations, it is assumed that non-specific complexes form, as some shift was observed also for the 3' UTR at higher ribosome concentrations.

IRESes are found widely in the viral world. For example, mono- and dicistronic mRNA translation was shown to depend on the region in the 5'NTR of encephalomyocarditis virus (EMCV) which allowed high translational efficiency and/or independent translation of the viral genes (Jang *et al.*, 1988, Raman and Brian, 2005, Araujo *et al.*, 2012). This has been reported recently for the Enterovirus 71 Internal Ribosome Entry Site (Leteane, 2014) and is also studied in as shown by studies of translational control during viral infection of Severe Acute Respiratory Syndrome (SARS) non-structural protein 1 (Leteane, 2014)

As noted previously, an IRES in the 5'-UTR of HCoV-HKU1 would not be unexpected given their importance in eukaryotic translation initiation in other pathogenic viruses such as hepatitis A virus (HAV) (Glass and Summers, 1992), hepatitis C virus (HCV) (Tsukiyama-Kohara *et al.*, 1992) and foot-and-mouth-disease virus (FMDV) (Kuhn *et al.*, 1990). The complex stem–loop structure of the IRES is capable of engaging the 40S ribosomal subunit in a factor-independent fashion, which allows cap-independent translation (Kieft, 2008, Ray *et al.*, 2012).

As noted the 5'UTR could be a site that can be used as a target gene for complementary non-coding micro RNAs who's binding in this specific reign might lead to a decrease in the virus protein levels and inhibition of virus growth.

Chapter 6. Functional analysis of frameshift element in HCoV-HKU1.

6.1. Introduction.

As stated earlier, the frameshift element of HCoV-HKU1 occurs within an open reading frame (ORF 1ab) which is the only translation event of the incoming +ve strand virus genome. Translation of ORF1 result in polyprotein 1a (pp1a) and a −1 ribosomal frameshifting event leads to translation of ORF1b which results in the production of pp1ab (Sawicki *et al.*, 2007, Chen *et al.*, 2009). These two polyproteins are co- and post translationally processed into 16 non-structural proteins (nsps), most of them driving viral genome replication and subgenomic mRNA (sgmRNA) synthesis (Masters, 2006, Firth and Brierley, 2012). However, details of the frameshift reaction, for example its rate of shift and the role of cis-acting sequences, are lacking for the HKU1 sequence and need to be examined as part of a description of the complete replication cycle of this recently discovered member of the coronavirus family. To formally measure function of the frameshift, two different reporter genes (a DNA sequence that is used to 'tag' another DNA sequence of interest) were used to measure the translation of ORFs upstream and downstream of the frameshift secondary RNA structure.

The reporter genes used were chosen as they are easily monitored, and permit the function the 'target' sequence, in his case the HKU1 pp1ab junction, to be analysed. The arrangement of mCherry with a HSV tag upstream of the 1ab-frameshift and eGFP with a His tag downstream were used in this study as described in chapter 3 sections 3.2.7 and 3.2.7.1 (Wan *et al.*, 2002). Initially, the designed plasmids were used with a cell free transcription and translation system for protein expression (Goren and Fox, 2008).

As discussed earlier, the TNT® quick coupled transcription/translation systems is a single-tube, coupled transcription/ translation reaction which uses eukaryotic *in vitro* translation components. When compared with the standard rabbit reticulocyte lysate systems, where mRNA is transcribed separately and then added to a rabbit reticulocyte lysate (Thompson and Pelham, 1979, Hemmer *et al.*, 1989), the time and process for achieving *in vitro* translation is improved as the RNA polymerase, nucleotides, salts and ribonuclease inhibitor are all included in a single master mix. Variants of the system make use of RNA synthesized *in vitro* from one of the SP6, T3 or T7 RNA promoters by inclusion of the requisite RNA polymerase.

Using this system protein production can be improved by two- to six-fold in a 60- to 90-minute reaction by comparison with a standard *in vitro* rabbit reticulocyte lysate reaction with either of two common configurations, either the use of T7 or SP6 RNA polymerase. In all cases the reactions use 0.2–2.0μg of circular plasmid DNA including a T7 or SP6 promoter, or a PCR-generated fragment containing a T7 promoter, which is added to an aliquot of the TNT® mix and incubated in a total reaction of 50μl for 60–90 minutes at 30°C. The translated proteins are then resolved by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and detected by western blot. Using this system, typical assays can be completed and analysed in 5–6 hours (Taylor *et al.*, 2013).

A second method of analysis was expression of the test constructs in the insect cell line (*Sf*9) as an alternative way for functional analysis of the 1ab-frameshifting event. This method uses recombinant baculovirus expressing the target frameshift element that are constructed with the same fluorescence reporter gene constructs (Kain *et al.*, 1995). This is possible as the base vector

used for the construction of the test plasmids was pTriEx1.1, designed for expression in insect cells in addition to its ability to be transcribed by the T7 polymerase. Unlike the TNT system the levels of protein synthesis in the insect cell system can be very high, allowing direct observation of protein tag fluorescence after illumination with UV light. When excited with UV light with a wavelength of 488nm the green fluorescent protein (eGFP) from the jellyfish *Aequorea victoria* emits bright green light (lambda max = 509 nm) and does not require a substrate or cofactor. eGFP fluorescence is also stable, species-independent and can be observed non-invasively in living cells and, in the case of transparent organisms, whole animals. Similarly, mCherry is the second generation monomeric red fluorescent protein from *Discosoma* sp. The protein is a 28.8 kDa monomer with 256 amino acids, pI: 6.23. Ex.= 587 nm (540-590 nm); Em.= 610 nm (550-650 nm). As a result, the use of these two-different colours provide an easily measured readout for translation before and after the frameshift sequence (Shaner *et al.*, 2004) and the ratio of the two colours can provide a measure of their relative levels, in turn the result of the frameshift rate.

6.2. Results.

6.2.1. Protein expression of the 1ab-FS HCoV-HKU1.

In order to measure the effect of 1ab-FS HCoV-HKU1 on protein expression, a comparison was made of the translation profiles of four matched constructs; HSV-mCherry, which is HSV-mCherry cloned into the pTriEx vector in the absence of any HKU1 sequence addition (positive control for mCherry expression), HSV-mCherry+1ab-FS+eGFP-His (the wild type HKU1 sequence at the 1 ab junction), HSV-mCherry+1ab-stopFS+eGFP-His (the wild type HKU1 sequence with a stop codon at the start of frameshift slippery sequence) and HSV-mCherry+eGFP-His (a direct fusion of the two reporter sequences with no intervening frameshift). A no plasmid addition control was also included.

An equal volume (2µl) from HSV-mCherry+1ab-FS+eGFP-His (383.7ng/µl) (WT), HSV-mCherry only (388.7ng/µl), HSV-mCherry+eGFP-His (without the frameshift) (297.6ng/µl) or HSV-mCherry+1ab-stopFS+eGFP-His (with the stop codon at the start of the frameshift) (330.4ng/µl) was added to each reaction mixture which includes 40 µl TNT® quick master mix, 1 µl Methionine (1mM) and 8µl nuclease-free water to 50 µl of total volume, and the mixture was incubated at 30°C (PCR thermocycler) for 90 minutes.

The quick coupled T7 TNT system used was unmodified and the resultant proteins were detected by western blot with an anti HSV antibody as detailed in chapter 2 sections 2.19 and 2.20. The western blot analysis showed that a single band of the molecular weight predicted for HSV-mCherry was present in the positive control, in the HSV-mCherry+1ab-FS+eGFP-His (WT) and HSV-mCherry+1ab-stopFS+eGFP-His (stop codon at the start of slippery sequence of

the frameshift (TTT to TAA) which was absent from the negative control lane.

In addition to the mCherry protein, a single band of the molecular weight predicted for the HSV-mCherry+eGFP-His was found in the translation of HSV-mCherry+1ab-FS+eGFP-His (WT) and HSV-mCherry+eGFP-His fusion protein (no frameshift). The intensity of the band for the HSV-mCherry+1ab-FS+eGFP-His (WT) was lower than that observed for the HSV-mCherry+eGFP-His fusion protein construct in keeping with the former being the product of a frameshift event and the latter being the result of direct fusion. Together the western blot analysis provides a clear demonstration of the function of the HKU1 1ab frameshift sequence with outcomes as predicted for the constructs made - **figure 58**. However, as this data was obtained *in vitro*, a second assay which examined activity *in vivo* was also completed.

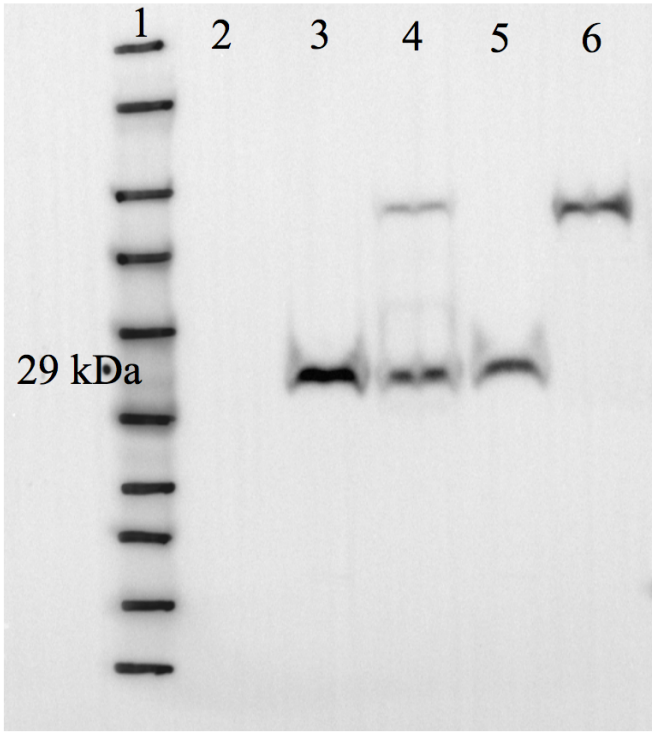

Figure 58. Western blotting analysis of 1ab-FS HCoV-HKU1 expression. Lane 1: See Blue™ Plus2 Pre-Stained Protein Standard (Invitrogen), lane 2: No-plasmid, lane 3: HSV-mCherry only, lane 4: HSV-mCherry+1ab-FS+eGFP-His (WT), lane 5: HSV-mCherry+1ab-stopFS+eGFP-His (stop codon at the start of slippery sequence of the frameshift (change TTT to TAA) and lane 6: HSV-mCherry+eGFP-His (direct fusion).

6.2.2. Function of the HCoV-HKU1 1ab-FS in the Insect Sf9 cell line.

Recombinant baculoviruses were constructed with each of the vectors described using *in vivo* recombination with baculovirus genomic DNA. The transfection mixture was prepared consistent with the manufacturer's protocol (Invitrogen) as described in chapter 2 section (2.25). Recombinant viruses were amplified by successive passage until the observed cytopathic effect was extensive and high titre baculovirus stocks were harvested and stored at 4^{O}C. To examine the *in vivo* fluorescent signal from each recombinant, a 6 well dish containing 5 x 10^6 insect cells was infected at a multiplicity of infection of 3 with each of the recombinant virus stocks by adding 0.5ml of virus stock to each well. After 1 hour at room temperature the wells were supplemented with 2mls of complete insect cell media and incubation continued at 27^{O}C for 3 days.

The infected cells were monitored daily by observation using an Evos FL digital microscope. At 3 days post infection the cells were harvested and the direct fluorescence in each culture measured by flow cytometry. Infected cells were washed once and resuspended in 1 ml of Facsflow buffer (Becton Dickinson) and then analysed by a Becton Dickinson FACScan, recording 10000 events. The data was analysed initially by CellQuest (BD) and the binary files also analysed using WinMDI 2.8 (Scripps Research Institute). The analysed data shows a clear fluorescent signal in both the red and green channels indicating the efficiency of 1ab ORF frameshift in these constructs.

In addition to the Flow cytometry, samples were recorded as visualised images using the fluorescence microscope showing two different fluorescence light emitted colours, red channel for mCherry and green channel for eGFP - **Figure 59**. In these images, symbol R shows red fluorescence expression for the mCherry reporter in the positive control (HSV-mCherry only), RG shows approximately an equal red and green fluorescence expression for the mCherry and eGFP in the HSV-mCherry+eGFP-His construct without the frameshift sequence. Symbol S shows red fluorescence expression for the mCherry only in the HSV-mCherry+1ab-FS+eGFP-His construct which includes a stop codon upstream of the slippery sequence of the frameshift (change TTT to TAA) and the RFG shows variable red and green fluorescence expression for the mCherry and eGFP reporters in the wild type of HSV-mCherry+1ab-FS+eGFP-His.

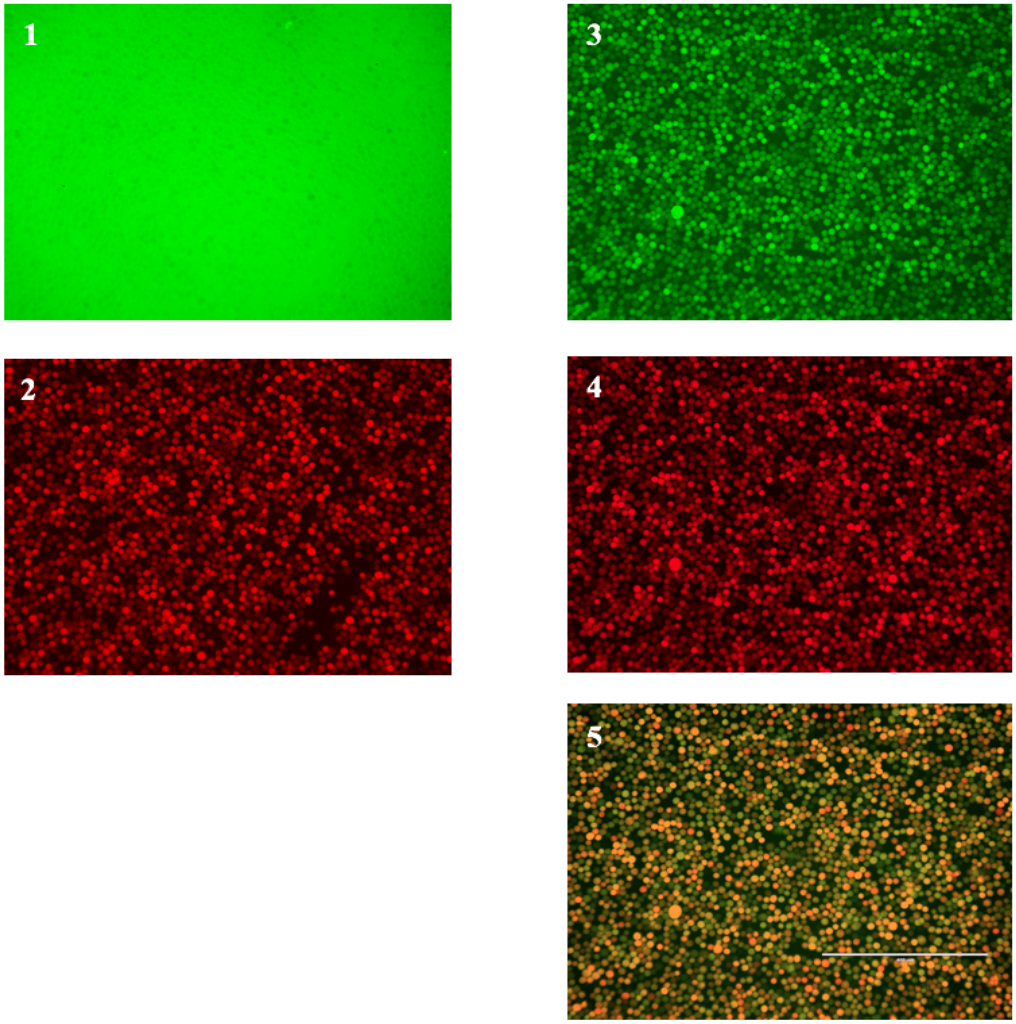

Figure 59. Fluorescent protein expression images of different 1ab-FS HCoV-HKU1 pTriEx 1.1 plasmids in to insect cell line. Image 1 and 2: HSV-mCherry only as positive control (green and red channel respectively), 3, 4 and 5: wild type of HSV-mCherry+1ab-FS+eGFP-His (green, red and orange overlying channels respectively), 6, 7 and 8: HSV-mCherry+1ab-FS+eGFP-His construct including a stop codon upstream the slippery sequence of the frameshift (green, red and orange overlying channels respectively), 9, 10 and 11: HSV-mCherry+eGFP-His construct without the frameshift sequence (green, red and orange overlying channels respectively) and all images captures at 10X magnification power.

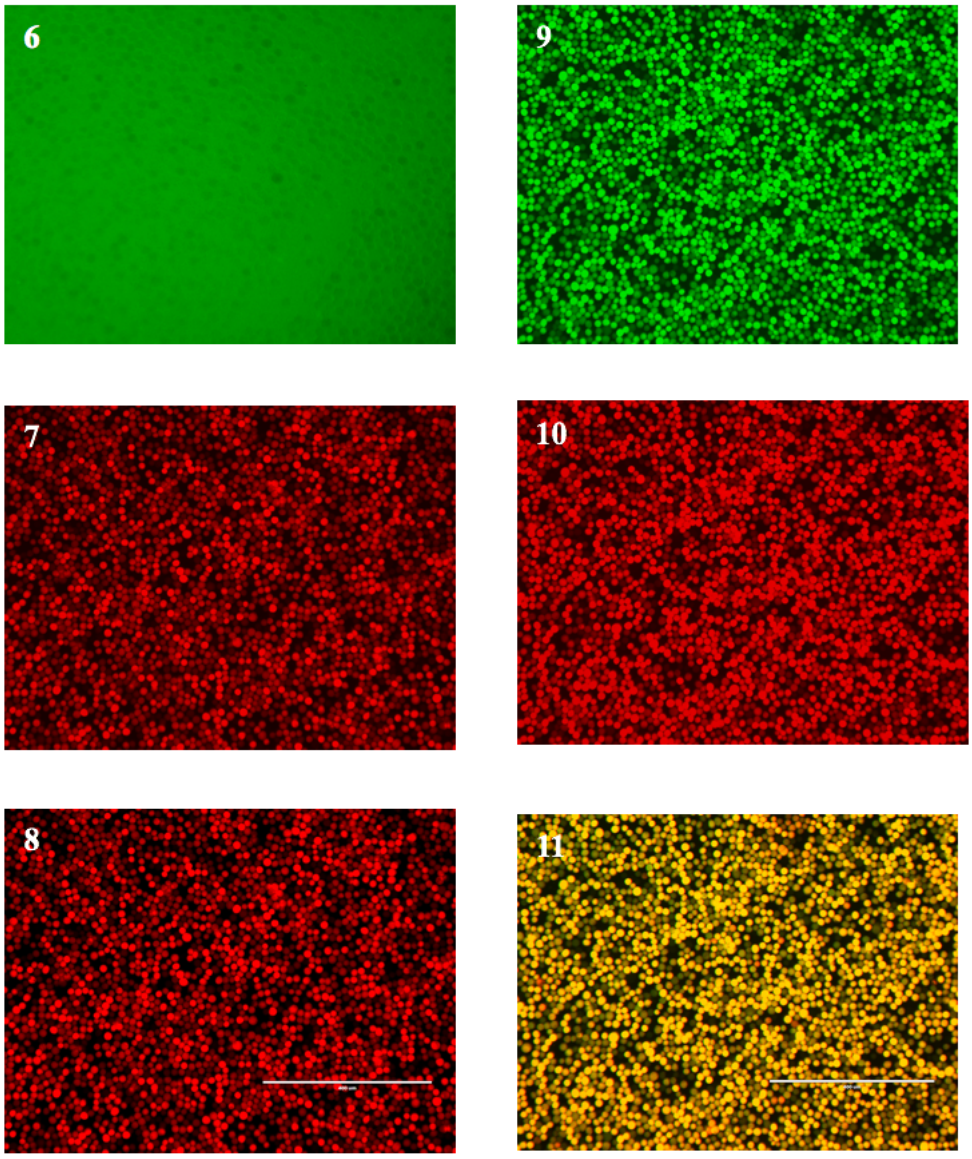

Figure 59 (continued). Fluorescent protein expression images of different 1ab-FS HCoV-HKU1 pTriEx 1.1 plasmids in to insect cell line. Image 1 and 2: HSV-mCherry only as positive control (green and red channel respectively), 3, 4 and 5: wild type of HSV-mCherry+1ab-FS+eGFP-His (green, red and orange overlying channels respectively), 6, 7 and 8: HSV-mCherry+1ab-FS+eGFP-His construct including a stop codon upstream the slippery sequence of the frameshift (green, red and orange overlying channels respectively), 9, 10 and 11: HSV-mCherry+eGFP-His construct without the frameshift sequence (green, red and orange overlying channels respectively) and all images captures at 10X magnification power.

6.2.3. Expression rate of different 1ab-FS HCoV-HKU1 pTriEx 1.1 plasmids in the infected *Sf*9 cells.

Regarding the fluorescence expression levels observed in the previous section and to investigate the 1ab-FS rate, the mean fluorescence level for each group was taken from the statistical data accompanying each FACS file and the values used to plot a graph of relative fluorescence level. The values show 0% and 96% of green and red fluorescence expression respectively for the positive control (R: HSV-mCherry only) and 100% and 98% of the green and red fluorescence expression respectively in the (RG) HSV-mCherry+eGFP-His fusion protein where there is no frameshift sequence. In the case of expression of (S) HSV-mCherry+1ab-StopFS+eGFP-His, 1% and 100% of the green and red fluoresce expression respectively were recorded while for the WT sequence, HSV-mCherry+1ab-FS+eGFP-His, the levels were 17% and 61% of the green and red fluorescence respectively - **Figure 60**. These are consistent with the predictions of the genetic constructs and the biochemical data above using *in vitro* protein translation. They confirm the activity of the 1ab intervening sequence as a *bona fide* frameshift *in vivo* and suggest an efficiency of shifting of 3.6:1 (approximately 4:1) based on the relative levels of mCherry to eGFP. In addition, these data establish an experimental system for the analysis of the role of sequences within the frameshift on its function through site directed changes to the FS sequence followed by re-analysis of the shift rate as described.

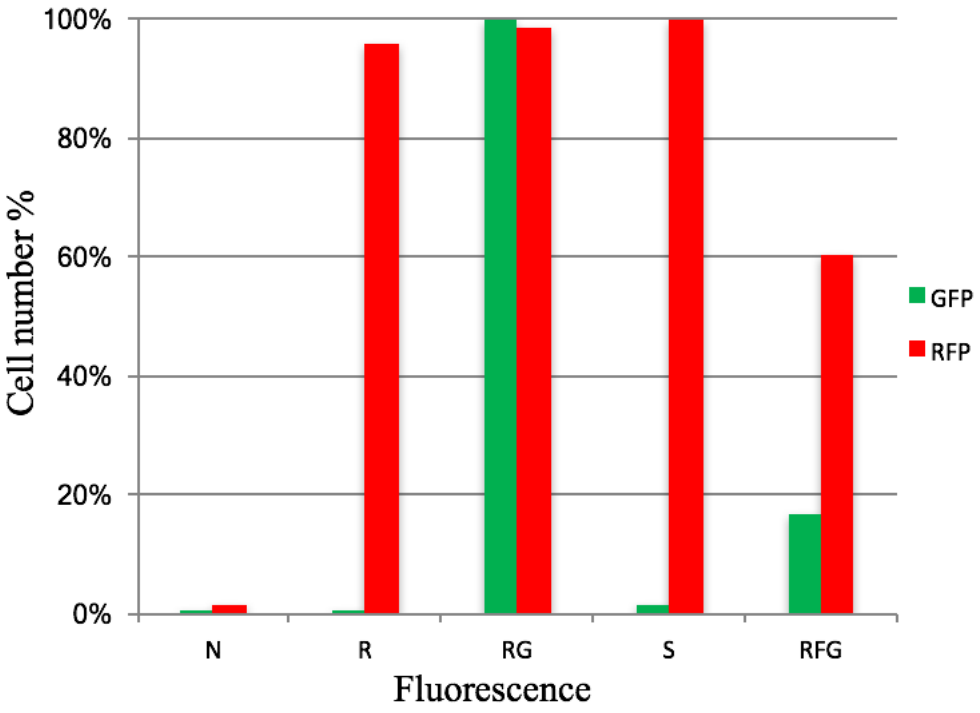

Figure 60. Fluorescent expression rate of different 1ab-FS HCoV-HKU1 pTriEx 1.1 plasmids in to insect cell line. N: negative control (no-plasmid), R: positive control (HSV-mCherry only), RG: HSV-mCherry+eGFP-His construct without the frameshift sequence, S: HSV-mCherry+1ab-FS+eGFP-His construct including a stop codon upstream the slippery sequence of the frameshift (change TTT to TAA) and RFG: wild type of HSV-mCherry+1ab-FS+eGFP-His.

6.2.4. Construction of stems mutation in 1ab-FS HCoV-HKU1 pTriEx 1.1 plasmid.

To probe the structure function relationship of the HKU1 FS a total of 14 oligonucleotide primers were designed for the introduction of 4 mutations into either side of stems 1 and 2 of the frameshift sequence and 2 mutations in the slippery sequence, in which each C or G at the desired mutation side was change to A in order to make a wobble pair. As a result of these changes there is a predicted change in the folding structure – see **Tables 9** and **11** and **Figure 61**, based on the bioinformatics analysis described in chapter 3. Mutations were introduced by using overlapping high-fidelity PCR as described in the previous section (as described in section 2.12) using one fragment amplified with a non-mutant forward primer with a mutant reverse primer and a second corresponding fragment amplified with a mutant forward primer and a non-mutant reverse primer. Following gel extraction and clean-up for each product, the fragments were mixed and an overlapping high-fidelity PCR was used to amplify the final product. As before, gel extraction, clean-up and subcloning into the requisite pTriEx1.1 based plasmid effectively replaced the wild type sequence with the mutant sequence, were completed as described (2.15-17).

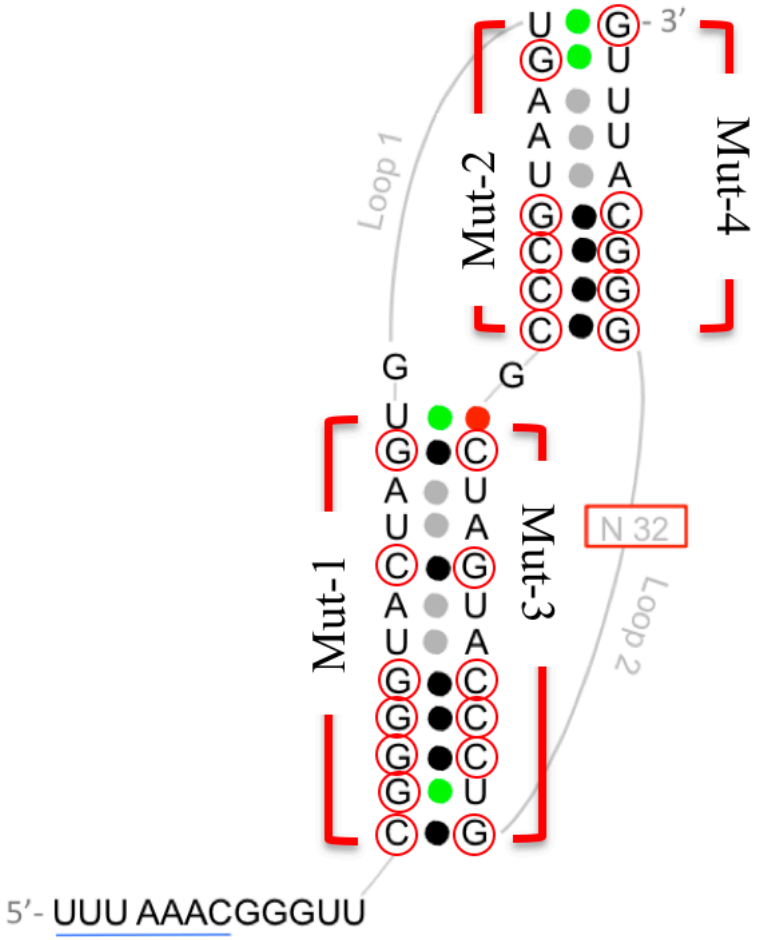

Figure 61. Construction of 4 desired mutations into 1ab-FS HCoV-HKU1 pTriEx 1.1 plasmid. Red coloured bracket includes the location of the 4 mutations 1ab-FS HCoV-HKU1 sequence and red colour circle includes the nucleotide change from G or C to A.

6.2.5. Sequence alignment for Constructed two stems mutations in the HCoV-HKU1 mCherry+1ab-FS+eGFP plasmid.

DNA sequence analysis of each of the derived mutants was performed according to Source BioScience requirements as described in chapter 2 (section 2.18), to ensure the existence of the desired mutation. For example, nucleotide changes from G and C in the target sequence to A will increase the number of wobble pairs in a predicated stem-loop and change the predicted stem or loop folding pattern on one side of the 1ab-FS HCoV-HKU1 structures. The sequences obtained for the desired mutations in the 1ab-FS HCoV-HKU1 sequence (the lower sequence in the figure) were aligned with the base vector 1ab-FS HCoV-HKU1 (the upper sequence) in each case - **Figure 62**. The colours of the trace files shown are standard; red, green, blue and black coloured peaks represent nucleotides T, A, C and G respectively. The data showed the presence of the desired mutations and confirmed the desired two stems mutations in both sides of the 1ab-FS HCoV-HKU1 sequence.

Figure 62. Confirmation of desired mutation into the constructed HCoV-HKU1 HSV-mCherry+1ab-FS+eGFP-His pTriEx 1.1 plasmid. Red coloured square represents nucleotide change from G or C to A compared to upper coloured sequence.

4.2.6. Effect of stems mutation on protein expression of 1ab-FS HCoV-HKU1.

In order to measure the effect of the 4 side stem mutations (in SL1 and SR2) of 1ab-FS HCoV-HKU1 on protein expression, a comparison was made next to the two controls previously described, one the HSV-mCherry+1ab-FS+eGFP-His which is the wild type 1ab FS sequence cloned into the pTriEx vector (positive control) and the other the direct mCherry-eGFP fusion protein. In addition a no plasmid addition was included (negative control). An equal amount of plasmid DNA from Mut-1 (136.5ng/µl), Mut-2 (390.8ng/µl), Mut-3 (363.5ng/µl), Mut-4 (276.8ng/µl), HSV-mCherry+1ab-FS+eGFP-His (direct read through) (186.3ng/µl), HSV-mCherry+1ab-Fs+eGFP-His (WT) (383.7ng/µl) and D.W. (negative control) were added to each reaction mixture which included 40 µl TNT® quick master mix, 1 µl Methionine (1mM) and 8µl nuclease-free water to 50 µl of total volume. The mixture was then incubated at 30°C (PCR thermocycler) for 90 minutes.

The quick coupled T7 TNT system used was unmodified (no labelled methionine) and the resultant proteins were detected by western blot with an anti HSV antibody as detailed in chapter 2 sections (2.19 and 2.20). The western blot analysis showed the characteristic two bands when the TNT mix was programmed by the WT plasmid, representing mCherry with a stop in the frameshift sequence and, at a lower intensity, the read-through product of mCherry-eGFP. By contrast the direct fusion construct HSV-mCherry+1ab-FS+eGFP-His gave a single band of the mCherry-eGFP fusion protein.

However, all four mutations tested in these reactions gave rise to only the mCherry product with no evidence of any frameshift product. The intensity of the band for the 4 mutations was similar suggesting all translated products stopped at the FS sequence and did not undo frameshifting - **Figure 63**.

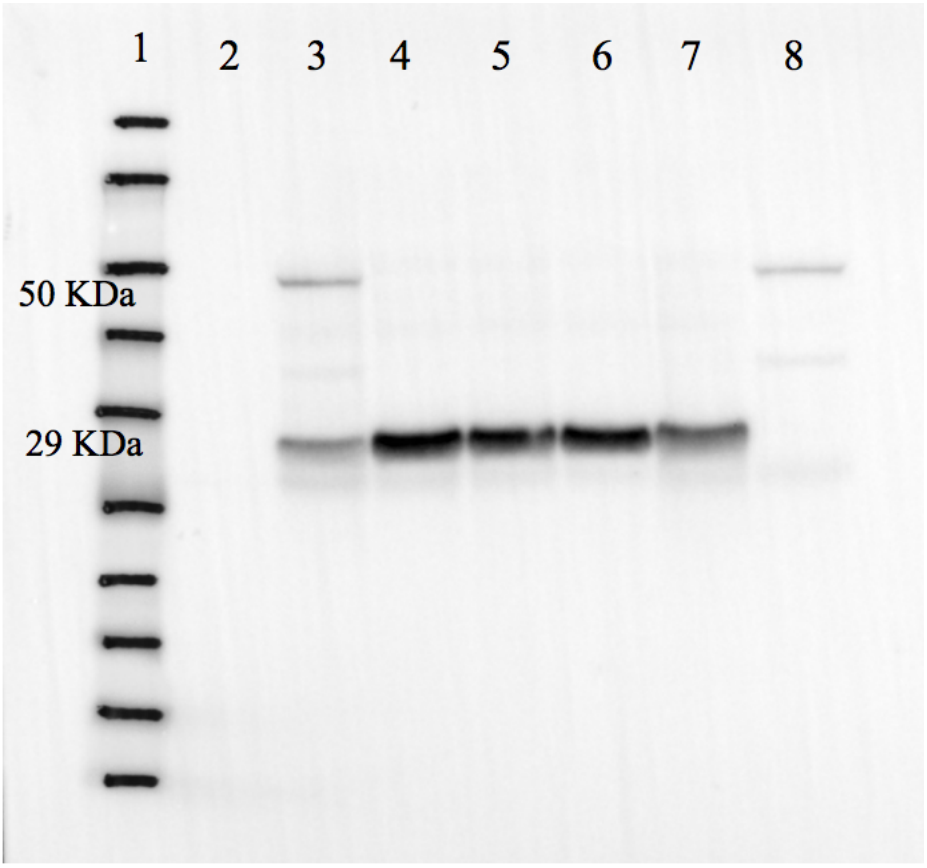

Figure 63. Western blotting analysis of 1ab-FS HCoV-HKU1 expression. Lane 1: See Blue™ Plus2 Pre-Stained Protein Standard pertained (Invitrogen), lane 2: No-plasmid, lane 3: HSV-mCherry+1ab-FS+eGFP-His (WT), lane 4-7: 4 mutations on both sides of stem 1 and 2 respectively, lane 8: HSV-mCherry+1ab-FS+eGFP-His (fusion protein).

6.2.7. Protein expression of the 4 mutation of 1ab-FS HCoV-HKU1 in the Insect cell line.

As before, all four mutants described above were also used to generate recombinant baculoviruses for expression of the reporter proteins in insect cells. This allowed an *in vivo* measure of the mutant frameshift sequence activity. Following the generation of the recombinants the viruses were amplified to high titre and used to infect a single confluent well of *Sf*9 cells (~10^6 cells) in a six well dish. Three days post infection the cells were harvested by centrifugation, resuspended into 1 ml of Facsflow buffer (Becton Dickinson) and then transferred into FACS tube. A total of 10000 events were captured using a FACScan flow cytometer according to the manufacturers protocol (Becton Dickinson) and the data was analysed using WinMDI 2.8 (Scripps Research Institute). All 4 mutants produced an abundant signal in the FL2 (red) channel while both controls produced signal in the FL2 (red) and FL1 (green) channels consistent with read-through of the mCherry protein into a mCherry-eGFP fusion protein.

In addition, the visual images were captured for the two channels of the fluorescence microscope, fluorescence light emitted as red channel for mCherry and green channel for eGFP **Figure 64**. In the R signal a red fluorescence expression shows for the mCherry reporter in the positive control-1 (HSV-mCherry only), RFG shows a variable red and green fluorescence expression for the mCherry and eGFP in the HSV-mCherry+1ab-FS+eGFP-His construct without any mutation (positive control-2). All the constructed mutations, Mut-1 to Mut-4, show a red fluorescence expression for the mCherry only encoded in the HSV-mCherry+1ab-FS+eGFP-His plasmid. All four mutations in both sides of Stem1 and Stem-2 of frameshift sequence effectively abolished green emissions,

however measured. The F shows an approximately equal red and green fluoresce expression for the mCherry and eGFP in the HSV-mCherry+1ab-FS+eGFP-His due to the removal of two (T) from the slippery sequence which leads to permanent read through of the entire sequence.

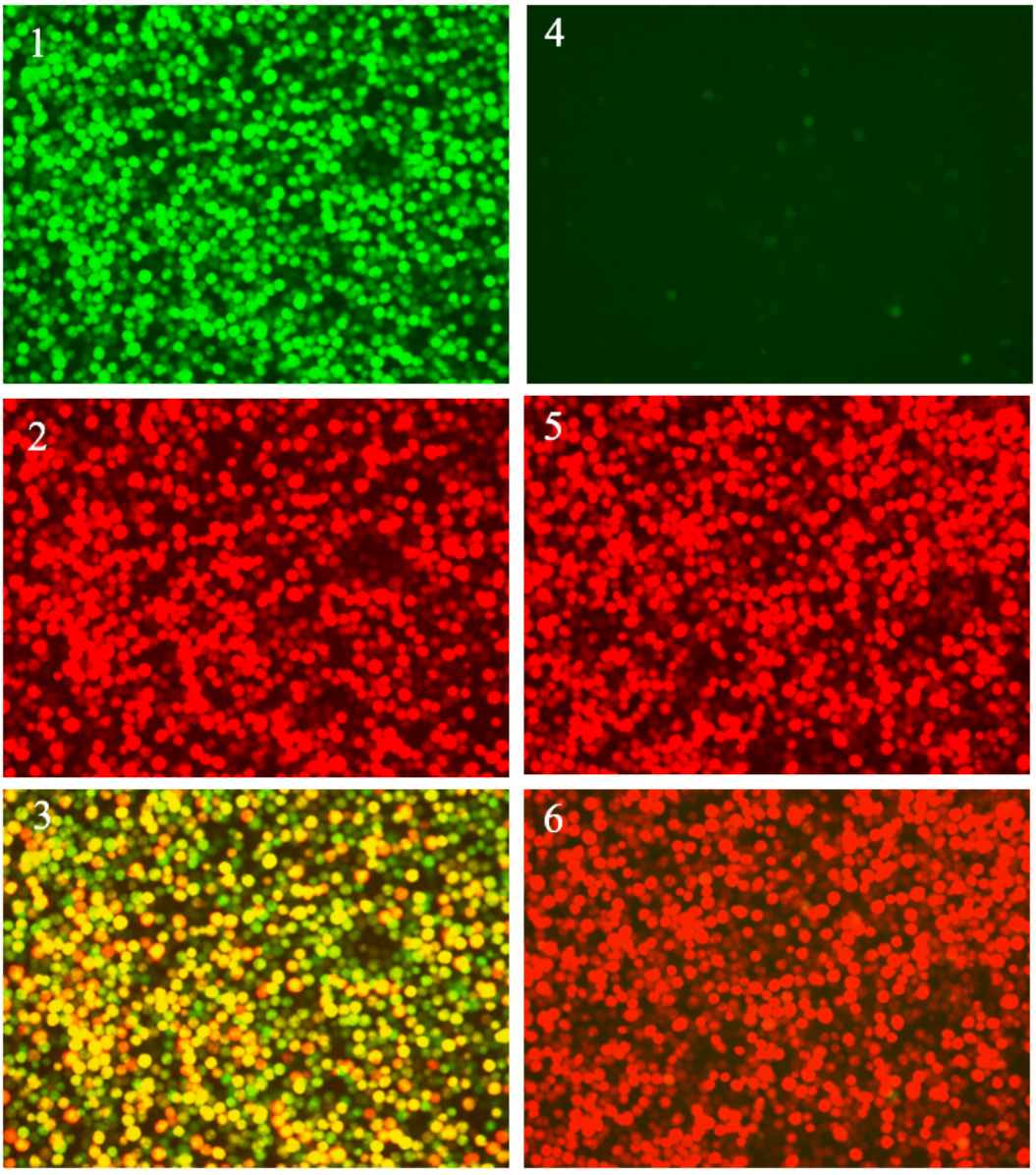

Figure 64. Fluorescent protein expression images of different 1ab-FS HCoV-HKU1 pTriEx 1.1 plasmids in to insect cell line. Image 1, 2 and 3 is RFG: wild type of HSV-mCherry+1ab-FS+eGFP-His, 4, 5 and 6 is Mut-1: left side mutation of stem-1 (green, red and orange overlying channels respectively), 7, 8 and 9 Mut-2: left side mutation of stem-2 (green, red and orange overlying channels respectively), 10, 11 and 12 Mut-3: right side mutation of stem-1 (green, red and orange overlying channels respectively), 13, 14 and 15 Mut-4: right side mutation of stem-2 (green, red and yellow overlying channels respectively), 16, 17 and 18 F: HSV-mCherry+1ab-FS +eGFP-His construct including (-TT) in the slippery sequence for frameshifting and all images captures at 10X magnification power.

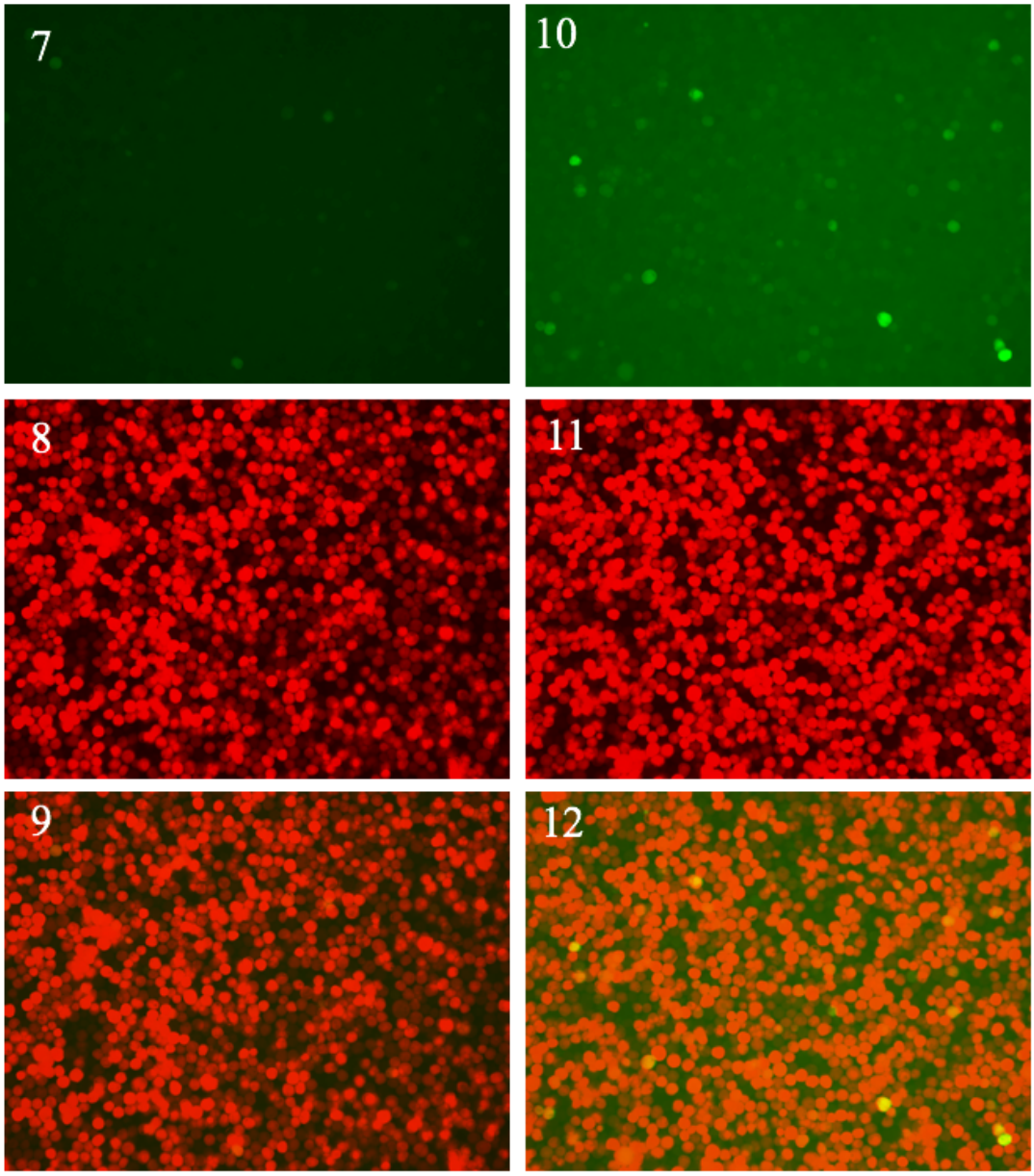

Figure 64 (continued). Fluorescent protein expression images of different 1ab-FS HCoV-HKU1 pTriEx 1.1 plasmids in to insect cell line. Image 1, 2 and 3 is RFG: wild type of HSV-mCherry+1ab-FS+eGFP-His, 4, 5 and 6 is Mut-1: left side mutation of stem-1 (green, red and orange overlying channels respectively), 7, 8 and 9 Mut-2: left side mutation of stem-2 (green, red and orange overlying channels respectively), 10, 11 and 12 Mut-3: right side mutation of stem-1(green, red and orange overlying channels respectively), 13, 14 and 15 Mut-4: right side mutation of stem-2 (green, red and yellow overlying channels respectively), 16, 17 and 18 F: HSV-mCherry+1ab-FS +eGFP-His construct including (-TT) in the slippery sequence for frameshifting and all images captures at 10X magnification power.

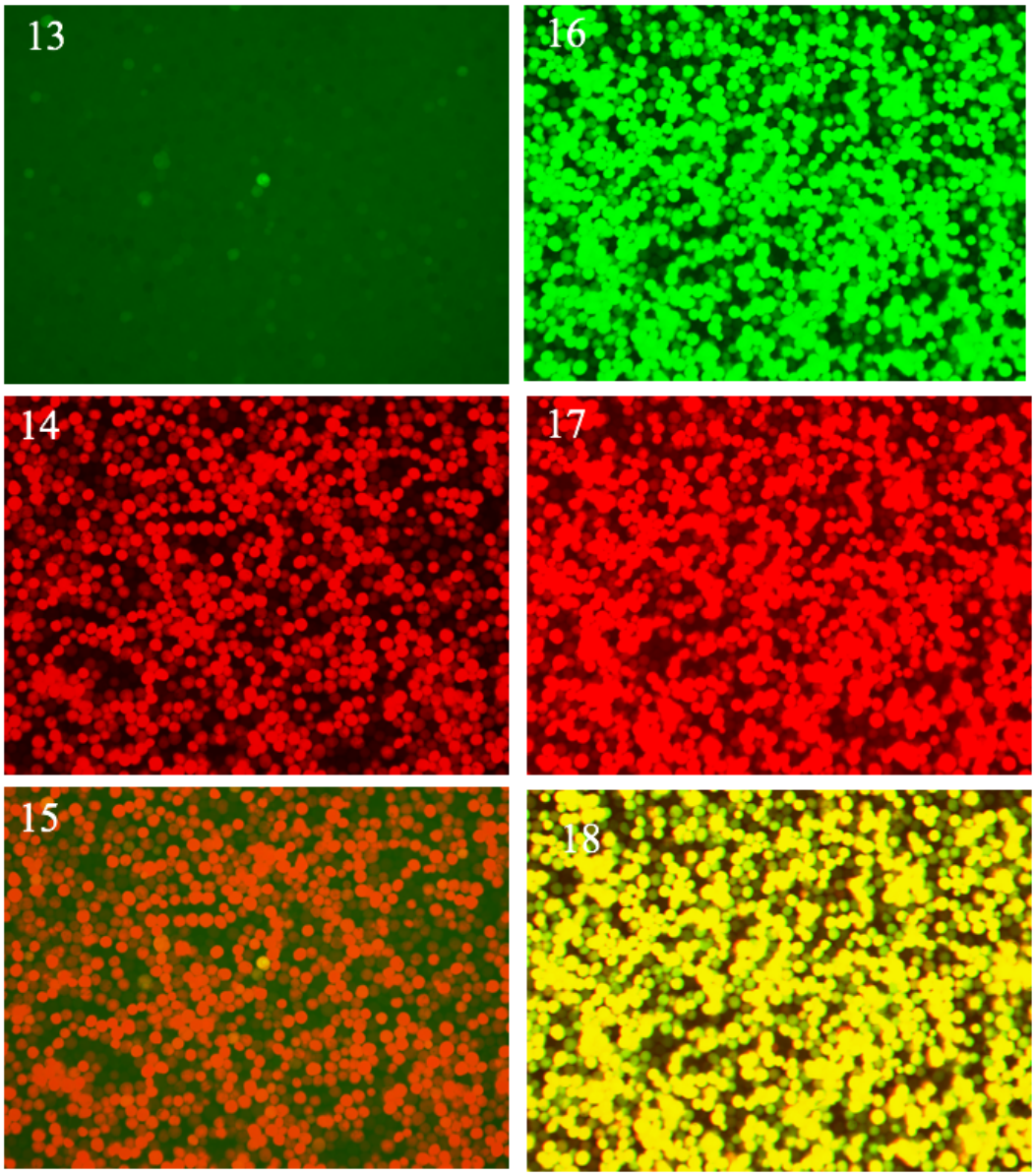

Figure 64 (continued). Fluorescent protein expression images of different 1ab-FS HCoV-HKU1 pTriEx 1.1 plasmids in to insect cell line. Image 1, 2 and 3 is RFG: wild type of HSV-mCherry+1ab-FS+eGFP-His, 4, 5 and 6 is Mut-1: left side mutation of stem-1 (green, red and orange overlying channels respectively), 7, 8 and 9 Mut-2: left side mutation of stem-2 (green, red and orange overlying channels respectively), 10, 11 and 12 Mut-3: right side mutation of stem-1 (green, red and orange overlying channels respectively), 13, 14 and 15 Mut-4: right side mutation of stem-2 (green, red and orange overlying channels respectively), 16, 17 and 18 F: HSV-mCherry+1ab-FS +eGFP-His construct including (-TT) in the slippery sequence for frameshifting and all images captures at 10X magnification power.

6.3. Discussion.

The studies in this chapter were done in order to formally demonstrate the function of, and to investigate the contribution of sequence identity to, the frameshift element of HCoV-HKU1 which is located at the junction of open reading frame 1ab. The translation of HKU1 ORF1 results in a high level of polyprotein 1a (pp1a), and, by virtue of a −1 ribosomal frameshifting event, the translation of ORF1b at a lower level as part of pp1ab (Sawicki *et al.*, 2007, Chen *et al.*, 2009). These two polyproteins are co- and post translationally processed into 16 non-structural proteins (nsps), most of them driving viral genome replication and subgenomic mRNA (sgmRNA) synthesis (Masters, 2006, Firth and Brierley, 2012). A reporter plasmid was constructed with mCherry acting in place of pp1a and eGFP acting in place of pp1b. In addition, control vectors including a vector in which a stop codon was added at the start of the slippery sequence, preventing a frameshift from occurring and a vector in which there was direct fusion of upstream and downstream reporters, were constructed. The successful construction of HSV-mCherry+1ab-FS+eGFP-His in the pTriEx 1.1 plasmid vector was achieved based on infusion cloning as described in chapter 3 sections (3.2.7 and 3.2.7.1 respectively) and the final construct was confirmed by DNA sequencing analysis.

In addition, the expressed proteins were designed to accommodate two different assays; one is an *in vitro* transcription/translation method (Hemmer *et al.*, 1989, Thompson and Pelham, 1979, Goren and Fox, 2008) and the other a flow cytometry assay. In the former case, analysis was by western blot of the translated product whereas in the latter case detection was of fluorescence using two different tagging proteins, green for eGFP and red for mCherry using the FL1-

H and FL2-H channels respectively. Expression *in vivo* was achieved using the recombinant baculovirus system leading to the expression of HSV-mCherry+1ab-FS+eGFP-His protein in infected *Sf*9 cells as described in chapter 3 sections (3.2.7 and 3.2.7.1 respectively) (Kain *et al.*, 1995).

The observations for the western blots and the flow cytometry results revealed different protein expression levels for the two proteins encoded by the HSV-mCherry+1ab-FS+eGFP-His (wild type) construct and this confirmed the efficiency of the 1ab-FS HCoV-HKU1 - **Figure 58**, **63** and **Images 1**, **2** and **3 (Figure 64)**. The relative fluorescence rates allowed a calculation of frequency of frameshift of 1a to 1b which was found to be 3.6 to 1 - **Figure 60**. In HKU1 coronavirus, as a consequence of the frameshift signal at the junction of ORF1a and 1b, the 1b reading frame is translated only about 25% of the frequency of the 1a reading frame from the same positive (+ve) strand genomic template.

Frameshift rates for other viruses range from 1 to 4% for Beet western yellows virus (BWYV) in plant cells to 25 and 50% for mouse mammary tumour virus (MMTV) and avian infectious bronchitis virus (IBV), respectively, in animal cells (Su *et al.*, 2005). As a result of the shift, two polyproteins are produced in different amounts in infected cells without the need to regulate two different ORFs, and similar mechanisms are used in many other viruses, for example retroviruses use a similar mechanism to effect the same differential regulation of protein expression (Giedroc and Cornish, 2009).

Individual frameshift rates have evidently evolved to be optimum for the virus concerned and it has been shown that changing the frameshifting efficiency downward results in significantly lower infectivity for both HIV-1 (Dulude *et al.*, 2006) and in murine Moloney leukemia virus, where the translational reading through signal was exchanged with the frameshifting signal from HIV-1 (Gaudin

et al., 2005). Viruses with identical codon changes affect both the structure of the frameshift signal, frameshifting efficiency and exposed to produce different ratios of genomic and subgenomic RNA (Plant *et al.*, 2013). Concerning the mutational analysis, six mutations were constructed in the 1ab-FS HCoV-HKU1 in which these mutations include; change in the slippery sequence (one change for stopping frameshifting (TTT to TAA) and a read through signal (-TT) as well as four mutations in the frameshift sequence as described in section 6.2.4, and confirmed by the bioinformatics analysis - **Figure 62**.

The results from western blot and the flow cytometry show that changes the first three bases in the slippery sequence TTT to TAA lead to a cessation of frameshifting as was also observed for all four mutations in both sides of the Stem-1 and 2 in the frameshift - **Figure 59** and **60**. In addition, a deletion nutation in the slippery sequence resulted in read through the frameshift signal and resulted in upregulation of downstream protein production **Figure 64**.

These results could be used for antisense antiviral therapy, as well as encouraging many researchers to work for the development of tightly binding molecules that target the stem-loop as a conceivable antiviral strategy (Biswas *et al.*, 2004, Dulude *et al.*, 2008, Park *et al.*, 2008).

7. General discussion.

In late 2003 a novel group 2 coronaviruses was discovered in patients with pneumonia, now known as HCoV HKU1 (Woo *et al.*, 2005b, Wan-Ji Lee *et al.*, 2013). This virus belongs to the Coronaviridae family (Gorbalenya *et al.*, 2006, Hulo *et al.*, 2011) which, for most human coronavirus infections, are commonly characterised as causing only mild upper respiratory infection (part of the range of viruses that cause the common cold). Visually, all family members are characterised by their characteristic spike protein resemblance to the solar corona or crown and they share the features of cytoplasmic replication, budding into the ER and Golgi complex, linear, non-segmented, positive-sense, ssRNA genome and a genome length of approximately 27.6-31kb (Neuman *et al.*, 2006a, Hulo *et al.*, 2011).

In general all coronaviruses genomes are organized into 5' non-structural protein coding regions comprising the replicase genes, which are two-thirds of the genome, and 3' structural and nonessential accessory protein coding regions, Infected cells contain seven to nine virus specific mRNAs with co-terminal 3’ ends, the largest of which is the genomic RNA (Masters, 2006).This 5’ end has a 5′ cap (cap-0) consisting of a guanine nucleotide connected to mRNA via an unusual 5′ to 5′ triphosphate linkage. This guanosine is methylated at the 7 position directly after capping in vivo by a methyl transferase effectively making the virus full length RNA appear like a cellular messenger RNA (Araujo *et al.*, 2012). The 5’-UTR of mouse hepatitis virus (MHV) includes a higher-order structural sequence that functions as a cis-active element that is vital for viral genome transcription and replication (Liu *et al.*, 2007). Additionally , the 5’-UTRs of porcine transmissible gastroenteritis virus (TGEV) and bovine coronavirus (BCoV) have been extensively studied and four stems–loops (I, II, III, and IV)

have been mapped within the 5'-UTR (210-nt) (Raman and Brian, 2005). The predicted SL-I has 11–42 nucleotides, contains just three contiguous Watson–Crick base pairs and has a large 16-nucleotide loop that is not conserved among group 2 coronaviruses.

The predicted SL-II has 51–84 nucleotides and is an A-U base pair-rich hairpin that folds into the transcription regulatory sequence (TRS, the core motif described above) present in the terminal loop (Liu *et al.*, 2007). This TRS is an intragenic sequence or leader-body junction site that is rich in Adenine and Uracil elements (Pasternak *et al.*, 2006). All of the mRNAs carry identical 70-90 nts leader sequences at their 5'ends (Burns *et al.*, 1981, Lai *et al.*, 1983, Lai *et al.*, 1984). All coronavirus TRSs include conserved 6-8 nucleotide core sequences (CS) plus variable 5' and 3' flanking sequences (Sola *et al.*, 2005). This short-repeated sequence of about 7-18 nucleotides is found in the leader TRS region and at many other sites, termed body sites, throughout the genome, all preceding translated sequences. The leader–body junction occurs at numerous sites and includes conserved 39-proximal nucleotides within the main TRS (Smits *et al.*, 2005) and the TRS immediately upstream of the initiating AUG for the ORF that follows (Sawicki *et al.*, 2007). The MHV TRSs are located at the 3' end of the leader (TRS-L), which contains consensus heptameric sequences of two to four 5'-UCUAA-3' repeats like other Betacoronaviruses, with the SARS-CoV TRS having 5'-ACGAAC-3' as the core sequence (Marra *et al.*, 2003, Rota *et al.*, 2003) and contains 5'-AAUCUAAAC-3' (or a closely related sequence depending on the virus strain) as the body (TRS-B) (Sola *et al.*, 2011).

These TRSs at the 5' end of each gene represents signals that regulate the discontinuous transcription of subgenomic mRNAs (sgmRNAs) and include a core sequence (CS; 5'-CUA AAC-3') and the 5' and 3' flanking sequences (5'

TRS and 3'TRS, respectively) that modulate transcription (Zuniga *et al.*, 2004, Dufour *et al.*, 2011).

On other hand, SL-III is phylogenetically conserved among group 2 coronaviruses and seems to have homologues in coronavirus groups 1 and 3 (Raman *et al.*, 2003). The predicted SL-IV of 186–215 nucleotides is also conserved among group 2 coronaviruses (Raman and Brian, 2005). The packaging signal has been characterized in MHV and maps within ORF1b approximately 20kb from the 5' end of the genome, which is sufficient for RNA to be incorporated into virions (Makino *et al.*, 1990, Fosmire *et al.*, 1992, Narayanan *et al.*, 2003, Kuo and Masters, 2013). The BCoV packaging signal exhibits 74% sequence identity to the MHV packaging signal and is located in a similar position (Cologna and Hogue, 2000). The TGEV packaging signal was originally mapped to the first 649 nts at the 5' end of the genome; subsequently this position was further delimited to the first 598 nts (Escors *et al.*, 2003, Morales *et al.*, 2013).

The 3' untranslated region of coronaviruses such as mouse hepatitis virus (MHV), bovine coronavirus (BCoV) and HCoV-HKU1 have a pseudoknot structure in the 3' untranslated region 63-115 nucleotides downstream of the N gene (nucleotide position 29708-29760), which is essential for viral replication (Woo *et al.*, 2005b). A bulge occurs due to an overlap between the last segment of the stem-loop and stem 1 in the pseudoknot and the structure has a conserved counterpart in every group 1 and group 2 coronavirus (Goebel *et al.*, 2004). A pseudoknot is a type of tertiary interaction that includes base pairing between nucleotides in a loop with nucleotides in a single-stranded region outside the loop. It has two double-helical stem regions and two loop regions. The common type of pseudoknot is the H-type which consists of a hairpin stem (stem 1) with a second stem formed by the downstream region which base pairs with the loop of

stem 1 (Giedroc *et al.*, 2000). Pseudoknots usually serve as structural elements helping to stabilize complex 3D structures and perform an active role as vital elements in the regulation of several biological processes such as binding of ribosomal proteins to RNA (Giedroc *et al.*, 2000), initiation of internal ribosome entry translation (Wang *et al.*, 1995), and controlling the translational frame during protein synthesis (Green *et al.*, 2008). In addition, a 3' stem-loop II-like motif is an RNA secondary structure motif was recognized in the 3' untranslated region (3'-UTR) of astrovirus, coronavirus and equine rhinovirus genomes (Jonassen *et al.*, 1998, Robertson *et al.*, 2005). It has an important role in viral replication whereby it might bind to the host's eukaryotic translation initiation factor 1A (eIF- 1A) to take-over the host translational machinery for use by the virus, or to bind other translational regulation proteins having similar folds for similar purposes (Hor *et al.*, 2013). It seems to be conserved at both nucleotide sequence and secondary structure folding levels indicating a strong evolutionary selection for its conservation. The presence of this conserved motif in three different viral families is suggested to be the result of at least two separate recombination events and they have been important for the design of anti-viral therapeutic agents (Hor *et al.*, 2013). Picornavirus RNA genomes, including that of poliovirus (PV), similar to many positive-strand RNA viruses (Picornavirales, Nidovirales, Togaviridae, Caliciviridae, and Astroviridae), have a covalently linked 5′-terminal protein called VPg (viral protein genome linked), a 5′ untranslated region (UTR), a single large open reading frame, a 3′ UTR, and a poly(A) tail of variable length (~20 to 150 adenosine residues) (Ahlquist and Kaesberg, 1979). The infectious salmon anaemia virus (ISAV) mRNAs have heterogeneous 5′-ends, and are polyadenylated from a signal sequence 13–14 nucleotides downstream of the 5′-end terminus of the vRNA which is similar to

the Orthomyxoviridae family which includes the genera Influenzavirus A, Influenzavirus B, Influenzavirus C and Thogotovirus, which all have segmented negative sense single stranded RNA genomes and have partially complementary 3'- and 5'-end noncoding terminal sequences that are conserved within and partially between the genera which are important for the classification of new orthomyxovirus-like virus isolates. The conserved terminal parts of each genomic RNA segment show partial self-complementarity, and form, together with some of the adjacent untranslated nucleotides, double stranded terminal panhandle structures. This structure is important for the transcriptional regulation of the viral RNA (vRNA), since these two structures 3'- and 5'-end terminal sequences are involved in the initiation of transcription (Sandvik *et al.*, 2000).

Programmed ribosomal frameshifting (PRF) allows for translation of two proteins encoded in overlapping reading frames from a single translation initiation site upstream of the 5′ open reading frame. −1 PRF has been documented to occur in bacteria (Tsuchihashi and Kornberg, 1990), yeast (Dinman *et al.*, 1991) and mammals (Clark *et al.*, 2007, Manktelow *et al.*, 2005); The ability of the ribosome to change translational reading frames in the −1 direction (−1 PRF) is used by many positive strand RNA viruses, including economically important plant viruses and many human pathogens, such as retroviruses including HIV-1 and related retroviruses and coronaviruses, including SARS-CoV (Brierley and Dos Ramos, 2006). The efficiency of −1 PRF is not 100%, but instead dictates the molar ratio of downstream and upstream proteins ultimately present in the assembling virus. In HIV-1, this ratio is tightly regulated by the virus; molecules that alter the frameshifting levels have significant effects on virus propagation and infectivity (Biswas *et al.*, 2004).

In addition, it is known that changing the frameshifting efficiency downward results in significantly lower infectivities in both HIV-1 (Dulude and Brakier-Gingras, 2006) and murine Moloney leukemia virus, where the translational readthrough signal was replaced with the frameshifting signal from HIV-1 (Gendron *et al.*, 2005). This, in turn has motivated efforts to develop tightly binding molecules as a potential anti-viral strategy (Biswas *et al.*, 2004, Dulude *et al.*, 2008, Park *et al.*, 2008). Two mRNA-encoded signals are required for stimulation of efficient frameshifting by an elongating ribosome. One is a heptanucleotide “slippery site” of the general sequence X XXY YYZ over which the ribosome pauses, e.g., G GGA AAC in pea enation mosaic virus-1 (PEMV-1) (Nixon *et al.*, 2002) and a downstream RNA pseudoknot, positioned six to eight nucleotides from the 3′ edge of the slip-site (Brierley, 1995). The slip-site alone dramatically increases the intrinsic level of frameshift from 0.00005 to ≈0.005 per codon depending on the sequence (Stahl *et al.*, 2002), with the pseudoknot further stimulating this process 10–30-fold more. Thus, the pseudoknot induces a subtle, yet critical, perturbation in the kinetic partitioning of the translocating ribosome into the −1 frame from the reference frame. A downstream H-type pseudoknot is a stimulatory element. A prominent exception to the pseudoknot as the downstream stimulator is in lentiviruses HIV-1 and SIV, where a bipartite stem–loop structure appears necessary and sufficient to stimulate −1 PRF at the gag-pol junction (Gaudin *et al.*, 2005, Marcheschi *et al.*, 2007, Staple and Butcher, 2005). This may be facilitated by the slippery sequence itself, which is UUUUUUA, a particularly shifty sequence (Brierley *et al.*, 1992).

The tropism of coronaviruses, the particular cells infected, is dictated by the receptors used for cell entry, which varies across the family as a whole. For the human viruses, angiotensin converting enzyme 2 (hACE2) is used by SARS-

CoV (Haga *et al.*, 2008) and NL63, dipeptidyl peptidase 4 (DPP4) is used by the newly described MERS-CoV (Cui *et al.*, 2013, Munster *et al.*, 2016) and aminopeptidase N is used by 229E (Vijgen *et al.*, 2004). By contrast OC43 and HKU1 use acetylated sialic acid containing glycans as their receptor and these are found extensively in the upper respiratory tract (Huang *et al.*, 2015).

Following cell entry it is assumed that all coronaviruses follow a similar replication cycle but the exact detail of many steps has only been published for selected members of the family, mostly those used extensively as a model (e.g. MHV and BCoV) and those causing serious pathology which, as a result, are considered a pandemic threat (e.g. SARS). The role of folded RNA structures is particularly obscure except where detailed above as there have been few dedicated studies of their function despite the fact that they can be predicted in all coronaviruses to date. The conservation suggests functional importance but knowledge of the RNA secondary structures of human coronavirus HKU1 for example, is inadequate notwithstanding there being 30 strains sequenced. As a result, their suggested functions, based on studies of other coronaviruses, need to be confirmed by dedicated study. A successful framework for such studies, focusing on HKU1, was established in **chapter 3** which described the design, construction and confirmation of genetic configurations suitable for later manipulation and read-out of RNA sequence function.

Some studies have been performed with other coronaviruses and there is a more extensive literature on similar structures in other viruses. For example, the 5'-UTR of Mouse Hepatitis Virus (MHV) includes a higher-order structural sequence that functions as a cis-active element that is vital for viral genome transcription and replication (Liu *et al.*, 2007). In MHV and Bovine Coronavirus (BCoV), as in HCoV-HKU1, there is a pseudoknot structure in the 3' untranslated

region 63-115 nucleotides downstream of the 3' end of the N gene (nucleotide position 29708-29760) which has been shown to be essential for viral replication (Woo *et al.*, 2005b). The presence of the 3' stem-loop II-like motif in the 3'-UTR of astrovirus, coronavirus and equine rhinovirus genomes is important in viral replication as it might bind to the host's eukaryotic translation initiation factor 1A (eIF-1A) to favour virus translation during host cell and the existence of this conserved motif in three different viral families has been suggested to be the result of at least two separate recombination events and they are considered important for the design of potential broadly active anti-viral therapeutic agents (Hor *et al.*, 2013).

Picornavirus genomic RNAs, also positive sense single stranded RNA have long 5' nontranslated regions (5'NTRs) which contain an internal ribosomal entry site, e.g. the IRES of Encephalomyocarditis virus is located between nucleotides 260 and 484 of the 5'NTR (Jang *et al*., 1988, Gale *et al.*, 2000). It was observed from the HKU1 sequences available in the database that four, nine and six strains have a unique sequences in the HCoV-HKU1 5'-UTR, 3'-UTR and frameshift respectively (see **Figures 19-21**). Within these strains there were changes in the particular sequences used which could have an effect on base pairing interactions predicted for the regions where these changes occur (Dirks *et al.*, 2004). As these changes could have functional consequences which would ultimately relate to virus fitness, they were noted for test once an assay system had been developed.

The dissimilarity of sequence even among closely related strains arises from the high level of mutations in coronaviruses. The mutations happen originally by incorrect copying by the virus encoded polymerase with more extensive changes occurring via template switching during replication, which can

lead to recombinant genomes in cells infected by different coronavirus strains (Zuniga *et al.*, 2010). The polymerase itself lacks a proofreading 3' to 5' exonuclease domain (Steinhauer *et al.*, 1992), which results in lack of repair of any mismatched bases and, given the extreme size of the genome (*circa*. 30kb), all progeny genomes will contain many mutations, a quasispecies.

Recently is has been recognised that there is, in fact, an error correction mechanism for coronaviruses encoded by the nsp14-ExoN protein, which acts as a 3'-to-5' exoribonuclease to remove mis-incorporated bases (Smith *et al.*, 2013). Through mutation and knock-out studies this enzyme function has been shown to be necessary to avoid error catastrophe in both MHV and SARS (Denison *et al.*, 2011). It is all the more remarkable therefore that despite the extensive dissimilarity at nucleotide level, the predicted secondary structure elements of coronaviruses in general and HKU1 in particular are found to be highly conserved, reflecting their importance for virus survival.

The RNA secondary structure prediction revealed several examples of base pairing which did not follow the Watson-Crick resulting in different predicted structural folding for the unique isolates. These differences included a repositioning of the transcriptional regulatory sequence (UCUAAAC) sequence located in SL-2 of the 5'-UTR which was predicted to be present in the SL-2 loop of all available HCoV-HKU1 except the reference sequence (NC_006577.2) where the entire SL-2 was predicted to be single-stranded **–** see **Figure 22**. Correspondingly, the frameshift sequence was shown in some strains to exhibit variation in a stem structure on loop 2 **Figure 23** (Woo *et al.*, 2005b). Likewise, the sequence of HCoV-HKU1 3'-UTR were predicted to vary in SL-3 and SL-4 in the Genotypes B and C respectively **Figure 24**.

Successful cloning and sequencing was concluded for all overlapping PCR and infusion cloning products resulting in full length HCoV-HKU1 5'-UTR, 3'-UTR and frameshift DNA fragments, all cloned into the pTriEx 1.1 plasmid – see **Figures 25**, **26** and **27** respectively. To these, additional sequences were added upstream and downstream of the HKU1 sequence as reporters to measure the activity of the fragment bearing the secondary structure **Figures 34 A**, **35 A** and **36 A** respectively (described in chapter 3). These recombinant plasmids were used in **chapter 4** to study the effect of the UTRs on translation.

A significant dissimilarity of protein expression was observed from the results of the western blot analysis and the statistical analysis thereof in which the addition of the 5'-UTR to eGFP led to a notable increase in protein expression compared to eGFP alone (Elfakess and Dikstein, 2008) while the addition of the 3'-UTR showed no effect on protein expression. Considering the experimental set-up and the possible explanations of the levels of translation observed, it was concluded that the most likely explanation for the enhanced expression caused by the 5'UTR was the existence of an internal ribosome entry site on the HCoV-HKU1 5'-UTR which increased the initiation rate of translation leading to more translated protein in a given time period. Ribosomes recruited to structured RNA elements, such as IRES elements, give rise to IRES-mediated translation and it has been found that the distance from the 3′end of the IRES to the initiating codon differs depending on the origin of the IRES element (Stoneley and Willis, 2004).

A ribosomal subunit attaches to the 3′ end of the IRES structure and scans in a 5' to 3' direction along the mRNA until it reaches the initiating codon. Once the 40S ribosomal subunit reaches the initiating codon, which is mostly AUG, the 60S ribosomal subunit attaches to form an 80S ribosome that then translates the downstream ORF. The 40S subunit carries the initiator methionine-tRNA and

certain eukaryotic initiation factors (eIFs). As a result of their function, IRES elements have been termed “ribosome landing pads” (RLP) and have been documented in several viral and cellular RNA elements (Le *et al.*, 1995).

In the case of virus infection they are specially used to usurp the translational apparatus of the host during viral infection when host specific translation is faltering as a result of host cell shutdown. Elsewhere they can initiate translation of specific mRNAs during cellular stress or other periods when general global translation is repressed (Thiel and Siddell, 1994, Hellen and Sarnow, 2001, Spahn *et al.*, 2004, Moon *et al.*, 2016).

The HCoV-HKU1 sequence contains a section of 13 nucleotides (AUUUAUUGUUUGG) which is very similar to the core IRES sequence described for RNA 5 of MHV (UUUUAUUCUUUUU) (Woo *et al.*, 2005b). Other than Coronaviridae, many examples of intra-molecular structures have been found in positive strand RNA viruses that are important for many viral processes. Some well characterised examples include the 5’-UTR of hepatitis C virus (HCV) which also contains an internal ribosomal entry site (IRES), required for efficient translation of the viral polyprotein (Masante *et al.*, 2015). IRESes are found widely in the viral world and have been shown previously for Severe Acute Respiratory Syndrome coronavirus (SARS CoV) (Leteane, 2014).

As noted previously, an IRES in the 5’-UTR of HCoV-HKU1 would not be unexpected given their importance in eukaryotic translation initiation in other pathogenic viruses such as hepatitis A virus (HAV) (Glass and Summers, 1992), hepatitis C virus (HCV) (Tsukiyama-Kohara *et al.*, 1992) and foot-and-mouth-disease virus (FMDV) (Kuhn *et al.*, 1990). The complex stem–loop structure of the IRES is capable of engaging the 40S ribosomal subunit in a factor-

independent fashion, which allows cap-independent translation (Kieft, 2008, Ray *et al.*, 2012).

Based on the observed stimulation of translation by the 5'UTR of HKU1 a mutagenic analysis of the predicted fold was undertaken. Analysis of 12 constructed mutations in the 5'-UTR of HCoV-HKU1 indicated a significant dissimilarity of protein expression among them. Mutations mostly clustered in the upstream half of the UTR were found to be stimulatory for translation whereas those in the downstream half were mostly inhibitory. As the introduced mutations were intended to destabilise stem-loop structures this could be interpreted as the ribosome accessing the messenger in this region and being stimulated by local secondary structure melting. Mutation No.10 in the downstream half of the sequence noticeably decreased protein expression when compared to the parental UTR construct, almost to the level of GFP expression alone, which might be consistent with failed IRES binding.

The target sequence for mutation number 10 is located on the right side of stem3 (GTTTAATCATAATCTTGT) to (ATTTAATAATAATATTAT) and it reduced the number of C:G base pairs, explaining its notable effect. The localisation of points of significant downregulation may indicate targets for antisense antiviral therapy (Moon *et al.*, 2016). In contrast to the 5' UTR, addition of the 3'UTR had no effect of the translation of eGFP. This would be consistent with a role in replication rather than translation as has been shown by others (Gale *et al.*, 2000, Yang and Leibowitz, 2015).

Further analysis using EMSA to identify protein complexes with RNA (Hellman and Fried, 2007, Alves and Cuha, 2012) suggested a binding by intact ribosomes to the 5'UTR at lower concentration when compared to the 3'-UTR.

These data were qualitative and preliminary and more detailed analysis is required but they suggest a direct binding of ribosomes to 5'UTR as would be expected of a 1:1 UTR RNA:ribosome complex necessary for stimulation of protein expression. An EMSA shift was observed also for the 3' UTR at higher ribosome concentrations, which is to be expected as translation did occur in these constructs although was not stimulated.

The rest of the constructed plasmids were used to analyse the frameshift element of HCoV-HKU1 which is located at the junction of open reading frame 1ab. High levels of polyprotein 1a (pp1a) result from translation of HKU1 ORF1 while the translation of ORF1b is at a lower level as part of pp1ab (Sawicki *et al.*, 2007, Chen *et al.*, 2009). These two polyproteins are co- and post translationally processed into 16 non-structural proteins (nsps), most of them driving viral genome replication and subgenomic mRNA (sgmRNA) synthesis (Masters, 2006, Firth and Brierley, 2012). The reporter plasmids used for 1ab FS analysis included a mCherry ORF acting in place of pp1a and an eGFP ORF acting in place of pp1b. Various control vectors including a vector in which a stop codon was added at the start of the slippery sequence, preventing a frameshift from occurring and a vector in which there was direct fusion of upstream and downstream reporters, were also designed, constructed and sequence verified.

Analysis of protein expression from these reporters by western blot and by flow cytometry revealed different protein expression levels for the two proteins and the relative fluorescence rates allowed a calculation of the frequency of frameshift of 1a to 1b which was found to be 3.6 to 1 - **Figure 60**. That's would mean that in HKU1 coronavirus, as a consequence of the frameshift signal at the junction of ORF1a and 1b, the 1b reading frame is translated at about 25% of the frequency of the 1a reading frame from the same positive (+ve) strand genomic

template. Documented frameshift rates for other viruses range from ~ 4% for Beet western yellows virus (BWYV) in plant cells to 25 and 50% for mouse mammary tumour virus (MMTV) and avian infectious bronchitis virus (IBV), respectively, in animal cells (van der Hoek *et al.*, 2005). The result of the shift is the formation of two polyproteins at different levels in infected cells without the need to regulate two different ORFs, and similar mechanisms are used in many other viruses, notably retroviruses (Giedroc and Cornish, 2009).

It has been shown that reducing the frameshifting efficiency results in significantly lower infectivity for both HIV-1 (Dulude *et al.*, 2006) and in Murine Moloney Leukemia Virus, where the resident frameshift was exchanged for the frameshifting signal from HIV-1 (Gaudin *et al.*, 2005). Viruses with changed frameshift signals and frameshifting efficiency produced different ratios of genomic and subgenomic RNA (Plant *et al.*, 2013). To probe its structure, six mutations were created in the 1ab-FS of HCoV-HKU1 including; a change in the slippery sequence (one change for stopping frameshifting (TTT to TAA) and a read through signal (-TT) as well as four mutations in the frameshift sequence which bioinformatics analysis suggested would affect the folded structure - **Figure 62**.

It was observed from the results of western blot and flow cytometry that changes the first three bases in the slippery sequence TTT to TAA lead to a cessation of frameshifting, a result that was also the case for all four mutations in both sides of Stems 1 and 2 - **Figure 59** and **60**. In addition, a deletion mutation in the slippery sequence resulted in read through of the frameshift signal and an upregulation of downstream protein production **Figure 64**. As for the 5'UTR data these results have the potential to act as a guide for the design of antisense antiviral therapy, as well as encouraging following researchers to work for the

development of tightly binding molecules that target the stem-loop as a conceivable antiviral strategy (Biswas *et al.*, 2004, Dulude *et al.*, 2008, Park *et al.*, 2008) (described in chapter 6).

As it's known that RNA virus infection involves delivery of the viral genome into cells, transcription of viral mRNA and consequent translation of viral proteins are essential for genome replication, viral assembly and budding. Targeting viral RNA sequences with ASOs or siRNAs is a theoretically appealing strategy for treating RNA virus infections for several reasons. First, preventing the synthesis of even one critical viral protein could potentially disrupt the viral life cycle. Secondly, the sequence specificity of an ASO and siRNA compound potentially allows viral genes to be targeted without affecting host genes, thus decreasing or eliminating unwanted side effects.

In addition, targeting viral sequences does not require an understanding of gene function. The possibility of using ASOs and siRNAs to disrupt RNA virus gene expression was first demonstrated using Rous Sarcoma Virus and Respiratory Syncytial Virus, respectively (Stephenson and Zamecnik, 1978, Bitko and Barik, 2001). This approach was supported by using the antisense peptide nucleic acid (PNA) and locked nucleic acid (LNA) oligomers that can bind key IRES sequences and block translation in the Hepatitis C virus (HCV) (Nulf and Corey, 2004), targeting the TRS region in the 5' UTR (53-72 nt, 56-76 nt) in SARS CoV by using Arg-rich peptide (Neuman *et al.*, 2005).

Similarly, phosphonodiamidite morpholino oligomers (PMOs) have been used to inhibit Flaviviridae infections (West Nile (WNV), Yellow Fever Virus (YFV), Japanese Encephalitis (JEV), and Dengue (DENV)) by targeting conserved sequences within the 5' and 3' untranslated regions of the viral

genome (Spurgers *et al.,* 2008). Finally, the translation start site region of PB1 or NP mRNA or the 3'-terminal region of NP viral RNA has been targeted in the Influenza A virus (Orthomyxoviridae) (Bai *et al.*, 1998).

In summary, this thesis investigated the controlling activity of three secondary RNA structures of HCoV-HKU1 as such knowledge was currently unavailable for this virus. The outcome of the thesis work is summarised in three points:

1- The 5'-UTR has a significant effect on translation probably due to an IRES sequence, whereas the 3'-UTR showed no effect on translation. This data was also supported by preliminary EMSA results.

2- The frameshift signal which is located at the junction of ORF1a and 1b, is functional with a translation efficiency of about 25%.

3- Significant downregulation of UTR and FS activity demonstrated the potential for targets for antisense antiviral therapy targeting each element.

Future work might include RNA mapping to confirm the RNA structures predicted and include structures for the mutations that inhibited the activity of the two sequences by using of RNAse fingerprinting or the Cry-electron microscopy as it knows that RNA folding occurs via a series of transitions between metastable intermediate states. Similarly, further study would include replication assays to investigate the involvement of the 3'-UTR in virus replication. Given the conservation of the folded RNA structures despite variation in the primary sequence it would be interesting to discover if any *trans* acting factors that inhibited 5' UTR or FS function were active against other coronaviruses in addition to HKU1.

In addition, repeating the current assays is physiologically relevant cell types such as human airway epithelial cells would be valuable. In fact, it this type of cell was attempted in the studies described here but the viability was limited and their use was curtailed during the study.

7. Appendix

Table 16. Standard letters and genetic codes for the amino acids.

Second Position

First Position	U	C	A	G	Third Position
U	UUU → Phe (F) UUC → Phe (F) UUA → Leu (L) UUG → Leu (L)	UCU → Ser (S) UCC → Ser (S) UCA → Ser (S) UCG → Ser (S)	UAU → Tyr (Y) UAC → Tyr (Y) UAA → * UAG → *	UGU → Cys (C) UGC → Cys (C) UGA → * UGG → Trp (W)	U C A G
C	CUU → Leu (L) CUC → Leu (L) CUA → Leu (L) CUG → Leu (L)	CCU → Pro (P) CCC → Pro (P) CCA → Pro (P) CCG → Pro (P)	CAU → His (H) CAC → His (H) CAA → Gln (Q) CAG → Gln (Q)	CGU → Arg (R) CGC → Arg (R) CGA → Arg (R) CGG → Arg (R)	U C A G
A	AUU → Ile (I) AUC → Ile (I) AUA → Ile (I) AUG → Met (M)	ACU → Thr (T) ACC → Thr (T) ACA → Thr (T) ACG → Thr (T)	AAU → Asn (N) AAC → Asn (N) AAA → Lys (K) AAG → Lys (K)	AGU → Ser (S) AGC → Ser (S) AGA → Arg (R) AGG → Arg (R)	U C A G
G	GUU → Val (V) GUC → Val (V) GUA → Val (V) GUG → Val (V)	GCU → Ala (A) GCC → Ala (A) GCA → Ala (A) GCG → Ala (A)	GAU → Asp (D) GAC → Asp (D) GAA → Glu (E) GAG → Glu (E)	GGU → Gly (G) GGC → Gly (G) GGA → Gly (G) GGG → Gly (G)	U C A G

Format: mRNA GCA → Ala (A)

Figure 65. An image of the mfold web server. Showing the mfold folding form used for RNA secondary structure prediction of HCoV-HKU1 5'-UTR, 3'-UTR and 1ab-FS.

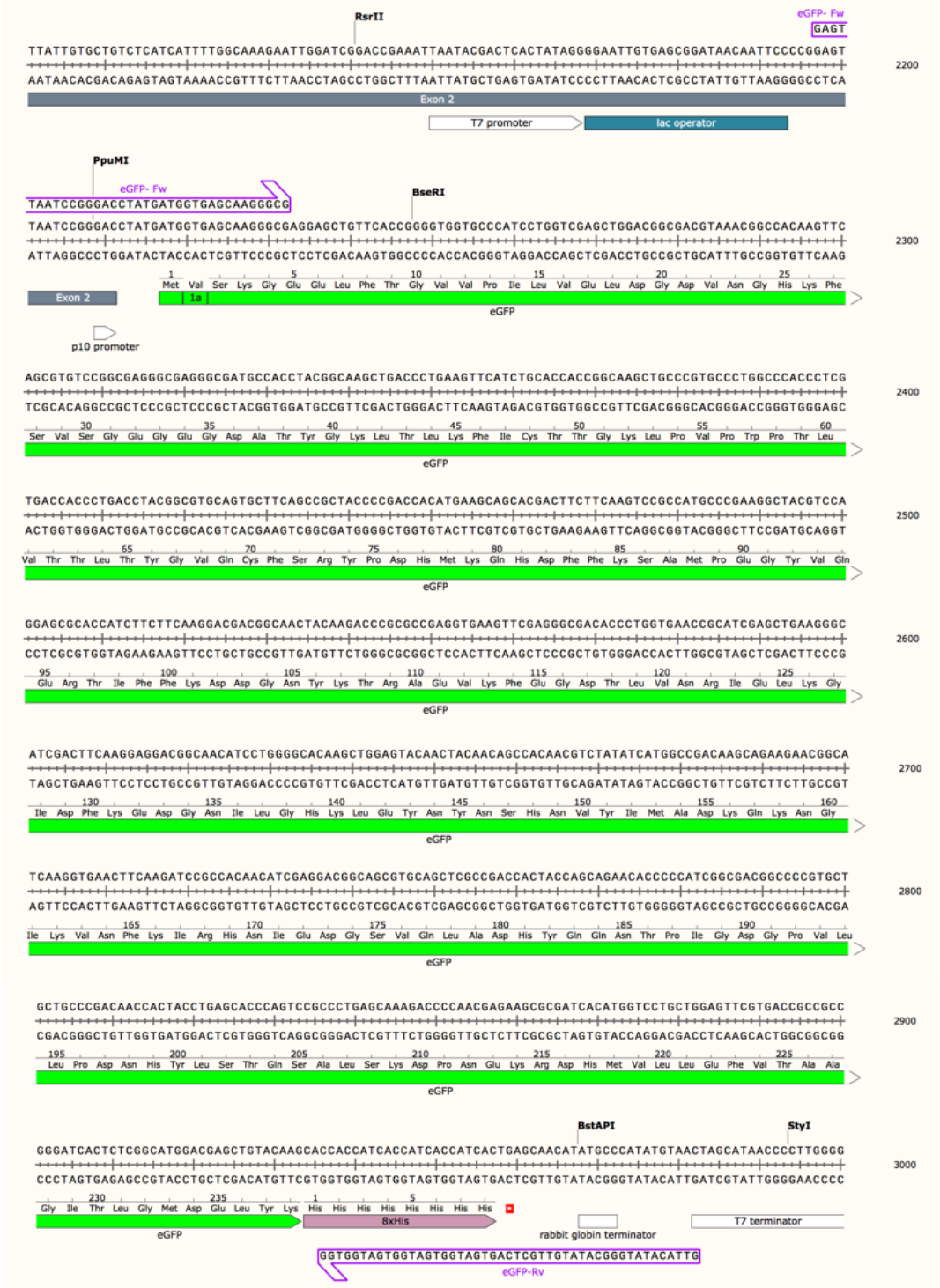

Figure 66. The cloning sequence of eGFP reporter gene. The eGFP contains 8X His tag at the C-terminus, coloured key features of the primers, restriction enzymes and expressed protein are marked.

Figure 67. The cloning sequence of HSVmCherry reporter gene. The mCherry contains HSV tag at the N-terminus, coloured key features of the primers, restriction enzymes and expressed protein are marked.

References

AGRAWAL, S. 1992. Antisense oligonucleotides as antiviral agents. *Trends Biotechnol,* 10**,** 152-8.

AHLQUIST, P. & KAESBERG, P. 1979. Determination of the length distribution of poly(A) at the 3' terminus of the virion RNAs of EMC virus, poliovirus, rhinovirus, RAV-61 and CPMV and of mouse globin mRNA. *Nucleic Acids Res,* 7**,** 1195-204.

ALVES, C. & CUHA, C. 2012. Electrophoretic Mobility Shift Assay: Analyzing Protein - Nucleic Acid Interactions. In: (ED.), S. M. (ed.) Gel Electrophoresis - Advanced Techniques. *InTech.* 12, 206-228.

AMINI, R., JAHANSHIRI, F., AMINI, Y., SEKAWI, Z. & JALILIAN, F. A. 2012. Detection of human coronavirus strain HKU1 in a 2 years old girl with asthma exacerbation caused by acute pharyngitis. *Virol J,* 9**,** 142.

ANAND, K. T., MANOHAR, S. R. K., HARIKRISHNAN, S. & NEELAKANDHAN, K. S. 2003. Right ventricular mass: A histopathological surprise. *Annals of Thoracic Surgery,* 75**,** 1969-1971.

ARAUJO, P. R., YOON, K., KO, D. J., SMITH, A. D., QIAO, M., SURESH, U., BURNS, S. C. & PENALVA, L. O. F. 2012. Before It Gets Started: Regulating Translation at the 5' UTR. *Comp Funct Genomics,* 2012**,** 475731.

BAI, H., YOU, Y., BO, X. & WANG, S. 1998. Antisense antivirals: future oligonucleotides-based therapeutics for viral infectious diseases. *Antivir. Chem. Chemother***,** 9, pp.253-262.

BAILEY, B. L., VISSCHER, K. & WATKINS, J. 2014. A stochastic model of translation with -1 programmed ribosomal frameshifting. *Phys Biol,* 11**,** 016009.

BAIRD, N. J., LUDTKE, S. J., KHANT, H., CHIU, W., PAN, T. & SOSNICK, T. R. 2010. Discrete structure of an RNA folding intermediate revealed by cryo-electron microscopy. *J Am Chem Soc,* 132**,** 16352-3.

BALBAS, P. & BOLIVAR, F. 1990. Design and construction of expression plasmid vectors in Escherichia coli. *Methods Enzymol,* 185**,** 14-37.

BARCENA, M., OOSTERGETEL, G. T., BARTELINK, W., FAAS, F. G., VERKLEIJ, A., ROTTIER, P. J., KOSTER, A. J. & BOSCH, B. J. 2009. Cryo-electron tomography of mouse hepatitis virus: Insights into the structure of the coronavirion. *Proc Natl Acad Sci U S A,* 106**,** 582-7.

BEALE, G. 1993. The discovery of mustard gas mutagenesis by Auerbach and Robson in 1941. *Genetics,* 134**,** 393-9.

BERGMANN, C. C., LANE, T. E. & STOHLMAN, S. A. 2006. Coronavirus infection of the central nervous system: host-virus stand-off. *Nat Rev Microbiol,* 4**,** 121-32.

BERUBE, K., PRYTHERCH, Z., JOB, C. & HUGHES, T. 2010. Human primary bronchial lung cell constructs: the new respiratory models. *Toxicology,* 278**,** 311-8.

BISWAS, P., JIANG, X., PACCHIA, A. L., DOUGHERTY, J. P. & PELTZ, S. W. 2004. The human immunodeficiency virus type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target for antiviral therapy. *J Virol,* 78**,** 2082-7.

BITKO, V. & BARIK, S. 2001. Phenotypic silencing of cytoplasmic genes using sequence-specific double-stranded short interfering RNA and its application in the reverse genetics of wild type negative-strand RNA viruses. *BMC Microbiol,* 1**,** 34.

BLAAS, D. 2016. Viral entry pathways: the example of common cold viruses. *Wien Med Wochenschr,* 166**,** 211-26.

BORKAR, P. S. & MAHAJAN, A. R. 2014. Different RNA Secondary Structure Prediction Methods. 228-230.

BOSIS, S., ESPOSITO, S., NIESTERS, H. G. M., TREMOLATI, E., PAS, S., PRINCIPI, N. & OSTERHAUS, A. D. M. E. 2007. Coronavirus HKU1 in an Italian pre-term infant with bronchiolitis. *Journal of Clinical Virology,* 38**,** 251-253.

BRIAN, D. A. & BARIC, R. S. 2005. Coronavirus genome structure and replication. *Curr Top Microbiol Immunol,* 287**,** 1-30.

BRIERLEY, I. 1995. Ribosomal frameshifting viral RNAs. *J Gen Virol,* 76 (Pt 8)**,** 1885-92.

BRIERLEY, I. & DOS RAMOS, F. J. 2006. Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res,* 119**,** 29-42.

BRIERLEY, I., JENNER, A. J. & INGLIS, S. C. 1992. Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J Mol Biol,* 227**,** 463-79.

BURKS, J. S., DEVALD, B. L., JANKOVSKY, L. D. & GERDES, J. C. 1980. Two coronaviruses isolated from central nervous system tissue of two multiple sclerosis patients. *Science,* 209**,** 933-4.

BURNS, A. L., SPENCE, S., KOSCHE, K., RAMIREZ, F., MEARS, J. G., SCHREINER, H., MILLER, C., BAIRD, M., LEIBOWITZ, D., GIARDINA, P., MARKENSON, A. & BANK, A. 1981. Isolation and characterization of cloned DNA: the delta and beta globin genes in homozygous beta + thalassemia. *Blood,* 57**,** 140-6.

CADWELL, R. C. & JOYCE, G. F. 1992. Randomization of genes by PCR mutagenesis. *PCR Methods Appl,* 2**,** 28-33.

CASAIS, R., THIEL, V., SIDDELL, S. G., CAVANAGH, D. & BRITTON, P. 2001. Reverse genetics system for the avian coronavirus infectious bronchitis virus. *Journal of Virology,* 75**,** 12359-12369.

CHAN JFW, LAU SKP, TO KKW, CHENG VCC, WOO PCY & K-Y., Y. 2015. Middle East Respiratory Syndrome Coronavirus: Another Zoonotic Betacoronavirus Causing SARS-Like Disease. . *Clinical Microbiology* 28(2)**,** 28(2):465-522. doi:10.1128/CMR.00102-14.

CHEN, Y., CAI, H., PAN, J., XIANG, N., TIEN, P., AHOLA, T. & GUO, D. Y. 2009. Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proceedings of the National Academy of Sciences of the United States of America,* 106**,** 3484-3489.

CHENG, Z. F. & DEUTSCHER, M. P. 2005. An important role for RNase R in mRNA decay. *Mol Cell,* 17**,** 313-8.

CLARK, M. B., JANICKE, M., GOTTESBUHREN, U., KLEFFMANN, T., LEGGE, M., POOLE, E. S. & TATE, W. P. 2007. Mammalian gene PEG10 expresses two reading frames by high efficiency -1 frameshifting in embryonic-associated tissues. *J Biol Chem,* 282**,** 37359-69.

COLOGNA, R. & HOGUE, B. G. 2000. Identification of a bovine coronavirus packaging signal. *J Virol,* 74**,** 580-3.

COLUSSI, T. M., COSTANTINO, D. A., ZHU, J., DONOHUE, J. P., KOROSTELEV, A. A., JAAFAR, Z. A., PLANK, T. D., NOLLER, H. F. & KIEFT, J. S. 2015. Initiation of translation in bacteria by a structured eukaryotic IRES RNA. *Nature,* 519**,** 110-3.

CUI, J., EDEN, J. S., HOLMES, E. C. & WANG, L. F. 2013. Adaptive evolution of bat dipeptidyl peptidase 4 (dpp4): implications for the origin and emergence of Middle East respiratory syndrome coronavirus. *Virol J,* 10**,** 304.

DAOU-CHABO, R. & CONDON, C. 2009. *RNase J1 endonuclease activity as a probe of RNA secondary structure,* France, Cold Spring Harbor Laboratory Press, 7, 1417–1425.

DAVID CAVANAGH , T. D. K. B. 1990. Coronaviruses and their diseases. *Advances in Experimental Medicine and Biology,* 276**,** 1.

DAVIES, H. A. & MACNAUGHTON, M. R. 1979. Comparison of the morphology of three coronaviruses. *Arch Virol,* 59**,** 25-33.

DE GUZMAN, R. N., WU, Z. R., STALLING, C. C., PAPPALARDO, L., BORER, P. N. & SUMMERS, M. F. 1998. Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element. *Science,* 279**,** 384-8.

DE JONGE, H. J., FEHRMANN, R. S., DE BONT, E. S., HOFSTRA, R. M., GERBENS, F., KAMPS, W. A., DE VRIES, E. G., VAN DER ZEE, A. G., TE MEERMAN, G. J. & TER ELST, A. 2007. Evidence based selection of housekeeping genes. *PLoS One,* 2**,** e898.

DE SOUZA LUNA L. K, HEISER V, REGAMEY N, PANNING M, DREXLER J.F, MULANGU S, POON L, BAUMGARTE S, HAIJEMA B.J, KAISER L & DROSTEN C 2007. Generic detection of coron- aviruses and differentiation at the prototype strain level by reverse transcription-PCR and nonfluorescent low-density microarray. *Clin Microbiol,* 45 (3)**,** 1049–1052.

DE WIT, J. J. 2000. Detection of infectious bronchitis virus. *Avian Pathol,* 29**,** 71-93.

DENISON, M. R., GRAHAM, R. L., DONALDSON, E. F., ECKERLE, L. D. & BARIC, R. S. 2011. Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol,* 8**,** 270-9.

DINMAN, J. D., ICHO, T. & WICKNER, R. B. 1991. A -1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein. *Proc Natl Acad Sci U S A,* 88**,** 174-8.

DIRKS, R. M., LIN, M., WINFREE, E. & PIERCE, N. A. 2004. Paradigms for computational nucleic acid design. *Nucleic Acids Res,* 32**,** 1392-403.

DOMINY, C. N. & ANDREWS, D. W. 2003. Site-directed mutagenesis by inverse PCR. *Methods Mol Biol,* 235**,** 209-23.

DOS RAMOS, F., CARRASCO, M., DOYLE, T. & BRIERLEY, I. 2004. Programmed -1 ribosomal frameshifting in the SARS coronavirus. *Biochem Soc Trans,* 32**,** 1081-3.

DOWLE & MIRIAM 2014. *High resolution electron microscopy of biological systems (Doctoral dissertation, University of Birmingham).*

DRAPER, D. E. 2004. A guide to ions and RNA structure. *RNA,* 10**,** 335-43.

DUFOUR, D., MATEOS-GOMEZ, P. A., ENJUANES, L., GALLEGO, J. & SOLA, I. 2011. Structure and functional relevance of a transcription-regulating sequence involved in coronavirus discontinuous RNA synthesis. *J Virol,* 85**,** 4963-73.

DULUDE, D., BERCHICHE, Y. A., GENDRON, K., BRAKIER-GINGRAS, L. & HEVEKER, N. 2006. Decreasing the frameshift efficiency translates into an equivalent reduction of the replication of the human immunodeficiency virus type 1. *Virology,* 345**,** 127-36.

DULUDE, D. & BRAKIER-GINGRAS, L. 2006. [The structure of the frameshift stimulatory signal in HIV-1 RNA: a potential target for the treatment of patients infected with HIV]. *Med Sci (Paris),* 22**,** 969-72.

DULUDE, D., THEBERGE-JULIEN, G., BRAKIER-GINGRAS, L. & HEVEKER, N. 2008. Selection of peptides interfering with a ribosomal frameshift in the human immunodeficiency virus type 1. *RNA,* 14**,** 981-91.

EBIHARA, T., ENDO, R., MA, X., ISHIGURO, N. & KIKUTA, H. 2005. Detection of human coronavirus NL63 in young children with bronchiolitis. *J Med Virol,* 75**,** 463-5.

EHRESMANN, C., BAUDIN, F., MOUGEL, M., ROMBY, P., EBEL, J. P. & EHRESMANN, B. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res,* 15**,** 9109-28.

ELFAKESS, R. & DIKSTEIN, R. 2008. A Translation Initiation Element Specific to mRNAs with Very Short 59UTR that Also Regulates Transcription. *PLoS ONE* 3(8): e3094.

ESCORS, D., IZETA, A., CAPISCOL, C. & ENJUANES, L. 2003. Transmissible gastroenteritis coronavirus packaging signal is located at the 5' end of the virus genome. *J Virol,* 77**,** 7890-902.

ESPER, F., WEIBEL, C., FERGUSON, D., LANDRY, M. L. & KAHN, J. S. 2005. Evidence of a novel human coronavirus that is associated with respiratory tract disease in infants and young children. *J Infect Dis,* 191**,** 492-8.

ESPER, F., WEIBEL, C., FERGUSON, D., LANDRY, M. L. & KAHN, J. S. 2006. Coronavirus HKU1 infection in the United States. *Emerg Infect Dis,* 12**,** 775-9.

FAN, H., OOI, A., TAN, Y. W., WANG, S., FANG, S., LIU, D. X. & LESCAR, J. 2005. The nucleocapsid protein of coronavirus infectious bronchitis virus:

crystal structure of its N-terminal domain and multimerization properties. *Structure,* 13**,** 1859-68.

FARABAUGH, P. J. 1996. Programmed translational frameshifting. *Annu Rev Genet,* 30**,** 507-28.

FIRTH, A. E. & BRIERLEY, I. 2012. Non-canonical translation in RNA viruses. *J Gen Virol,* 93**,** 1385-409.

FLINDERS, J. & DIECKMANN, T. 2006. NMR spectroscopy of ribonucleic acids. *Progress in Nuclear Magnetic Resonance Spectroscopy,* 48**,** 137-159.

FOSMIRE, J. A., HWANG, K. & MAKINO, S. 1992. Identification and characterization of a coronavirus packaging signal. *J Virol,* 66**,** 3522-30.

FOUCHIER, R. A., HARTWIG, N. G., BESTEBROER, T. M., NIEMEYER, B., DE JONG, J. C., SIMON, J. H. & OSTERHAUS, A. D. 2004. A previously undescribed coronavirus associated with respiratory disease in humans. *Proc Natl Acad Sci U S A,* 101**,** 6212-6.

FRIEBE, P., BOUDET, J., SIMORRE, J. P. & BARTENSCHLAGER, R. 2005. Kissing-loop interaction in the 3' end of the hepatitis C virus genome essential for RNA replication. *J Virol,* 79**,** 380-92.

FULCHER, M. L., GABRIEL, S., BURNS, K. A., YANKASKAS, J. R. & RANDELL, S. H. 2005. Well-differentiated human airway epithelial cell cultures. *Methods Mol Med,* 107**,** 183-206.

GALE, M., JR., TAN, S. L. & KATZE, M. G. 2000. Translational control of viral gene expression in eukaryotes. *Microbiol Mol Biol Rev,* 64**,** 239-80.

GAUDIN, C., MAZAURIC, M. H., TRAIKIA, M., GUITTET, E., YOSHIZAWA, S. & FOURMY, D. 2005. Structure of the RNA signal essential for translational frameshifting in HIV-1. *J Mol Biol,* 349**,** 1024-35.

GENDRON, K., DULUDE, D., LEMAY, G., FERBEYRE, G. & BRAKIER-GINGRAS, L. 2005. The virion-associated Gag-Pol is decreased in chimeric Moloney murine leukemia viruses in which the readthrough region is replaced by the frameshift region of the human immunodeficiency virus type 1. *Virology,* 334**,** 342-52.

GERNA, G., PERCIVALLE, E., SARASINI, A., CAMPANINI, G., PIRALLA, A., ROVIDA, F., GENINI, E., MARCHI, A. & BALDANTI, F. 2007. Human respiratory coronavirus HKU1 versus other coronavirus infections in Italian hospitalised patients. *Journal of Clinical Virology,* 38**,** 244-250.

GIEDROC, D. P. & CORNISH, P. V. 2009. Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res,* 139**,** 193-208.

GIEDROC, D. P., THEIMER, C. A. & NIXON, P. L. 2000. Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J Mol Biol,* 298**,** 167-85.

GLASS, M. J. & SUMMERS, D. F. 1992. A cis-acting element within the hepatitis A virus 5'-non-coding region required for in vitro translation. *Virus Res,* 26**,** 15-31.

GOEBEL, S. J., HSUE, B., DOMBROWSKI, T. F. & MASTERS, P. S. 2004. Characterization of the RNA components of a putative molecular switch in the 3' untranslated region of the murine coronavirus genome. *J Virol,* 78**,** 669-82.

GOEBEL, S. J., MILLER, T. B., BENNETT, C. J., BERNARD, K. A. & MASTERS, P. S. 2007. A hypervariable region within the 3 ' cis-acting element of the murine coronavirus genome is nonessential for RNA synthesis but affects pathogenesis. *Journal of Virology,* 81**,** 1274-1287.

GOLDSMITH, C. S. & MILLER, S. E. 2009. Modern uses of electron microscopy for detection of viruses. *Clin Microbiol Rev,* 22**,** 552-63.

GORBALENYA, A. E., ENJUANES, L., ZIEBUHR, J. & SNIJDER, E. J. 2006. Nidovirales: evolving the largest RNA virus genome. *Virus Res,* 117**,** 17-37.

GOREN, M. A. & FOX, B. G. 2008. Wheat germ cell-free translation, purification, and assembly of a functional human stearoyl-CoA desaturase complex. *Protein Expr Purif,* 62**,** 171-8.

GREEN, L., KIM, C. H., BUSTAMANTE, C. & TINOCO, I., JR. 2008. Characterization of the mechanical unfolding of RNA pseudoknots. *J Mol Biol,* 375**,** 511-28.

GRIFFITHS-JONES, S., MOXON, S., MARSHALL, M., KHANNA, A., EDDY, S. R. & BATEMAN, A. 2005. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res,* 33**,** D121-4.

GROSSOEHME, N. E., LI, L. C., KEANE, S. C., LIU, P. H., DANN, C. E., LEIBOWITZ, J. L. & GIEDROC, D. P. 2009. Coronavirus N Protein N-Terminal Domain (NTD) Specifically Binds the Transcriptional Regulatory Sequence (TRS) and Melts TRS-cTRS RNA Duplexes. *Journal of Molecular Biology,* 394**,** 544-557.

HAGA, S., YAMAMOTO, N., NAKAI-MURAKAMI, C., OSAWA, Y., TOKUNAGA, K., SATA, T., YAMAMOTO, N., SASAZUKI, T. & ISHIZAKA, Y. 2008. Modulation of TNF-alpha-converting enzyme by the spike protein of SARS-CoV and ACE2 induces TNF-alpha production and facilitates viral entry. *Proc Natl Acad Sci U S A,* 105**,** 7809-14.

HELLEN, C. U. & SARNOW, P. 2001. Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes Dev,* 15**,** 1593-612.

HELLMAN, L. M. & FRIED, M. G. 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc,* 2**,** 1849-61.

HEMMER, O., DEMANGEAT, G., GREIF, C. & FRITSCH, C. 1989. Detection of a proteolytic activity in the micrococcal nuclease used for preparation of messenger-dependent reticulocyte lysates. *FEBS Letters,* 250**,** 357-361.

HENGEN, P. 1995. Purification of His-Tag fusion proteins from Escherichia coli. *Trends Biochem Sci,* 20**,** 285-6.

HINE, R. & MARTIN, E. A. 2015. A dictionary of biology. Seventh edition / ed. Oxford: Oxford University Press.

HOLDEN, N. S. & TACON, C. E. 2011. Principles and problems of the electrophoretic mobility shift assay. *J Pharmacol Toxicol Methods,* 63**,** 7-14.

HOR, C. Y., YANG, C. B., CHANG, C. H., TSENG, C. T. & CHEN, H. H. 2013. A Tool Preference Choice Method for RNA Secondary Structure Prediction by SVM with Statistical Tests. *Evol Bioinform Online,* 9**,** 163-84.

HUANG, X., DONG, W., MILEWSKA, A., GOLDA, A., QI, Y., ZHU, Q. K., MARASCO, W. A., BARIC, R. S., SIMS, A. C., PYRC, K., LI, W. & SUI, J. 2015. Human Coronavirus HKU1 Spike Protein Uses O-Acetylated Sialic Acid as an Attachment Receptor Determinant and Employs Hemagglutinin-Esterase Protein as a Receptor-Destroying Enzyme. *J Virol,* 89**,** 7202-13.

HULO, C., DE CASTRO, E., MASSON, P., BOUGUELERET, L., BAIROCH, A., XENARIOS, I. & LE MERCIER, P. 2011. ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res,* 39**,** D576-82.

ITO, T. & LAI, M. M. 1997. Determination of the secondary structure of and cellular protein binding to the 3'-untranslated region of the hepatitis C virus RNA genome. *J Virol,* 71**,** 8698-706.

IVANOV, K. A., THIEL, V., DOBBE, J. C., VAN DER MEER, Y., SNIJDER, E. J. & ZIEBUHR, J. 2004. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J Virol,* 78**,** 5619-32.

JANG, S. K., KRAUSSLICH, H. G., NICKLIN, M. J., DUKE, G. M., PALMENBERG, A. C. & WIMMER, E. 1988. A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J Virol,* 62**,** 2636-43.

JOHNSON, R. F., FENG, M., LIU, P., MILLERSHIP, J. J., YOUNT, B., BARIC, R. S. & LEIBOWITZ, J. L. 2005. Effect of mutations in the mouse hepatitis virus 3'(+)42 protein binding element on RNA replication. *J Virol,* 79**,** 14570-85.

JONASSEN, C. M., JONASSEN, T. O. & GRINDE, B. 1998. A common RNA motif in the 3' end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. *J Gen Virol,* 79 (Pt 4)**,** 715-8.

KADARE, G. & HAENNI, A. L. 1997. Virus-encoded RNA helicases. *J Virol,* 71**,** 2583-90.

KAIN, S. R., ADAMS, M., KONDEPUDI, A., YANG, T. T., WARD, W. W. & KITTS, P. 1995. Green fluorescent protein as a reporter of gene expression and protein localization. *Biotechniques,* 19**,** 650-5.

KANJANAHALUETHAI, A., CHEN, Z., JUKNELIENE, D. & BAKER, S. C. 2007. Membrane topology of murine coronavirus replicase nonstructural protein 3. *Virology,* 361**,** 391-401.

KERTESZ, M., WAN, Y., MAZOR, E., RINN, J. L., NUTTER, R. C., CHANG, H. Y. & SEGAL, E. 2010. Genome-wide measurement of RNA secondary structure in yeast. *Nature,* 467**,** 103-7.

KIEFT, J. S. 2008. Viral IRES RNA structures and ribosome interactions. *Trends Biochem Sci,* 33**,** 274-83.

KING, A. M., ADAMS, M. J., CARSTENS, E. B. & LEFKOWITZ, E. J. 2012. *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, Elsevier.

KISTLER, A., AVILA, P. C., ROUSKIN, S., WANG, D., WARD, T., YAGI, S., SCHNURR, D., GANEM, D., DERISI, J. L. & BOUSHEY, H. A. 2007. Pan-viral screening of respiratory tract infections in adults with and without asthma reveals unexpected human coronavirus and human rhinovirus diversity. *Journal of Infectious Diseases,* 196**,** 817-825.

KUHN, R., LUZ, N. & BECK, E. 1990. Functional analysis of the internal translation initiation site of foot-and-mouth disease virus. *J Virol,* 64**,** 4625-31.

KUNKEL, T. A. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci U S A,* 82**,** 488-92.

KUO, L. & MASTERS, P. S. 2013. Functional analysis of the murine coronavirus genomic RNA packaging signal. *J Virol,* 87**,** 5182-92.

KUPFER, B., SIMON, A., JONASSEN, C. M., VIAZOV, S., DITT, V., TILLMANN, R. L., MULLER, A., MATZ, B. & SCHILDGEN, O. 2007. Two cases of severe obstructive pneumonia associated with an HKU1-like coronavirus. *European Journal of Medical Research,* 12**,** 134-138.

LAI, M. M. 1990. Coronavirus: organization, replication and expression of genome. *Annu Rev Microbiol,* 44**,** 303-33.

LAI, M. M., BARIC, R. S., BRAYTON, P. R. & STOHLMAN, S. A. 1984. Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proc Natl Acad Sci U S A,* 81**,** 3626-30.

LAI, M. M., PATTON, C. D., BARIC, R. S. & STOHLMAN, S. A. 1983. Presence of leader sequences in the mRNA of mouse hepatitis virus. *J Virol,* 46**,** 1027-33.

LAING, C. & SCHLICK, T. 2011. Computational approaches to RNA structure prediction, analysis, and design. *Curr Opin Struct Biol,* 21**,** 306-18.

LAU, S. K. P., WOO, P. C. Y., YIP, C. C. Y., TSE, H., TSOI, H. W., CHENG, V. C. C., LEE, P., TANG, B. S. F., CHEUNG, C. H. Y., LEE, R. A., SO, L. Y., LAU, Y. L., CHAN, K. H. & YUEN, K. Y. 2006. Coronavirus HKU1 and other coronavirus infections in Hong Kong. *Journal of Clinical Microbiology,* 44**,** 2063-2071.

LE, S. Y., SONENBERG, N. & MAIZEL, J. V., JR. 1995. Unusual folding regions and ribosome landing pad within hepatitis C virus and pestivirus RNAs. *Gene,* 154**,** 137-43.

LETEANE, M. 2014. *Translational control during viral infection, investigating the role of Severe Acute Respiratory Syndrome non-structural protein 1 and Enterovirus 71 Internal Ribosome Entry Site.* PhD, SURREY.

LIN, Y. J., LIAO, C. L. & LAI, M. M. 1994. Identification of the cis-acting signal for minus-strand RNA synthesis of a murine coronavirus: implications for the role of minus-strand RNA in RNA replication and transcription. *J Virol,* 68**,** 8131-40.

LIU, H. & NAISMITH, J. H. 2008. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol,* 8**,** 91.

LIU, P., LI, L., MILLERSHIP, J. J., KANG, H., LEIBOWITZ, J. L. & GIEDROC, D. P. 2007. A U-turn motif-containing stem-loop in the coronavirus 5' untranslated region plays a functional role in replication. *RNA,* 13**,** 763-80.

LIU, Y., WIMMER, E. & PAUL, A. V. 2009. Cis-acting RNA elements in human and animal plus-strand RNA viruses. *Biochim Biophys Acta,* 1789**,** 495-517.

LUKAVSKY, P. J., KIM, I., OTTO, G. A. & PUGLISI, J. D. 2003. Structure of HCV IRES domain II determined by NMR. *Nat Struct Biol,* 10**,** 1033-8.

MAKINO, S., YOKOMORI, K. & LAI, M. M. 1990. Analysis of efficiently packaged defective interfering RNAs of murine coronavirus: localization of a possible RNA-packaging signal. *J Virol,* 64**,** 6045-53.

MANKTELOW, E., SHIGEMOTO, K. & BRIERLEY, I. 2005. Characterization of the frameshift signal of Edr, a mammalian example of programmed -1 ribosomal frameshifting. *Nucleic Acids Res,* 33**,** 1553-63.

MARCHESCHI, R. J., STAPLE, D. W. & BUTCHER, S. E. 2007. Programmed ribosomal frameshifting in SIV is induced by a highly structured RNA stem-loop. *J Mol Biol,* 373**,** 652-63.

MARRA, M. A., JONES, S. J., ASTELL, C. R., HOLT, R. A., BROOKS-WILSON, A., BUTTERFIELD, Y. S., KHATTRA, J., ASANO, J. K., BARBER, S. A., CHAN, S. Y., CLOUTIER, A., COUGHLIN, S. M., FREEMAN, D., GIRN, N., GRIFFITH, O. L., LEACH, S. R., MAYO, M., MCDONALD, H., MONTGOMERY, S. B., PANDOH, P. K., PETRESCU, A. S., ROBERTSON, A. G., SCHEIN, J. E., SIDDIQUI, A., SMAILUS, D. E.,

STOTT, J. M., YANG, G. S., PLUMMER, F., ANDONOV, A., ARTSOB, H., BASTIEN, N., BERNARD, K., BOOTH, T. F., BOWNESS, D., CZUB, M., DREBOT, M., FERNANDO, L., FLICK, R., GARBUTT, M., GRAY, M., GROLLA, A., JONES, S., FELDMANN, H., MEYERS, A., KABANI, A., LI, Y., NORMAND, S., STROHER, U., TIPPLES, G. A., TYLER, S., VOGRIG, R., WARD, D., WATSON, B., BRUNHAM, R. C., KRAJDEN, M., PETRIC, M., SKOWRONSKI, D. M., UPTON, C. & ROPER, R. L. 2003. The Genome sequence of the SARS-associated coronavirus. *Science,* 300**,** 1399-404.

MASANTE, C., JAUBERT, C., PALAU, W., PLISSONNEAU, J., BESNARD, L., VENTURA, M. & DI PRIMO, C. 2015. Mutations of the SL2 dimerization sequence of the hepatitis C genome abrogate viral replication. *Cell Mol Life Sci,* 72**,** 3375-85.

MASTERS, P. S. 2006. The molecular biology of coronaviruses. *Adv Virus Res,* 66**,** 193-292.

MATROSOVICH, M. N., MATROSOVICH, T. Y., GRAY, T., ROBERTS, N. A. & KLENK, H. D. 2004. Human and avian influenza viruses target different cell types in cultures of human airway epithelium. *Proc Natl Acad Sci U S A,* 101**,** 4620-4.

MCINTOSH, K., CHAO, R. K., KRAUSE, H. E., WASIL, R., MOCEGA, H. E. & MUFSON, M. A. 1974. Coronavirus infection in acute lower respiratory tract disease of infants. *J Infect Dis,* 130**,** 502-7.

MEMISH, Z. A., ZUMLA, A. I., AL-HAKEEM, R. F., AL-RABEEAH, A. A. & STEPHENS, G. M. 2013. Family cluster of Middle East respiratory syndrome coronavirus infections. *N Engl J Med,* 368**,** 2487-94.

MOHAMMED & SEGNI 2014. Microwave synthesis characterisation and antimicrobiol studies of noval ligand (2,6-Diacetylpyridine and 5-Amino-1,3,4-Thiadiazole-2-Thiol) and its metal complex. *World journal of pharmaceutical research,* 3**,** 467-480.

MOON, J. S., LEE, S. H., KIM, E. J., CHO, H., LEE, W., KIM, G. W., PARK, H. J., CHO, S. W., LEE, C. & OH, J. W. 2016. Inhibition of Hepatitis C Virus in Mice by a Small Interfering RNA Targeting a Highly Conserved Sequence in Viral IRES Pseudoknot. *PLoS One,* 11**,** e0146710.

MORALES, L., MATEOS-GOMEZ, P. A., CAPISCOL, C., DEL PALACIO, L., ENJUANES, L. & SOLA, I. 2013. Transmissible gastroenteritis coronavirus genome packaging signal is located at the 5' end of the genome and promotes viral RNA incorporation into virions in a replication-independent process. *J Virol,* 87**,** 11579-90.

MUNSTER, V. J., ADNEY, D. R., VAN DOREMALEN, N., BROWN, V. R., MIAZGOWICZ, K. L., MILNE-PRICE, S., BUSHMAKER, T., ROSENKE, R., SCOTT, D., HAWKINSON, A., DE WIT, E., SCHOUNTZ, T. & BOWEN, R. A. 2016. Replication and shedding of MERS-CoV in Jamaican fruit bats (Artibeus jamaicensis). *Sci Rep,* 6**,** 21878.

NARAYANAN, K., CHEN, C. J., MAEDA, J. & MAKINO, S. 2003. Nucleocapsid-independent specific viral RNA packaging via viral envelope protein and viral RNA signal. *J Virol,* 77**,** 2922-7.

NEUMAN, B. W., ADAIR, B. D., YOSHIOKA, C., QUISPE, J. D., MILLIGAN, R. A., YEAGER, M. & BUCHMEIER, M. J. 2006a. Ultrastructure of SARS-CoV, FIPV, and MHV revealed by electron cryomicroscopy. *Adv Exp Med Biol,* 581**,** 181-5.

NEUMAN, B. W., ADAIR, B. D., YOSHIOKA, C., QUISPE, J. D., ORCA, G., KUHN, P., MILLIGAN, R. A., YEAGER, M. & BUCHMEIER, M. J. 2006b. Supramolecular architecture of severe acute respiratory syndrome coronavirus revealed by electron cryomicroscopy. *J Virol,* 80**,** 7918-28.

NEUMAN, B. W., JOSEPH, J. S., SAIKATENDU, K. S., SERRANO, P., CHATTERJEE, A., JOHNSON, M. A., LIAO, L., KLAUS, J. P., YATES, J. R., 3RD, WUTHRICH, K., STEVENS, R. C., BUCHMEIER, M. J. & KUHN, P. 2008. Proteomics analysis unravels the functional repertoire of coronavirus nonstructural protein 3. *J Virol,* 82**,** 5279-94.

NEUMAN, B. W., STEIN, D. A., KROEKER, A. D., CHURCHILL, M. J., KIM, A. M., KUHN, P., DAWSON, P., MOULTON, H. M., BESTWICK, R. K., IVERSEN, P. L. & BUCHMEIER, M. J. 2005. Inhibition, escape, and attenuated growth of severe acute respiratory syndrome coronavirus treated with antisense morpholino oligomers. *J Virol,* 79**,** 9665-76.

NIXON, P. L., RANGAN, A., KIM, Y. G., RICH, A., HOFFMAN, D. W., HENNIG, M. & GIEDROC, D. P. 2002. Solution structure of a luteoviral P1-P2 frameshifting mRNA pseudoknot. *J Mol Biol,* 322**,** 621-33.

NULF, C. J. & COREY, D. 2004. Intracellular inhibition of hepatitis C virus (HCV) internal ribosomal entry site (IRES)-dependent translation by peptide nucleic acids (PNAs) and locked nucleic acids (LNAs). *Nucleic Acids Res,* 32**,** 3792-8.

OSTROWSKI, L. E., STEWART, D. & HAZUCHA, M. 2012. Interferon gamma stimulates accumulation of gas phase nitric oxide in differentiated cultures of normal and cystic fibrosis airway epithelial cells. *Lung,* 190**,** 563-71.

PARK, S. J., JUNG, Y. H., KIM, Y. G. & PARK, H. J. 2008. Identification of novel ligands for the RNA pseudoknot that regulate -1 ribosomal frameshifting. *Bioorg Med Chem,* 16**,** 4676-84.

PASTERNAK, A. O., SPAAN, W. J. M. & SNIJDER, E. J. 2006. Nidovirus transcription: how to make sense ... ? *Journal of General Virology,* 87**,** 1403-1421.

PEIRIS, J. S., YUEN, K. Y., OSTERHAUS, A. D. & STOHR, K. 2003. The severe acute respiratory syndrome. *N Engl J Med,* 349**,** 2431-41.

PICKLES, R. J., MCCARTY, D., MATSUI, H., HART, P. J., RANDELL, S. H. & BOUCHER, R. C. 1998. Limited entry of adenovirus vectors into well-differentiated airway epithelium is responsible for inefficient gene transfer. *J Virol,* 72**,** 6014-23.

PLANT, E. P., PEREZ-ALVARADO, G. C., JACOBS, J. L., MUKHOPADHYAY, B., HENNIG, M. & DINMAN, J. D. 2005. A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. *PLoS Biol,* 3**,** e172.

PLANT, E. P., SIMS, A. C., BARIC, R. S., DINMAN, J. D. & TAYLOR, D. R. 2013. Altering SARS coronavirus frameshift efficiency affects genomic and subgenomic RNA production. *Viruses,* 5**,** 279-94.

POMERANTZ, S. C., KOWALAK, J. A. & MCCLOSKEY, J. A. 1993. Determination of oligonucleotide composition from mass spectrometrically measured molecular weight. *J Am Soc Mass Spectrom,* 4**,** 204-9.

PRENDERGAT, F. G. & MANN, K. G. 1978. Chemical and physical properties of aequorin and the green fluorescent protein isolated from Aequorea forskalea. *Biochemistry***,** 17

PRYTHERCH, Z., JOB, C., MARSHALL, H., OREFFO, V., FOSTER, M. & BERUBE, K. 2011. Tissue-Specific stem cell differentiation in an in vitro airway model. *Macromol Biosci,* 11**,** 1467-77.

PYRC, K., SIMS, A. C., DIJKMAN, R., JEBBINK, M., LONG, C., DEMING, D., DONALDSON, E., VABRET, A., BARIC, R., VAN DER HOEK, L. & PICKLES, R. 2010. Culturing the unculturable: human coronavirus HKU1 infects, replicates, and produces progeny virions in human ciliated airway epithelial cell cultures. *J Virol,* 84**,** 11255-63.

RAJKOWITSCH, L., CHEN, D., STAMPFL, S., SEMRAD, K., WALDSICH, C., MAYER, O., JANTSCH, M. F., KONRAT, R., BLASI, U. & SCHROEDER, R. 2007. RNA chaperones, RNA annealers and RNA helicases. *RNA Biol,* 4**,** 118-30.

RAJKOWITSCH, L., SEMRAD, K., MAYER, O. & SCHROEDER, R. 2005. Assays for the RNA chaperone activity of proteins. *Biochem Soc Trans,* 33**,** 450-6.

RAMAN, S., BOUMA, P., WILLIAMS, G. D. & BRIAN, D. A. 2003. Stem-loop III in the 5 ' untranslated region is a cis-acting element in bovine coronavirus defective interfering RNA replication. *Journal of Virology,* 77**,** 6720-6730.

RAMAN, S. & BRIAN, D. A. 2005. Stem-loop IV in the 5' untranslated region is a cis-acting element in bovine coronavirus defective interfering RNA replication. *J Virol,* 79**,** 12434-46.

RAY, U., RAY, P. S. & DAS, S. 2012. Ribosome–RNA interaction: a potential target for developing antiviral against hepatitis C virus. *Current science (Bangalore)***,** 102, 405-412.

RESTA, S., LUBY, J. P., ROSENFELD, C. R. & SIEGEL, J. D. 1985. Isolation and propagation of a human enteric coronavirus. *Science,* 229**,** 978-81.

REUTER, J. S. & MATHEWS, D. H. 2010. RNAstructure: software for RNA secondary structure prediction and analysis. *Bmc Bioinformatics,* 11.

RISCO, C., ANTON, I. M., ENJUANES, L. & CARRASCOSA, J. L. 1996. The transmissible gastroenteritis coronavirus contains a spherical core shell consisting of M and N proteins. *J Virol,* 70**,** 4773-7.

ROBERTSON, M. P., IGEL, H., BAERTSCH, R., HAUSSLER, D., ARES, M., JR. & SCOTT, W. G. 2005. The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol,* 3**,** e5.

ROTA, P. A., OBERSTE, M. S., MONROE, S. S., NIX, W. A., CAMPAGNOLI, R., ICENOGLE, J. P., PENARANDA, S., BANKAMP, B., MAHER, K., CHEN, M. H., TONG, S., TAMIN, A., LOWE, L., FRACE, M., DERISI, J. L., CHEN, Q., WANG, D., ERDMAN, D. D., PERET, T. C., BURNS, C., KSIAZEK, T. G., ROLLIN, P. E., SANCHEZ, A., LIFFICK, S., HOLLOWAY, B., LIMOR, J., MCCAUSTLAND, K., OLSEN-RASMUSSEN, M., FOUCHIER, R., GUNTHER, S., OSTERHAUS, A. D., DROSTEN, C., PALLANSCH, M. A., ANDERSON, L. J. & BELLINI, W. J. 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science,* 300**,** 1394-9.

RYDER, S. P., RECHT, M. I. & WILLIAMSON, J. R. 2008. Quantitative analysis of protein-RNA interactions by gel mobility shift. *Methods Mol Biol,* 488**,** 99-115.

SANDVIK, T., RIMSTAD, E. & MJAALAND, S. 2000. The viral RNA 3'- and 5'-end structure and mRNA transcription of infectious salmon anaemia virus resemble those of influenza viruses. *Arch Virol,* 145**,** 1659-69.

SAWICKI, S. G., SAWICKI, D. L. & SIDDELL, S. G. 2007. A contemporary view of coronavirus transcription. *J Virol,* 81**,** 20-9.

SEMRAD, K. 2011. Proteins with RNA chaperone activity: a world of diverse proteins with a common task-impediment of RNA misfolding. *Biochem Res Int,* 2011**,** 532908.

SHANER, N. C., CAMPBELL, R. E., STEINBACH, P. A., GIEPMANS, B. N., PALMER, A. E. & TSIEN, R. Y. 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol,* 22**,** 1567-72.

SHARMIN, R. & ISLAM, A. B. 2014. A highly conserved WDYPKCDRA epitope in the RNA directed RNA polymerase of human coronaviruses can be used as epitope-based universal vaccine design. *BMC Bioinformatics,* 15**,** 161.

SHEHU-XHILAGA, M., HILL, M., MARSHALL, J. A., KAPPES, J., CROWE, S. M. & MAK, J. 2002. The conformation of the mature dimeric human immunodeficiency virus type 1 RNA genome requires packaging of pol protein. *J Virol,* 76**,** 4331-40.

SLOOTS, T. P., MCERLEAN, P., SPEICHER, D. J., ARDEN, K. E., NISSEN, M. D. & MACKAY, I. M. 2006. Evidence of human coronavirus HKU1 and human bocavirus in Australian children. *Journal of Clinical Virology,* 35**,** 99-102.

SMITH, E. C., BLANC, H., SURDEL, M. C., VIGNUZZI, M. & DENISON, M. R. 2013. Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. *PLoS Pathog,* 9**,** e1003565.

SMITS, S. L., VAN VLIET, A. L., SEGEREN, K., EL AZZOUZI, H., VAN ESSEN, M. & DE GROOT, R. J. 2005. Torovirus non-discontinuous transcription:

mutational analysis of a subgenomic mRNA promoter. *J Virol,* 79**,** 8275-81.

SOLA, I., MATEOS-GOMEZ, P. A., ALMAZAN, F., ZUNIGA, S. & ENJUANES, L. 2011. RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA Biol,* 8**,** 237-48.

SOLA, I., MORENO, J. L., ZUNIGA, S., ALONSO, S. & ENJUANES, L. 2005. Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis. *J Virol,* 79**,** 2506-16.

SOUSA, A. A. & LEAPMAN, R. D. 2012. Development and application of STEM for the biological sciences. *Ultramicroscopy,* 123**,** 38-49.

SPAHN, C. M., JAN, E., MULDER, A., GRASSUCCI, R. A., SARNOW, P. & FRANK, J. 2004. Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor. *Cell,* 118**,** 465-75.

SPURGERS, K. B., SHARKEY, C. M., WARFIELD, K. L. & BAVARI, S. 2008. Oligonucleotide antiviral therapeutics: antisense and RNA interference for highly pathogenic RNA viruses. *Antiviral Res,* 78**,** 26-36.

STAHL, G., MCCARTY, G. P. & FARABAUGH, P. J. 2002. Ribosome structure: revisiting the connection between translational accuracy and unconventional decoding. *Trends Biochem Sci,* 27**,** 178-83.

STAPLE, D. W. & BUTCHER, S. E. 2005. Solution structure and thermodynamic investigation of the HIV-1 frameshift inducing element. *J Mol Biol,* 349**,** 1011-23.

STEAD, J. A. & MCDOWALL, K. J. 2007. Two-dimensional gel electrophoresis for identifying proteins that bind DNA or RNA. *Nat Protoc,* 2**,** 1839-48.

STEINHAUER, D. A., DOMINGO, E. & HOLLAND, J. J. 1992. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene,* 122**,** 281-8.

STEPHENSON, M. L. & ZAMECNIK, P. C. 1978. Inhibition of Rous sarcoma viral RNA translation by a specific oligodeoxyribonucleotide. *Proc Natl Acad Sci U S A,* 75**,** 285-8.

STONELEY, M. & WILLIS, A. E. 2004. Cellular internal ribosome entry segments: structures, trans-acting factors and regulation of gene expression. *Oncogene,* 23**,** 3200-7.

STUDIER, F. W., ROSENBERG, A. H., DUNN, J. J. & DUBENDORFF, J. W. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol,* 185**,** 60-89.

SU, M. C., CHANG, C. T., CHU, C. H., TSAI, C. H. & CHANG, K. Y. 2005. An atypical RNA pseudoknot stimulator and an upstream attenuation signal for -1 ribosomal frameshifting of SARS coronavirus. *Nucleic Acids Res,* 33**,** 4265-75.

TAYLOR, S. C., BERKELMAN, T., YADAV, G. & HAMMOND, M. 2013. A defined methodology for reliable quantification of Western blot data. *Mol Biotechnol,* 55**,** 217-26.

THELLIN, O., ZORZI, W., LAKAYE, B., BORMAN, B. D., COUMANS, B., HENNEN, G., GRISAR, T., IGOUT, A. & HEINEN, E. 1999. Housekeeping genes as internal standards: use and limits. *Biotechnology,* 75**,** 291–295.

THIEL, V. & SIDDELL, S. G. 1994. Internal ribosome entry in the coding region of murine hepatitis virus mRNA 5. *J Gen Virol,* 75 (Pt 11)**,** 3041-6.

THOMPSON, M. & PELHAM, A. 1979. *Energy policy : timely reports to keep journalists, scholars, and the public abreast of developing issues, events, and trends,* Washington, Congressional Quarterly, 131, 3A-119A

TSUCHIHASHI, Z. & KORNBERG, A. 1990. Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proc Natl Acad Sci U S A,* 87**,** 2516-20.

TSUKIYAMA-KOHARA, K., IIZUKA, N., KOHARA, M. & NOMOTO, A. 1992. Internal ribosome entry site within hepatitis C virus RNA. *J Virol,* 66**,** 1476-83.

TZAKOS, A. G., EASTON, L. E. & LUKAVSKY, P. J. 2007. Preparation of large RNA oligonucleotides with complementary isotope-labeled segments for NMR structural studies. *Nat Protoc,* 2**,** 2139-47.

VABRET, A., DINA, J., GOUARIN, S., PETITJEAN, J., CORBET, S. & FREYMUTH, F. 2006. Detection of the new human coronavirus HKU1: A report of 6 cases. *Clinical Infectious Diseases,* 42**,** 634-639.

VABRET, A., MOUREZ, T., GOUARIN, S., PETITJEAN, J. & FREYMUTH, F. 2003. An outbreak of coronavirus OC43 respiratory infection in Normandy, France. *Clin Infect Dis,* 36**,** 985-9.

VABRET, A., MOUTHON, F., MOUREZ, T., GOUARIN, S., PETITJEAN, J. & FREYMUTH, F. 2001. Direct diagnosis of human respiratory coronaviruses 229E and OC43 by the polymerase chain reaction. *J Virol Methods,* 97**,** 59-66.

VALE F., CORREIA AC, MATOS B, NUNES JFM & APA, D. M. 2010. *Applications of transmission electron microscopy to virus detection and identification. Microscopy: Science, Technology, Applications and Education.,* Formatex Research Centre.

VALEGARD, K., MURRAY, J. B., STONEHOUSE, N. J., VAN DEN WORM, S., STOCKLEY, P. G. & LILJAS, L. 1997. The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J Mol Biol,* 270**,** 724-38.

VAN DEN HOOGEN, B. G., DE JONG, J. C., GROEN, J., KUIKEN, T., DE GROOT, R., FOUCHIER, R. A. & OSTERHAUS, A. D. 2001. A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat Med,* 7**,** 719-24.

VAN DER HOEK, L., PYRC, K., JEBBINK, M. F., VERMEULEN-OOST, W., BERKHOUT, R. J., WOLTHERS, K. C., WERTHEIM-VAN DILLEN, P. M., KAANDORP, J., SPAARGAREN, J. & BERKHOUT, B. 2004. Identification of a new human coronavirus. *Nat Med,* 10**,** 368-73.

VAN DER HOEK, L., SURE, K., IHORST, G., STANG, A., PYRC, K., JEBBINK, M. F., PETERSEN, G., FORSTER, J., BERKHOUT, B. & UBERLA, K. 2005. Croup is associated with the novel coronavirus NL63. *Plos Medicine,* 2**,** 764-770.

VIJGEN, L., KEYAERTS, E., ZLATEVA, K. & VAN RANST, M. 2004. Identification of six new polymorphisms in the human coronavirus 229E receptor gene (aminopeptidase N/CD13). *Int J Infect Dis,* 8**,** 217-22.

WAN, H., HE, J., JU, B., YAN, T., LAM, T. J. & GONG, Z. 2002. Generation of two-color transgenic zebrafish using the green and red fluorescent protein reporter genes gfp and rfp. *Mar Biotechnol (NY),* 4**,** 146-54.

WAN-JI LEE, YOON-SEOK CHUNG, HEE SOOK YOON, CHUN KANG & KIM, K. 2013. Prevalence and Molecular Epidemiologyof Human Coronavirus

HKU1 in PatientsWith Acute Respiratory Illness. *Medical Virology* 85:309–314

WAN., H., HE., J., JU., B., YAN., T., LAM., T. J. & GONG., Z. 2002. Generation of Two-color Transgenic Zebrafish Using the Green and Red Fluorescent Protein Reporter Genes gfp and rfp. *Mar. Biotechnol.,* 4**,** 146-154.

WANG, C., LE, S., ALI, N. & SIDDIQUI, A. 1995. An RNA pseudoknot is an essential structural element of the internal ribosome entry site located within the hepatitis C virus 5'noncoding region. *Rna,* 1**,** 526-537.

WEISS, S. R. & NAVAS-MARTIN, S. 2005. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol Mol Biol Rev,* 69**,** 635-64.

WILLIAMS, A. K., WANG, L., SNEED, L. W. & COLLISSON, E. W. 1992. Comparative analyses of the nucleocapsid genes of several strains of infectious bronchitis virus and other coronaviruses. *Virus Res,* 25**,** 213-22.

WOO, P. C., HUANG, Y., LAU, S. K., TSOI, H. W. & YUEN, K. Y. 2005a. In silico analysis of ORF1ab in coronavirus HKU1 genome reveals a unique putative cleavage site of coronavirus HKU1 3C-like protease. *Microbiol Immunol,* 49**,** 899-908.

WOO, P. C., LAU, S. K., CHU, C. M., CHAN, K. H., TSOI, H. W., HUANG, Y., WONG, B. H., POON, R. W., CAI, J. J., LUK, W. K., POON, L. L., WONG, S. S., GUAN, Y., PEIRIS, J. S. & YUEN, K. Y. 2005b. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol,* 79**,** 884-95.

WU, R., SATO, G. H. & WHITCUTT, M. J. 1986. Developing differentiated epithelial cell cultures: airway epithelial cells. *Fundam Appl Toxicol,* 6**,** 580-90.

YAMAUCHI, Y. & GREBER, U. F. 2016. Principles of Virus Uncoating: Cues and the Snooker Ball. *Traffic,* 17**,** 569-92.

YANG, D. & LEIBOWITZ, J. L. 2015. The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res,* 206**,** 120-33.

ZHOU, M. & COLLISSON, E. W. 2000. The amino and carboxyl domains of the infectious bronchitis virus nucleocapsid protein interact with 3' genomic RNA. *Virus Res,* 67**,** 31-9.

ZUCKERMAN, A. J., TAYLOR, P. E. & ALMEIDA, J. D. 1970. Presence of particles other than the Australia-SH antigen in a case of chronic active hepatitis with cirrhosis. *Br Med J,* 1**,** 262-4.

ZUKER, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research,* 31**,** 3406-3415.

ZUNIGA, S., CRUZ, J. L., SOLA, I., MATEOS-GOMEZ, P. A., PALACIO, L. & ENJUANES, L. 2010. Coronavirus nucleocapsid protein facilitates template switching and is required for efficient transcription. *J Virol,* 84**,** 2169-75.

ZUNIGA, S., SOLA, I., ALONSO, S. & ENJUANES, L. 2004. Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *J Virol,* 78**,** 980-94.

ZUNIGA, S., SOLA, I., CRUZ, J. L. & ENJUANES, L. 2009. Role of RNA chaperones in virus replication. *Virus Res,* 139**,** 253-66.

ZUNIGA, S., SOLA, I., MORENO, J. L., SABELLA, P., PLANA-DURAN, J. & ENJUANES, L. 2007. Coronavirus nucleocapsid protein is an RNA chaperone. *Virology,* 357**,** 215-27.