THE UNIVERSITY OF READING

RNA structure and function signals in a novel Nidovirus sequence from the Sea Hare *Aplaysia californica*

A thesis submitted for the degree of Doctor of Philosophy in Microbial sciences

By

Khulud Ibrahim Bukhari

Supervisor: Professor. Ian Jones

School of Biological Science

March 2019

Declaration

I confirm that this is my own work and that it has not been previously submitted for a 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.

Khulud Ibrahim Bukhari

Abstract

Nidoviruses are enveloped, positive stranded RNA viruses that are among the most complex plus-strand RNA viruses known. My work has discovered the sequence of a new, highly divergent virus, tentatively named Abyssovirus, in a pool of sequenced intracellular RNA from a metazoan. Based on comparison of the genomic organization and replicase subunits, Abyssovirus appears to belong to a new family of nidoviruses. The aim of the present project was to validate the sequence data by studying the function of the predicted unusual translational stop-start signal located in the replicase gene, and to express predicted viral proteases and test their catalytic activity in Sf9 insect cells, *E. coli* and mammalian expression systems. The translational stop-start signal was found to allow ribosomes to readthrough the natural stop codon in the expression systems tested. Additionally, Abyssovirus Main Protease (M^{Pro}) showed evidence of cleavage of fragments from both ends of the expressed protease, as is typical for polyprotein-embedded Nidovirus proteases. These results suggest that the Abyssovirus RNA sequence encodes biologically functional proteins, and therefore that the RNA sequence likely does represent an extant virus with some Nidovirus-like features and an Abyssovirus-specific mechanism to translate the second half of the replicase open reading frame.

Dedication

This thesis work is dedicated to my parents, Ibrahim and Shaha, who have been constant source of support and encouragement during the challenges of graduate school and life. I am truly thankful for having you in my life. This work is also dedicated to my sisters and brothers, Sarah, Ohoud, Wadha, Mohammed, Alanood, Alhanoof, Hadeel, Ahad and Khalid, and my friends, Hanan Alqassem and Hanan Alshehri who have always loved me unconditionally and whose good examples have taught me to work hard for the things that I aspire to achieve.

Acknowledgments

Firstly, I would like to express my sincere gratitude to Prof. Ian Jones for his constant support, advice and useful scientific discussions. His guidance helped me in all the time of research. I could not have imagined having a better advisor and mentor for my PhD study. I specially thank Dr. Mulley for the assistance she provided at all levels of the research project.

I must express my gratitude to Ahmad Saif, my friend, for his continued support and encouragement. A very special thanks goes out to Lubna Al-Dohyan, Dr. Ban Abdulsattar and Dr. Saad Mutlk who always kindly advised and helped me. I would also like to thank Entedar Alsaadi, Amenah Alotaibi and Nourah Almimoni for their help and support.

I would also like to thank all the staff in Knight building for their kindness and help, with special thanks to Dr. Silvia Loureiro, Hemanta Maity, Mai Uchida and Dr. Sinead Lyons.

My sincere gratitude goes to my sponsor Ministry of Higher Education in Saudi Arabia; I recognize that this research would not have been possible without their financial support.

Finally, I would like to thank my parents, my brothers and sisters and my friends. They were always supporting me and encouraging me with their best wishes. Your prayer for me was what sustained me thus far.

Table of Contents

Declaration ..... **I**

Abstract ..... **II**

Dedication ..... **III**

Acknowledgments ..... **IV**

Table of Contents ..... **V**

Index of Figures ..... **XI**

Index of Tables ..... **XVII**

Chapter 1 ..... **1**

1. Introduction **1**

1.1 Taxonomy **2**

1.2 Morphology **3**

1.3 *Nidovirales* genome **6**

1.4 *Nidovirales* replication **11**

1.5 Proteins **14**

1.5.1 Structural Proteins **14**

1.5.2 Non-structural Proteins **16**

1.5.2.1 ORF1a activities **17**

1.5.2.2 ORF1b activities **19**

1.6 The Discovery of Abyssovirus **21**

1.6.1 Main protease (M^{Pro}) **23**

1.6.2 Readthrough at the1a/1b junction **24**

1.6.3 Structural protease (S^{Pro}) **25**

1.7 Aims of this research **26**

Chapter 2 28

2. Materials and methods **28**

2.1 Plasmid construction and cloning of desired DNA fragment Oligonucleotides **28**

2.1.1 Primers **28**

2.1.2 PTriEx1.1 vector map **28**

2.1.3 Source and amplification of Abyssovirus DNA fragments **29**

2.1.4 Double digest and gel extraction of pTriEx1.1 vector **30**

2.1.5 In-Fusion cloning of fragments into the pTriEx1.1 vector **31**

2.1.6 Generating mutant Abyssovirus genes using overlap PCR **31**

2.1.7 Ligation of PCR fragments with pTriEx1.1 ..... **32**

2.1.8 Transformation of *E. coli* competent cells **32**

2.2 DNA agarose electrophoresis **33**

2.3 Colony PCR screening of transformant bacteria **33**

2.3.1 Colony PCR analysis **33**

2.3.2 Plasmid DNA purification and sequencing **34**

2.4 Transformation of different bacterial strains **35**

2.4.1 Preparing competent cells for the expression system BL21 (DE3)-pLysS ... **35**

2.4.2 Other *E. coli* strains ... **35**

2.5 Induction of protein expression in *E. coli* T7 compliant strains ... **37**

2.6 Transfection of Mammalian cells (17clone-1) ... **37**

2.7 Transfection of insect cells (Sf9) ... **38**

2.8 Protein purification from insect cells ... **38**

2.9 Cell culture ... **39**

2.9.1 17clone-1 mouse fibroblast cells ... **39**

2.9.2 Sf9 insect cells ... **40**

2.10 Protein analysis ... **40**

2.10.1 SDS Polyacrylamide gel electrophoresis (PAGE) ... **40**

2.10.2 Protein staining with Coomassie brilliant blue ... **41**

2.10.3 Western blot ... **41**

Chapter 3 ... 43

1. Cloning and expression of Abyssovirus main protease fragments in different expression systems ... **43**

3.1 Introduction ... **43**

3.2 Results ... **48**

3.2.1 Design and construction of tagged Abyssovirus main protease ... **48**

3.2.1.1 Generation of mutants in the Abyssovirus main proteases by overlap PCR ... **53**

3.2.2 Colony PCR screening of transformed bacteria **60**

3.2.3 Expression and proteolytic activity of Abyssovirus main protease **62**

3.2.3.1 Main protease expression in *E. coli* **63**

3.2.3.1.1 Abyssovirus M^{Pro} wild type and mutant expression in *E. coli* BL21 (DE3)-pLysS **66**

3.2.3.1.2 Abyssovirus M^{Pro} wild type and mutant expression in *E. coli* Rosetta **70**

3.2.3.2 Protein expression in the insect Sf9 cells **74**

3.2.3.3 Expression and optimization of Abyssovirus main protease in mammalian cells using two transfection reagents **81**

3.3 Discussion **88**

Chapter 4 94

4. Stimulation and translation program of Abyssovirus 1a/1b junction sequence …..... **94**

4.1 Introduction **94**

4.2 Results **98**

4.2.1 Construction of tagged Abyssovirus stop suppression vectors **98**

4.2.1.1 Generating of mutant Abyssovirus stop suppression (TGA-C) by using overlap PCR **103**

4.2.2 Colony PCR screening of transformed bacteria **109**

4.2.3 Analysis of Abyssovirus stop codon readthrough signal **110**

4.2.3.1 Abyssovirus TGA-C readthrough in *E. coli* BL21 (DE3)- pLysS strain **111**

4.2.3.2 Abyssovirus TGA-C readthrough expression in the insect Sf9 cells **115**

4.2.3.3 Expression of Abyssovirus readthrough stop suppression site signaled in mammalian cells **118**

4.3 Discussion **120**

Chapter 5 128

5. The biological function of Abyssovirus Structural protease **128**

5.1 Introduction **128**

5.2 Results **131**

5.2.1 Construction of tagged Abyssovirus structural protease vectors **131**

5.2.2 Generating of mutant Abyssovirus structural protease constructs **135**

5.2.3 Screening of transformed bacteria using *Xho*I restriction enzyme **136**

5.2.4 Generating of mutant Abyssovirus Structural protease by using overlap PCR method **137**

5.2.5 Colony PCR screening of transformed bacteria **141**

5.2.6 Expression and proteolytic activity of Abyssovirus structural protease **143**

5.2.6.1 Abyssovirus structural protease expression in different *E. coli* strains **143**

5.2.6.1.1 Structural protease protein expression in *E. coli* BL21 (DE3) pLysS **144**

5.2.6.1.2 Structural protease protein expression in *E. coli* Rosetta strain **149**

5.2.6.2 Abyssovirus structural protease expression in the insect Sf9 cells **152**

5.2.7 Abyssovirus structural protease purification using IMAC from insect cells **155**

5.3 Discussion **161**

Chapter 6 164

6. General discussion **164**

References 174

Appendix 208

Appendix 1: Publications 208

Appendix 2: Poster Presentations ……..…….. 220

Appendix 3: Synthetic DNA Fragments. ……221

Index of Figures

Figure 1.1 Schematic representations of coronavirus, torovirus, arterivirus and ronivirus which belong to the order *Nidoviridae* **4**

Figure 1.2. Schematic diagram of Mesonivirus showing the conserved structural components …..... **6**

Figure 1.3. Shows four nidovirus genomic organizations and key encoded domains **9**

Figure 1.4. The life cycle of an arterivirus as a typical member of the *Nidovirales* **13**

Figure 2.1. Map of pTriEx1.1 vector **29**

Figure 3.1. The genome of Murine hepatitis virus (MHV) as a typical coronavirus with nsp5 location shown **44**

Figure 3.2. HHprep output showing the identification of the Abyssovirus M^{Pro} sequence within the submitted translated 1a region of the TSA sequence GBBW01007739.1 as a member of the main coronavirus polyprotein processing enzymes **47**

Figure 3.3 Location of nsp5 protease reading frame on the Abyssovirus genome and choice of mutation ……..... **49**

Figure 3.4. Plasmid map showing the complete Abyssovirus M^{Pro} sequence assembled into the pTriEx1.1 vector **52**

Figure 3.5. Generation of mutant Abyssovirus main protease fragments by overlap extension PCR **54**

Figure 3.6. Gel electrophoresis of PCR products following amplification of Abyssovirus M^{Pro} first active site mutation using specific primers **57**

Figure 3.7. Gel electrophoresis of PCR products following amplification of the Abyssovirus M^{Pro} second active site mutation using specific primers **57**

Figure 3.8. Gel electrophoresis of PCR products following amplification of four individual cleavage site mutations using mutation-specific primers **58**

Figure 3.9. Agarose gel electrophoresis of overlap product of Abyssovirus two active sites and four-cleavage sites mutants amplified by PCR **59**

Figure 3.10. Main protease mutations screened by colony PCR **61**

Figure 3.11. Western blot analysis of the wild type Abyssovirus M^{Pro} expression in *E. coli* **65**

Figure 3.12. Western blot analysis of wild and mutant Abyssovirus M^{Pro} in *E. coli* BL21 (DE3)-pLysS with HSV tag antibody..... **67**

Figure 3.13. Western blot analysis of wild type and mutant Abyssovirus M^{Pro} protease in *E. coli* BL21 (DE3)-pLysS with HIS tag antibody **69**

Figure 3.14. Western blot analysis by HSV tag antibody of wild type and mutant Abyssovirus M^{Pro} in *E. coli* Rosetta **71**

Figure 3.15. Western blot analysis by HIS tag antibody of wild type and mutant Abyssovirus M^{Pro} in *E. coli* Rosetta **73**

Figure 3.16. Abyssovirus main protease wild type and active site mutants expressed in insect cells **76**

Figure 3.17. Abyssovirus main protease cleavage site mutants expressed in insect cells **78**

Figure 3.18. Expression of wild Abyssovirus main protease, mutant active sites and cleavage sites in insect cells **80**

Figure 3.19. *In vitro* transfection of 17clone-1 cells with pTriEx1.1-GFP using Lipofectin and Lipofectamine 3000 transfection reagents **83**

Figure 3.20. Main protease proteins expressed in mammalian cells **85**

Figure 3.21. Western blot analysis of main protease proteins expressed in mammalian cells using Polyethylenimine (PEI) transfection reagent **87**

Figure 4.1. Elements required for ribosomal frameshifting in MHV **95**

Figure 4.2. Sequence from the Abyssovirus genome showing stop suppression sequence located in the frame at the junction of 1a and 1b **100**

Figure 4.3. Plasmid map showing the complete Abyssovirus stop suppression site and pseudoknots, located in the readthrough frame, sequence in the pTriEx1.1 vector **102**

Figure 4.4. Gel electrophoresis of PCR products following amplification of the Abyssovirus stop suppression signal TGA-C or TAA-A to introduce deletions of the pseudoknot **105**

Figure 4.5. Agarose gel electrophoresis of overlap products of Abyssovirus stop suppression site TGA-C and TAA-A deletion mutations **107**

Figure 4.6. Gel electrophoresis of digested mutant Abyssovirus stop suppression site (TGA-C and TAA-A) **108**

Figure 4.7. Double digest of pTriEx1.1 ..... **108**

Figure 4.8. Mutants of Abyssovirus stop suppression sequences (TGA-C and TAA-A with no pseudoknot) screened by PCR **109**

Figure 4.9. Western blot analysis of recombinant expression of the wild (TGA-C) and mutant (no pseudoknot) Abyssovirus readthrough stop suppression site signal

proteins **112**

Figure 4.10. Western blot analysis of recombinant expression of the wild (TAA-A) and mutant (no pseudoknot) Abyssovirus readthrough stop suppression site proteins **114**

Figure 4.11. The Abyssovirus readthrough stop suppression signal and mutant proteins expressed in Sf9 insect cells **117**

Figure 4.12. Abyssovirus stop suppression site signal proteins expressed in mammalian cells **119**

Figure 4.13. Percentage of Abyssovirus readthrough proteins expression in the insect cells Sf9 ..... **125**

Figure 5.1. Sindbis Virus, one of alphavirus members, structural proteins (capsid, E3, E2, 6K and E1) and their positions through an endoplasmic reticulum membrane **129**

Figure 5.2. The genome of Abyssovirus and the location of the structural protease ..**132**

Figure 5.3. Plasmid map showing the complete Abyssovirus structural protease sequence in the pTriEx1.1 vector **134**

Figure 5.4. Gel electrophoresis of different mutants in pTriEx1.1 plasmid digested with *Xho*I enzyme **137**

Figure 5.5. Gel electrophoresis of PCR products following amplification of the Abyssovirus structural protease gene to create active site mutation (S243A) using specific overlap primers **139**

Figure 5.6. Agarose gel electrophoresis of overlap product of Abyssovirus structural protease mutation amplified by PCR **140**

Figure 5.7. Mutant Abyssovirus structural protease screened by PCR **142**

Figure 5.8. Western blot analysis by HSV tag antibody of recombinant S^{Pro} expression of expression in *E. coli* BL21 (DE3)-pLysS **146**

Figure 5.9. Western blot analysis by HIS tag antibody of recombinant S^{Pro} expression of expression in *E. coli* BL21 (DE3)-pLysS **148**

Figure 5.10. Western blot analysis of recombinant expression of wild structural and first active site mutation S^{Pro} in Rosetta cells using HSV and HIS antibodies **151**

Figure 5.11. Abyssovirus structural protease proteins expressed in Sf9 insect cells . **154**

Figure 5.12. Elution profile of the Abyssovirus structural protease extract after affinity chromatography **158**

Figure 5.13. Analysis of elution fractions following IMAC of Abyssovirus S^{Pro} extracts from Sf9 insect cells **159**

Figure 5.14. Western blot of elution fractions from IMAC purification of structural protease in Sf9 cells **160**

Figure 6.1. Description of Abyssovirus host organism and genome **165**

Figure 6.2. Investigation overview of selected Abyssovirus sequences assayed for biological function **167**

Index of Tables

Table 2.1. *E.coli* strains used for plasmid transformation and protein expression **36**

Table 3.1. Oligonucleotides used for in-Fusion cloning of wild Abyssovirus main protease in pTriEx1.1 plasmid **51**

Table 3.2. Oligonucleotides used for cloning in pTriEx1.1 plasmid **55**

Table 3.3. Flanking oligonucleotides used to generate mutant main protease in all cases **56**

Table 4.1. Oligonucleotides used for in-Fusion of wild type and mutant Abyssovirus stop suppression fragments (TGA-C and TAA-A) into pTriEx1.1 plasmid **101**

Table 4.2. Oligonucleotides used for cloning in pTriEx1.1 plasmid **104**

Table 4.3. Oligonucleotides used to generate mutant Abyssovirus stop suppression site sequence cases **105**

Table 5.1. Oligonucleotides used for the in-Fusion cloning of the Abyssovirus structural protease in pTriEx1.1 ..... **133**

Table 5.2. Oligonucleotides used for mutation of the S^{Pro} gene in pTriEx1.1 plasmid **138**

Table 5.3. Oligonucleotides used to generate mutant structural protease case ……. **139**

Chapter 1

1. Introduction

Nidoviruses are enveloped, positive stranded RNA viruses that are among the most complex plus-strand RNA viruses known (Seybert *et al.,* 2005; Posthuma *et al.,* 2005; Pasternak *et al.,* 2006). Nidoviruses are grouped together based on similarities in their genome organization, similar strategies for nonstructural and structural protein expression, and structurally and functionally similar replicative enzymes (Cavanagh *et al.,* 1997; Cowley *et al.,* 2000; den Boon *et al.,* 1991; Gorbalenya *et al.,* 1989; Snijder *et al.,* 2005). Nidoviruses mostly infect mammals (coronaviruses, toroviruses and arteriviruses) however nidoviruses can also infect avian (coronaviruses) or invertebrate (roniviruses, mesoniviruses) hosts. This order causes a variety of diseases that can range from asymptomatic, persistent carrier-states to fatal infections. In 2003, Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) was discovered and spread worldwide (Drosten *et al.,* 2003; Ksiazek *et al.,* 2003; Peiris *et al.,* 2003), followed by the discovery and ongoing spread of Middle Eastern Respiratory Syndrome coronavirus (MERS-CoV) in 2012 (Zaki *et al.,* 2012). In the wake of SARS-CoV, a number of novel coronaviruses, including two that infect humans, HCoV-NL63 and HCoV-HKU1 (van der Hoek *et al.,* 2004; Fouchier *et al.,* 2004; Woo *et al.,* 2005), were recognized and added to the growing list of nidoviruses first characterized during the past two decades. This list continues to grow and it is likely there are many more nidoviruses than are currently known. In this study we used BLAST searches to scan the publicly available transcriptomes and expressed sequence tag libraries available at the US National Center for Biotechnology Information to reveal novel nido-like virus

sequences from several transcriptome studies dealing with the marine gastropod *Aplysia californica,* the sea hare (Fiedler *et al.,* 2010; Heyland *et al.,* 2011; Moroz *et al.,* 2006).

1.1 Taxonomy

Nidoviruses encode up to thirty proteins that are produced by ribosomal frameshifting, cleavage by multiple proteases and subgenomic RNA synthesis, making them among the most complex single stranded RNA (ssRNA) viruses. The name of the order, *Nidovirales,* is derived from the Latin word *nidus* which means "nest" as all viruses in this order synthesize a nested set of subgenomic RNAs (sgRNAs) during their replication process. Nidovirus genomes possess a linear 5'-capped positive sense single stranded, non-segmented RNA with two large open reading frames (ORFs) 1a and 1b located at the 5'-end of the virus genome. A set of smaller open reading frames are encoded in the 3' end of the genome. In the past, the order *Nidovirales* included three families, *Coronaviridae*, *Roniviridae* and *Arteriviridae*, the first two families with large genome size (26.3-31.3 kb) and the third family with a smaller genome size (12.7-15.7 kb). In recent times however a fourth family has been added to the *Nidovirales,* the *Mesonviridae,* with an intermediate genome size between those of the *Coronaviridae* and *Arteriviridae* (Gorbalenya *et al.,* 2006;Lauber *et al.,* 2012; Nauwynck *et al.,* 2012; ICTV:http://ictvonline.org/virusTaxonomy.asp).

1.2 Morphology

All members of the order *Nidovirales* are enveloped viruses with rod-shaped or pleomorphic architecture (Fauquet *et al.,* 2005). It has been shown that two genera of the family *Coronviridae* (Coronavirus and Torovirus), as well as members of the families *Roniviridae* and *Mesoniviridae,* have large projections (peplomers) protruding from the virus envelope that are formed, especially in the *Coronaviridae,* by trimers of the spike protein S (Fauquet *et al.,* 2005). This gives the virion a crown-like morphology in electron microscope images, hence the name coronavirus. The typical virions of coronaviruses are broadly spherical but fairly pleomorphic with an average diameter of 50-150 nm (Neuman, 2008). In Toroviruses, the nucleocapsid has a torus-shape, hence the name (Duckmanton *et al.,* 1997; Dervas *et al.,* 2017) and the virion size, along with the related bafiniviruses, is similar to coronaviruses. Roniviruses however have a bacilliform shape with a long nucleocapsid (150-200 nm) with helical symmetry and a diameter of 20-30 nm (Fauquet *et al.,* 2005). Virions of the arteriviruses are spherical and significantly smaller than the other members of the nidoviruses with a complete particle of 50-70 nm diameter and a nucleocapsid of 25-23 nm in diameter. While arteriviruses were originally described as possibly having an icosahedral core shell (Fauquet *et al.,* 2005), recent cryo-electron tomography has shown that the arterivirus Porcine Reproductive and Respiratory Syndrome Viruses (PRRSV) is pleomorphic (Snijder *et al.,* 2013). Heterotrimeric spikes can be occasionally found on the surface of arteriviruses (**Figure 1.1**) (Fauquet et al., 2005).

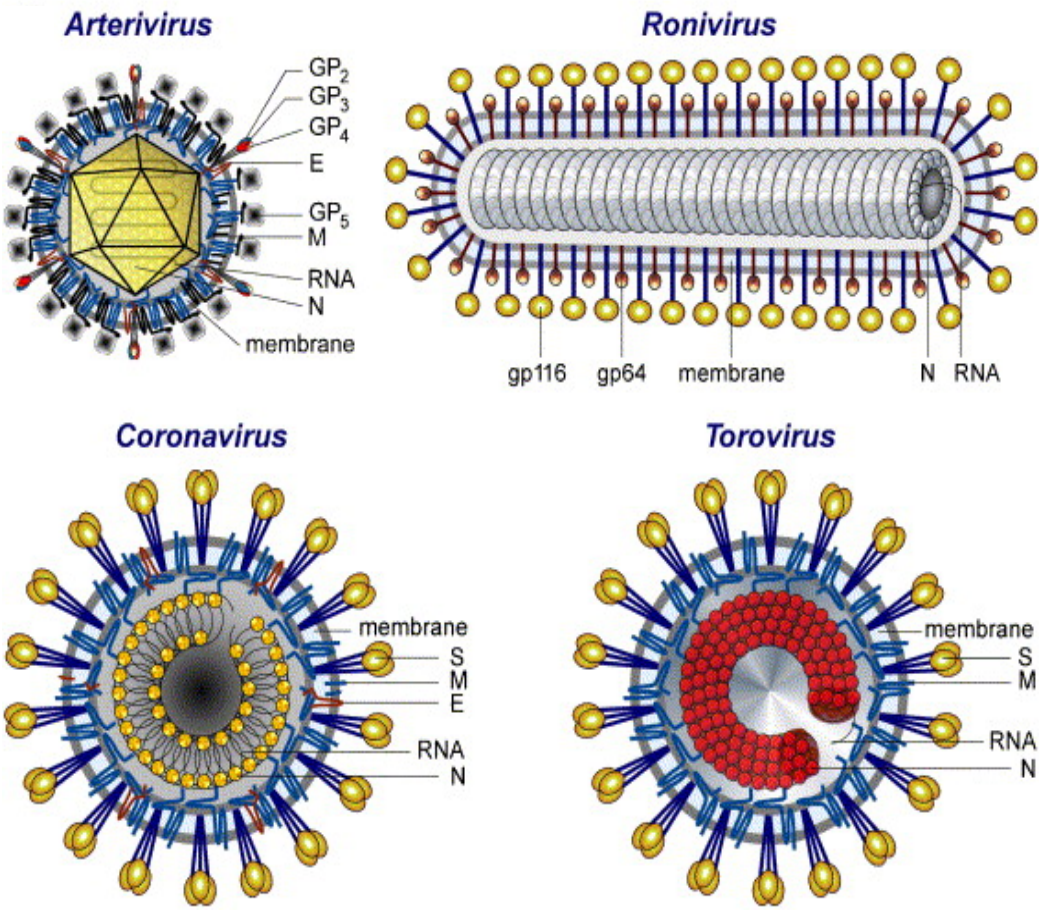

Figure 1.1 Schematic representations of coronavirus, torovirus, arterivirus and ronivirus which belong to the order *Nidoviridae*. N, nucleocapsid protein; S, spike protein; M, membrane protein; E, envelope protein; HE, hemagglutinin-esterase. The M protein in coronavirus cooperates with the N protein. In arterivirus, GP5 and M are major envelope proteins, however GP2, GP3, GP4, and E are minor envelope proteins. Both of toro- and roniviruses are lacks of the E protein, which is present in Corona and arteriviruses. An equivalent of the M protein has not yet been recognized in roniviruses (Taken from: Gorbalenya *et al.,* 2006).

The *Mesoniviridae* were founded with one species, which is represented by two viruses, Cavally Virus (CavV) and Nam Dinh Virus (NDiV). CavV represented the first insect associated nidovirus (Zirkel *et al.,* 2011), while NDiV was recognized in mosquitoes in Vietnam. Together, these viruses are the prototypes of a novel virus family in the *Nidovirales* (Nga *et al.,* 2011). Electron microscopy images of mesoniviruses show prominent club-shaped surface projections protruding from spherical enveloped virions (**Figure 1.2**) (Zirkel *et al*., 2013) and immature particles, possibly representing CavV nucleocapsids, 50 to 60 nm in diameter, have been observed in infected cells. NDiV virions were described as enveloped and spherical but with a lack of projections (Nga *et al.,* 2011). The genome size of the *Mesoniviridae* (CavV and NDiV) is approximately 20 kb, and falls between the large Nidoviruses (Coronovirus (CoV), Torovirus (ToV), and Ronivirus (RoV), 26 to 32 kb) and the small Nidoviruses (Arterivirus (ArV), 13 to 16 kb), hence the name *Mesoniviridae* (Lauber *et al.,* 2012).

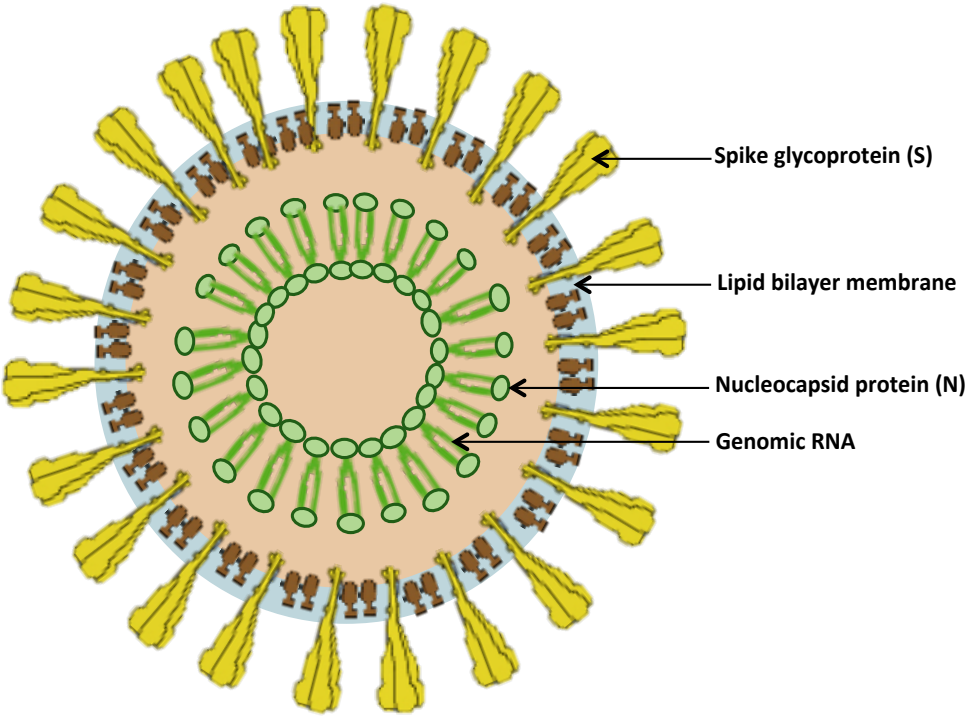

Figure 1.2. Schematic diagram of Mesonivirus showing the conserved structural components. (Adapted from Viral Zone http://viralzone.expasy.org/all_by_species/4776.html).

1.3 *Nidovirales* genome

All the members of the *Nidovirales* order have a 5'-capped, 3'-polyadenylated single stranded genome of positive polarity that can function as mRNA upon entry into the infected cell (Nga *et al.,* 2011). More than a hundred genetically distinct full-length nidovirus genomes and over a thousand partial genome sequences have been identified (Gonzalez *et al.,* 2003). The genome also includes untranslated regions at the 5' and 3' termini (Gorbalenya *et al.,* 2006).

The genome of coronaviruses contains up to 15 genes with multiple ORFs encoding both structural and non-structural proteins (Sawicki *et al.,* 2005; Ziebuhr, 2005). At the 5' end there is a leader sequence followed by an untranslated region (UTR). The first gene (the replicase) comprises two-thirds of the genome with two large ORFs (ORF 1a and ORF 1b) that are translated later to polyprotein 1a and polyprotein 1ab through a programmed ribosomal -1 frameshift mechanism (Gorbalenya *et al.,* 2006). The replicase is the only translated product derived from the genomic RNA. The last third of the genome encodes the structural proteins with the order S-E-M-N, which are expressed following the production of several subgenomic RNAs. Between these genes there are a variable number of ORFs, depending on the virus, encoding accessory proteins (Sawicki *et al.,* 2007). The transcriptional regulatory sequences (TRSs) that are present at the 3' end of each gene represent a signal for subgenomic RNA transcription. At the 3'end of the genome there are 270-500 nucleotides of untranslated region (UTR) followed by a poly A tail (**Figure 1.3.**).

The arterivirus genome is also a polycistronic positive sense RNA, capped at the 5' terminus and polyadenylated at the end 3' but is only 12-16 kb long (Snijder *et al.,* 2013). Whole genome sequences of many arterivirus are available for European and North American EAV (Equine Arteritis Virus) isolates, a large number of genotype 1 and 2 PRRSVs (Porcine Reproductive and Respiratory Syndrome Viruses), two LDV (Lactate Dehydrogenase Elevating Virus) strains, five distantly related monkey viruses, tentatively grouped under the name SHFV (Simian Hemorrhagic Fever Virus) and Wobbly Possum nidovirus (Snijder *et al.,* 2013). The genome contains 5' (156–224 nt) and 3' (59–117 nt) NTRs (nontranslated regions). The genome sequence contains 10-15 known ORFs. ORF 1a and 1b are the large

replicase that is located in the 5' proximal three-quarters of the genome (Snijder *et al*., 2013) (**Figure 1.3.**). The 3' of the genome contains sequences for the structural proteins, which are translated from a nested set of subgenomic mRNAs (Kuhn *et al.,* 2016).

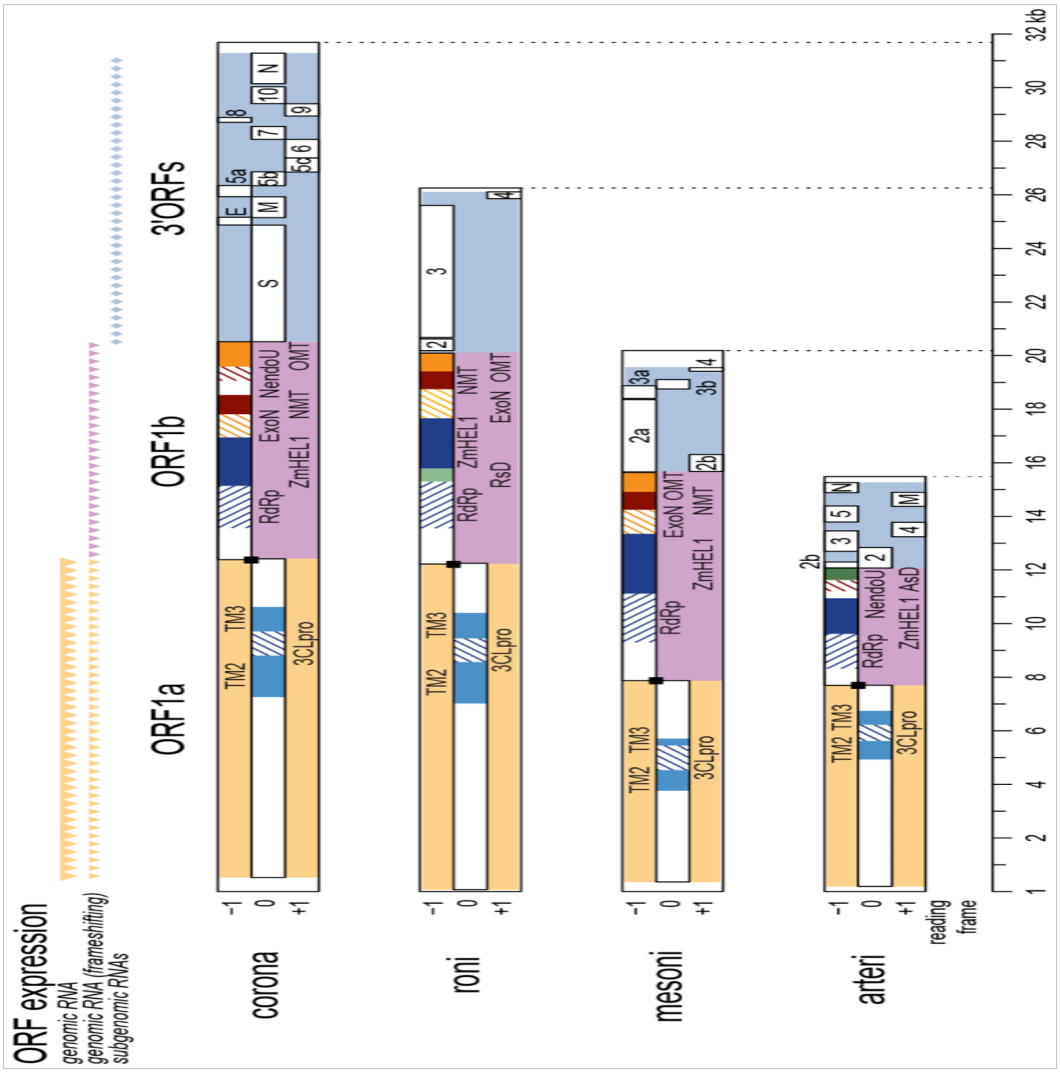

Figure 1.3. Shows four nidovirus genomic organizations and key encoded domains. The coding regions on the genomes are divided into ORF1a (yellow), ORF1b (violet) and the 3'ORFs (blue). Black squares represent ribosomal frameshifting sites. Inside ORFs (white rectangles) and coloured patterns highlight domains recognized in: all nidoviruses [TM2, TM3 (transmembrane domains), 3CLpro (or M^{Pro}), RdRp, and Zn-cluster binding domain fused with HEL1 (ZmHEL1) (van Dinten *et al.,* 2000)- light and dark blue], large nidoviruses (ExoN (exoribonuclease), OMT (2'-*O*-methyltransferase) - orange), certain clades (NMT (N7-methyltransferase), NendoU (uridylate-specific endonuclease) - red; Ronivirus-specific domain (RsD) - light green; arterivirus-specific domain (AsD) - dark green) (taken from Lauber *et al.,* 2013).

The genome of both CAVV and NDIV mesoniviruses are similarly single stranded RNA, positive sense and contain seven ORFs. The bigger ORF1a and ORF1b in the 5' of the genome and the smaller ORF2a, ORF2b, ORF3a, ORF3b and ORF4 located at the 3' of the genome (Zirkel *et al.,* 2013). As in other nidoviruses, ORF1a and ORF1b overlap by a few nucleotides and encode the two replicase polyproteins which are processed into the viral polymerase (RdRp) and other non-structural proteins. (http://viralzone.expasy.org/all_by_species/4776.html; Zirkel *et al.,* 2011; Nga *et al.,* 2011). The translation of ORF1a and ORF1b is regulated by a ribosomal frameshift sequence GGAUUUU (Zirkel *et al.,* 2011; Nga *et al.,* 2011). At the 5' of the genome there is a leader sequence, as in some *Nidoviridae* families (Zirkel *et al.,* 2011; Pasternak, 2006; Schutze *et al.,* 2006). The 3' of the CavV and NDiV genomes encode the structural proteins including a spike (S) glycoprotein (ORF2a), a nucleocapsid (N) protein (ORF2b), two proteins with membrane-spanning regions (ORF3a and -3b), and a small protein with unknown function (ORF4) (**Figure 1.3**) (Zirkel *et al.,* 2011; Nga *et al.,* 2011).

Okavirus is a genus of the family *Roniviridae* which includes two genotypes, Gill-Associated Virus (GAV) and Yellow Head Virus (YHV). As in other nidoviruses, the Okavirus genome is a linear positive sense single stranded RNA encoding a large replicase polyprotein that is expressed from ORF1a and ORF1b by ribosomal frameshifting (Firth and Atkins, 2009). As before the structural proteins are located at the 3' end of the genome (**Figure 1.3**). Unusually, in the Okavirus case the ORFs encoded in the 3' end of the gnome are in the form of a polyprotein (ORF3; ~1650 codons) that generates 3 mature proteins (envelope glycoproteins gp116 and gp64,

and an ~25 kDa N-terminal fragment), a nucleocapsid (ORF2; ~145 codons), and an extra short ORF (ORF4) located at the 3' sequence of ORF3.

1.4 *Nidovirales* replication

The genome of all nidoviruses acts as a messenger RNA (mRNA), which means that the genome is translated directly into viral proteins by host ribosomes. The common feature of nidoviruses, and all positive stranded RNA viruses, is that replication takes place in the cytoplasm of the infected cell, in cooperation with the modification of intracellular membranes which act as a platform for the RNA viral replication complexes (RCs) that are involved in RNA synthesis (Knoops *et al.,* 2012; Ulasli *et al.,* 2010).

The nidovirus genome is polycistronic and employs a unique mechanism of discontinuous RNA synthesis in order to express the ORFs that occur downstream of ORF1ab. These ORFs, encoding the structural and some accessory proteins are translated from a nested set of subgenomic mRNAs (sgRNAs). The sgRNAs contain a short common 5' leader sequence and are 3' co-terminal. The 3' end of the leader sequence includes the transcriptional regulatory sequence (TRS-L), which is also present in the genome just upstream of the coding sequence for each transcription unit (TRS-B (body)), where it acts as a *cis*-regulator of transcription. The leader and body segments of the sgRNAs are linked during discontinuous negative strand RNA synthesis to produce a subgenome-length template for each sgRNA (Sawicki *et al.,* 1995; Pasternak *et al.,* 2006). After entry, the nidovirus genome is translated to produce ppa1 and pp1ab via a -1 ribosomal frameshift (RFS) that is located at the 3' end of ORF1a (den Boon *et al.,* 1991; Brierley *et al.,* 1989). Both pp1a and pp1ab polyproteins are processed into subunits by nsp3 (a

papain-like proteinase) and nsp5 (the main proteinase). The RNA-dependent RNA polymerase (RdRp) is contained within nsp12 and together up to 16 mature proteins derive from ORF 1ab, 11 non-structural proteins from pp1a and 5 additional non-structural proteins from the 1b region of pp1ab (Fung and Liu, 2014).

The product of replication, full length negative strand RNA acts as the template to synthesize both more full-length genomic RNA and the sgRNAs that translate to give the structural proteins (Sawicki *et al.,* 2007). Discontinuous RNA synthesis is governed by the interaction of the replicase with the transcription-regulating sequences, TRSs or intergenic sequences, IGs, that are repeated at the transcriptional start sites for each sgRNA (La Monica *et al.,* 1992). However, only one protein is translated from each sgRNA even if the sequence of other ORFs is present. The 5' ORF is translated while and the downstream ORFs are silent and the sgRNAs are therefore functionally monocistronic (Lia and Anderson, 2007). After translation of the structural proteins (S, E and M proteins), N protein wraps the newly formed genomic RNA in order to form the nucleocapsid and the other structural proteins (include the HE protein in some species) coalesce in areas of the endoplasmic reticulum (ER). The structural proteins and the nucleocapsid come together in the endoplasmic reticulum Golgi intermediate compartment (ERGIC) where virion assembly occurs, driven by the M and E proteins (Hsieh *et al.,* 2005; Tseng *et al.,* 2010). At the end of the assembly process, virions are released from the cell by the secretory pathway (**Figure 1.4**; Snijder *et al.,* 2013).

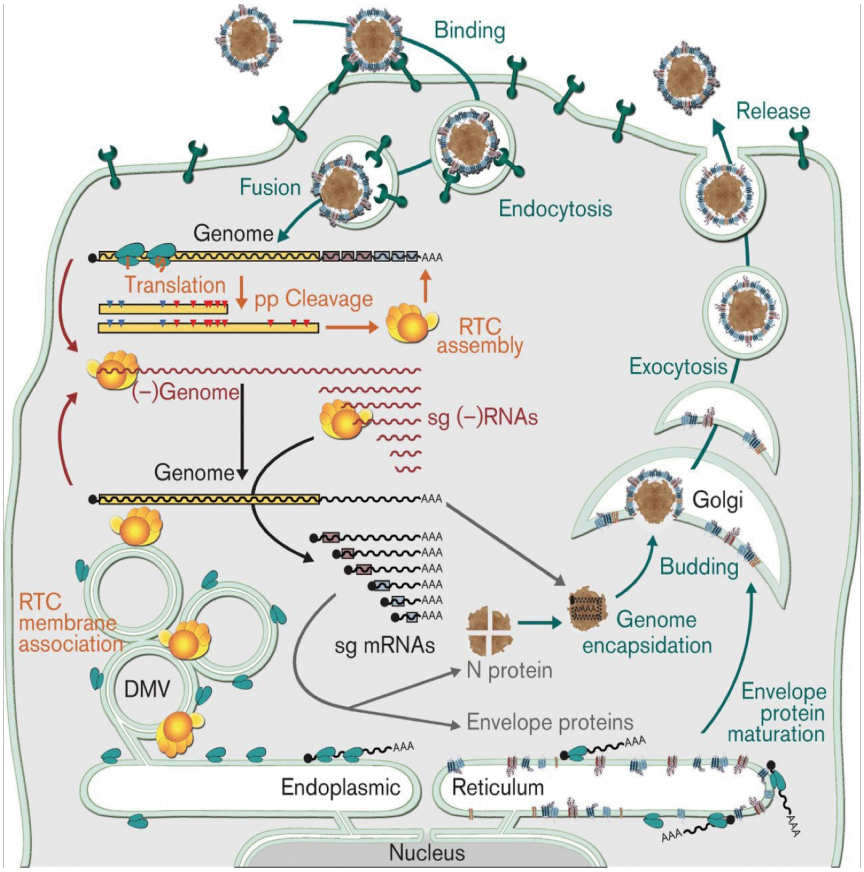

Figure 1.4. The life cycle of an arterivirus as a typical member of the *Nidovirales*. Following entry by receptor-mediated endocytosis and release of the genome into the cytosol of the host cell, genome translation produces the replicase polyproteins (pp1a and pp1ab; yellow bars). After the cleavage of the polyproteins by multiple proteases, the viral nonstructural proteins assemble into a replication and transcription complex (RTC) that performs minus-strand RNA synthesis. Both full-length and subgenome-length minus strands are produced, with the latter templating the synthesis of subgenomic messenger RNAs (mRNAs) required to express the structural protein genes encoded in the 3′of the genome. Finally, novel genomes are packaged into nucleocapsids that become enveloped by budding from smooth intracellular membranes, after which the new virions leave the cell by following the exocytic pathway (taken from Snijder *et al.,* 2013).

1.5 Proteins

1.5.1 Structural Proteins

The virion envelope of coronavirus contains at least 3 viral proteins including the spike protein (S), the membrane protein (M) and the envelope protein (E). Some coronaviruses also contain a Hemagglutinin-Esterase (HE). The spike protein (S) consists of a single protein with type I transmembrane topology, ranging from 1,160 to 1,400 amino acids in length that is highly glycosylated with 21 to 35 N-glycosylation sites (Belouzard *et al.,* 2012). The spike protein (S) is located on the surface of the virion and is essential for the virus to attach and enter the host cell (Kleine-weber *et al.,* 2018). The *Coronavirinae* are further divided into four groups, the alpha, beta, gamma and delta coronaviruses (Fehr and Perlman, 2015). In most Alphacoronaviruses and the Betacoronaviruses the spike protein (S) is uncleaved, however, in some Betacoronaviruses and all Gammacoronaviruses the spike protein is cleaved to two domains S1 and S2 through furin or another Golgi-resident host protease (Belouzard *et al.,* 2012). The S1 subunit forms the globular head of the spike, which binds to host receptors while S2 forms the stalk region of the spike and is anchored in the membrane. It contains two heptad repeats HR1 and HR2, which are typical features of class I viral fusion proteins, a putative fusion peptide FP and one short transmembrane domain (Kirchdoerfer *et al.,* 2016, Shang *et al.,* 2018). S2 mediates fusion of viral and host membranes and syncytium formation (Belouzard *et al.,* 2012; Li, 2015). The Hemagglutinin-Esterase protein (HE) present in some coronaviruses may also play a role in the attachment of the virus to host cells (Bakkers *et al.,* 2016).

The nucleocapsid protein (N) is a multifunctional phosphoprotein (Mcbride *et al.,* 2014) with a typical molecular mass of 50-60 kDa and can form dimers and higher order oligomers. In MHV (Mouse Hepatitis Virus) the N protein has 3 domains, an RNA binding domain, an oligomerization domain and a flexible linker which connects the two (Tan *et al.,* 2006). The primary function of the N protein is packaging the RNA genome into a helical ribonucleoprotein (Barcena *et al.,* 2009). N protein also has an important role in replication and discontinuous transcription (Wu *et al.,* 2014) and the multifunctional properties of coronavirus N proteins have led to suggestions that it represents an attractive target for the development of antiviral compounds (Grossoehme *et al.,* 2009).

The envelope protein (E) is a small (76–109 amino acid) protein with a single predicted hydrophobic domain (HD). E protein has ion channel activity *in vitro* suggesting that the E protein has an essential function for efficient trafficking of virions through the secretory pathway (Ruch and Machamer, 2012). The expression of E protein only or E together with M protein is sufficient for the synthesis of virus like particles (VLP) although other reports also suggest that M and N (Huang *et al.,* 2004), M and E (Hsieh *et al.,* 2005) and M protein only are able to drive the release of vesicles in different expression systems and cell types (Tseng *et al.,* 2010).

The membrane protein (M) is the most abundant protein in the virus particle and in infected cells and contains a short amino terminus located outside the virion, followed by three transmembrane domains (TM), and a large carboxy-terminal domain located inside the virion (Hogue and Machamer, 2008). As noted above,

coronavirus M proteins play key roles in virion assembly, budding and maturation (Wang *et al.,* 2009).

Toroviruses and mesoniviruses encode a coronavirus-like array of structural proteins but the structural apparatus of the *Arteriviridae* is quite different. Some arteriviruses such as EAV and PRRSV contain up to 6 envelope proteins, which might be common to all arteriviruses. These include a disulfide-linked heterodimer of non-glycosylated membrane proteins M (16-20 kDa) and GP5, cross the membrane 3 times, and are functionally analogous to the M protein of coronaviruses (Zhengchun *et al.,* 2012; Fauquet *et al.,* 2005).

Yellow Head Virus (YHV) contains 3 structural proteins including gp116 protein (110-135 kDa), gp64 protein (63-67 kDa) and p20 protein (20-22 kDa). The first two proteins are glycosylated and appear to be envelope proteins that form peplomers on the virion surface while p20 is related to the nucleoprotein and appears to function as the nucleocapsid, binding the RNA genome (Seibert *et al.,* 2012).

1.5.2 Non-structural proteins

ORF1 encodes the viral replicase proteins which are essential for RNA synthesis and virus replication (Freeman *et al.,* 2014). The translation of both ORF1a and ORF1b, which encode all the required proteins for RNA synthesis, produces 2 large polyproteins pp1a and pp1ab, the translation of the pp1ab polyprotein being achieved via a programmed ribosomal frameshifting event (Nakagawa *et al.,* 2016). The frameshift site consists of a ‘‘slippery site’’ (UUUAAAC) followed by a pseudoknot structure (Hu *et al.,* 2016). The two polyproteins, pp1a and pp1ab, are

processed by a papain-like proteinase(s) (PLpro) and the 3C-Like main proteinase (M^{Pro}) to produce 16 non-structural proteins (nsp) 1 to 16 (Lei & Hilgenfeld, 2017), nsp1–nsp11 encoded in ORF1a and nsp12–16 encoded in ORF1b (Tanaka *et al.,* 2013).

1.5.2.1 ORF1a activities

The first protein in ORFa1 is nsp1 and it has several functions including inhibition of host gene translation by binding to the 40S ribosomes to inactivate their translation functions and induce host mRNA degradation (Narayanan *et al.,* 2015). This is part of the virus suppression of the host antiviral signaling pathways (Jauregui *et al.,* 2013) and prevents type I interferon production in infected cells (Zhang *et al.,* 2016). Nsp2 acts as a cofactor for the function of nsp4 (Rascon-Castelo et al., 2015) while nsp3 is a large multidomain protein that has the ability to interact with other proteins involved in the replication and transcription processes (Lei *et al.,* 2018). Several domains have been identified in nsp3 including an N terminal-acidic domain (AC), ADP-ribose1-phosphatase (ADRP) and poly-ADP-ribose binding, PL2pro and Y domain (Angelini *et al.,* 2013). The function of PLPs (papain-like protease) is cleavage of the three earliest processing events in the replicase (Niemeyer *et al.,* 2018; Gadlage *et al.,* 2010). The ADRP is associated with proteins involved in ADP-ribosylation or poly (ADP-ribose) polymerization and ATP-dependent chromatin re-modeling (Egloff *et al.,* 2004; Saikatendu *et al.,* 2005). By contrast there is no clear function for the Y domain (Lei *et al.,* 2018).

Nsp4 is about 500 amino acids in length and contains four transmembrane helices and an internal carboxy-terminal domain (Sakai *et al.,* 2017; Gonzalez-Ochoa *et al.,*

2013). This protein is fundamental for cytoplasmic membrane modification to form the replication-transcription complex (RTC), with the assistance of nsp2 and nsp6 (Angelini *et al.,* 2013).

Nsp5, also known as $3CL^{Pro}$ or M^{Pro}, contains 3 domains including a chymotrypsin-like fold, which is located in the first two domains, while the function of the third domain is still unclear (Stobart *et al.,* 2013).

Nsp6 plays an important role in the generation of the RTCs, also described as double membrane vesicles (DMVs). It has also been associated with the formation of autophagosomes from the ER (endoplasmic reticulum) via an omegosome intermediate (Cottam *et al.,* 2014). Nsp8 (22kDa) has up to seven subdomains and has been reported to bind to the eight subunit nsp7 (10 kDa) to form hexadecameric structures that have a positively charged channel which mediates RNA binding (Velthuis *et al.,* 2012). Similarly, it has been shown that nsp9 binds to RNA within the viral replication complex and interacts with other complex associated proteins (Zeng *et al.,* 2018). Nsp10 contains two zinc fingers, has an identifiable helicase domain in the N-terminus, a ß-strand and a loop in the C-terminus but despite these structural features no distinct enzymatic activity has been reported for it (Krishna *et al.,* 2003; Bouvet *et al.,* 2014). Recently however, it has been shown that nsp10 activates nsp16-mediated 2′-O-MTase activity (Bouvet *et al.,* 2014).

1.5.2.2 ORF1b activities

ORF1b is translated as part of the pp1ab polyprotein with the frameshift event leading to a level of 1b products being about 25% of the domains translated from 1a. Nsp12 clearly contains the RNA-dependent RNA polymerase (RdRp) and is essential in the virus life cycle for the production of negative strand RNA (–RNA), new genomes and subgenomic (sg) messenger RNAs (mRNAs) (Sexton *et al.,* 2016; Velthuis *et al.,* 2010). Nonstructural protein 13 (nsp13) contains an N-terminal zinc-binding domain and a C-terminal superfamily 1 helicase domain which is able to unwind both dsRNA and dsDNA in a 5'-to-3' direction (Lee *et al.,* 2017). Nsp14 has a reported 3'→5'exonuclease domain (ExoN) that is capable of hydrolyzing both ssRNA (single stranded RNA) and dsRNA (double stranded RNA) to final oligoribonucleotide products of ≈8–12 nucleotides length (Becares *et al.,* 2016). Inactivation of Nsp14 in the context of a replicating virus leads to a higher mutation rate consistent with it acting directly or indirectly as part of a proof reading mechanism for coronaviruses (Denison *et al.,* 2011).

Nsp15 encodes an endoribonuclease (Bhardwaj et al., 2012) and forms hexamers (Deng *et al.,* 2017). The endoribonuclease function cleaves RNAs immediately 3' of uridylates and is stimulated by Mn^{2+} (Zhang *et al.,* 2018).

Nsp16 has an S-adenosyl-L-methionine (AdoMet)-dependent RNA (nucleoside-2'O) methyltransferase (2'OMTase) activity which is able to form the cap-1 methylated structures typical of host mRNA (Aouadi *et al.,* 2017). Inactivation of nsp16 leads to enhanced interferon production in infected cells suggesting that the

cap structure acts as a cloaking device to avoid triggering innate receptors (Aouadi *et al.,* 2017) which, in turn, has suggested nsp16 mutants as rationally attenuated strains for vaccine use (Menachery *et al.,* 2018).

The diversity and depth of the *nidovirales* family shows them to be a highly flexible platform enabling rapid adaptation to new hosts. This is consistent with the species jumping observed for new coronaviruses like SARS-CoV and MERS-CoV but it also suggests there may be many more members of the family yet to be found. Indeed, screening programs based on primers to conserved proteins, such as the RdRp and the main protease, have been used to discover new family members in birds (Lau *et al.,* 2018) and bats (Gouilh *et al.,* 2011). With the advent of cheaper and faster sequencing technologies, transcriptomic data has also been used for the identification of novel nidoviruses and this approach is likely to increase with time (Shi *et al.,* 2018). In it, no selection is made at the time of sequence generation, all nucleic acid present in a sample is sequenced - the metagenome. Bioinformatics is then used to discover target virus-like sequences. Thus, transcriptomics has the potential to discover new RNA virus genomes following the sequencing of total intracellular RNA pools.

Previously, the discovery of new RNA viruses preceded based on the appearance of an unexpected disease, antibody cross-reactivity or enough conserved motifs for successful amplification by reverse transcriptase polymerase chain reaction. However, with improvements in RNA transcriptome sequencing and homology-based search methods, complete infecting RNA viromes of an organism can be

discovered by trawls through deep-sequenced total intracellular RNA pools (Miranda *et al.,* 2016; Shi *et al.,* 2018, 2016).

These recent sequencing and bioinformatics methods have brought a great change to the order *Nidovirales* and many studies have revealed several highly divergent nido-like viruses (Lauck *et al.,* 2015; O'dea *et al.,* 2016; Saberi *et al.,* 2018; Shi *et al.,* 2018, 2016; Tokarz *et al.,* 2015; Vasilakis *et al.,* 2014; Wahl-Jensen *et al.,* 2016), which have been proposed, subject to ICTV review, to add four new virus families to the *Nidovirales* (Gorbalenya *et al.,* 2017).

1.6 The Discovery of Abyssovirus

At the outset of the work described in this thesis bioinformatic analysis had been used on a pool of intracellular RNA sequences derived from the deep sea metazoan *Aplysia californica*, the Californian sea hare. This analysis revealed the sequence of what appeared to be a new, highly divergent member of the nidoviruses. Fortuitously, a complete virus sequence could be assembled from the contigs deposited suggesting the likelihood of an extant virus, in addition to which two independent manuscripts described the findings, effectively corroborating each other (Bukhari *et al*., 2018; Debat, 2018). The virus was tentatively named Aplysia Abyssovirus by one of the discovering groups (Bukhari *et al*., 2018).

The etymology of the name is from the word abyss, a reference to the aquatic environment where Aplysia lives, to the Sumerian god of watery depths Abzu, and to its discovery in an RNA transcriptome obtained by “deep” sequencing technology.

The new nido-like virus sequence was identified by tBLASTn analysis of the transcriptome shotgun assembly (TSA) and expressed sequence tag (EST) databases for sequences similar to the main proteinase (M^{Pro}), polymerase and helicase of several nidovirus strains. This led to the discovery of a 35.9 kb transcript and 243 other fragments from Aplysia. Putative virus transcripts were then compared to DNA sequences from the same organisms by nucleotide BLAST and no evidence of either virus was found, supporting the likelihood that the Abyssovirus transcript comes from a novel RNA viruses associated with its host transcriptome.

The Abyssovirus genome is represented in its longest and most complete available form by the transcriptome shotgun assembly sequence GBBW01007738 which represents a reverse-complementary genomic sequence. The Abyssovirus genome organization has several alphavirus-like features (see below) that have not been detected in the nidoviruses before, including a replicase gene interrupted by a stop codon rather than a nidovirus-like ribosomal frameshift signal, and a single structural polyprotein gene. Further differences were present at the junctions of the domains likely cleaved by the M^{Pro}. Nidovirus main proteases generally recognize a glutamine or occasionally a glutamate-containing motif at the cleavage site (Ulferts *et al.,* 2011) but few of these were present, although a possible glutamate-alanine cleavage site occurs near the beginning of the conserved M^{Pro} region, suggesting that an alternate cleavage motif was recognized in Abyssovirus.

A further unusual feature was the presence of an in-frame stop codon separating the pp1a and pp1b genes, rather than the usual ribosomal frameshift signal (see above) suggesting that Abyssovirus might use a translational termination-suppression

signal to control expression of the pp1b region. As the discovery of Abyssovirus was virtual, the question of functionality cannot be addressed by culture of the virus directly. Accordingly, these unusual features, the main protease, the stop readthough sequence and the trypsin-like protease or Structural Protease (S^{Pro}), were noted for experiments designed to test their biological activity.

1.6.1 Main Protease (M^{Pro})

Since its discovery (Rota *et al.,* 2003, Marra *et al.,* 2003), Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) has become a very well characterized member of the *Coronaviridae*. As the proteolytic processing of both nonstructural and structural polyproteins is a vital step for virus replication, the nonstructural protein 5 (nsp5) (main protease (M^{Pro}) also known as 3C-Like protease ($3CL^{Pro}$)) (Anand *et al.,* 2003, Yang *et al.,* 2003) has been extensively analyzed with a view to inhibition of the enzyme for the production of an antiviral compound (Deng *et al.,* 2014). SARS-CoV M^{Pro} employs a Cys-His as a catalytic dyad (Cys145 and His41), similar to the porcine Transmissible Gastroenteritis Virus (TGEV) M^{Pro} and the Human coronavirus 229E M^{Pro} (HCoV 229E) which use Cys144 and His41 (Anand *et al.,* 2003). The structure contains three domains (I, II, and III) and the active site is located between domain I and II. The crystal structures of TGEV, HCoV 229E and SARS-CoV M^{Pro} (Anand *et al.,* 2003, Yang *et al.,* 2003, Anand *et al.,* 2002, Zhu *et al.,* 2017) show common features for CoV M^{Pro} with two chymotrypsin-like β-domains (domain I & II) and one a-helical dimerization domain (domain III) which is lacking from the Picornavirus M^{Pro} enzymes and chymotrypsin (Shi *et al.,* 2004). These structures and biochemical details were used

as guides for the design of mutations in the Abyssovirus M^{Pro} to investigate the structure and function of this enzyme.

1.6.2 Readthrough at the 1a/1b junction

As described above, most nidoviruses use a programmed ribosomal frameshift to connect the 1a and 1b open reading frames (Namy *et al.,* 2004, Stahl *et al.,* 2002, Dinman *et al.,* 1991, Licznar *et al.,* 2003 and Baranov *et al.,* 2002). In this process, the ribosome switches to an alternative frame at a specific site in response to signals in the messenger RNA (mRNA). The frameshifting process has two types, but the most common is a -1 frameshift, in which the ribosome slips a single nucleotide in the upstream direction. Programmed −1 ribosomal frameshifting (−1 PRF) is a mechanism known to regulate the expression of many viral genes and some cellular genes (Wang *et al.,* 2016). Another type of the frameshifting, +1 frameshifts where the ribosome is guided toward a triplet nucleotide by skipping one nucleotide downstream, is also known (Ketteler, 2012) but it is much less common than -1 frameshifts. The -1 frameshift signal contains a slippery sequence and a pseudoknot downstream (Ivanov *et al.,* 2000). These canonical sequences are not found in the Abyssovirus genome suggesting that an alternate mechanism, stop codon readthrough, is used. Protein translation commonly stops at termination codons including UAG, UGA and UAA. However, termination can sometimes be weak and the translation process continues until a later termination codon is reached, producing an extended polypeptide. This event is known as translational readthrough (RT) and occurs in both prokaryotic and eukaryotic systems (Li, G. and Rice, 1993). This feature was also highlighted for test in the studies described here.

1.6.3 Structural Protease (S^{Pro})

Alphaviruses are another enveloped, positive-sense, single-stranded RNA virus that cause many serious human and livestock diseases. The genus Alphavirus includes members Chikungunya Virus (CHIKV), Eastern, Western, and Venezuelan Equine Encephalitis Virus, Aura Virus, Semliki Forest Virus (SFV), and Sindbis Virus (SINV) (Rupp *et al.,* 2015).

The genome of alphaviruses consists of two ORFs, the second of which is expressed through production of a subgenomic mRNA from an internal promoter on the minus-strand RNA replication intermediate (Strauss *et al.,* 1984). The structural proteins are located in the second ORF whereas the non-structural proteins, required for RNA synthesis, occur in the first ORF, which is translated directly from the genomic RNA (Rupp *et al.,* 2015).

The structural polyproteins include capsid protein (CP), E3, E2, 6K, and E1 (Strauss and Strauss, 1994) and unusually the CP has a chymotrypsin-like serine protease activity. The first step in structural polyprotein processing is the autocatalytic cleavage of CP in order to release itself from the rest of the polyprotein (Choi *et al.,* 1991, Melancon and Garoff, 1987), an activity which is much earlier than its role in viral RNA encapsidation and budding (Strauss *et al.,* 1984, Hardy and Strauss, 1989, Melancon and Garoff, 1987).

When identifying Abyssovirus through bioinformatics, the pp2 was shown to contain a chymotrypsin-like serine proteinase (Birktoft and Blow, 1972), a feature similar to the alphavirus capsid proteinase (Melancont and Garoff, 1987). This was termed the structural proteinase (S^{Pro}). The Abyssovirus S^{Pro} was also targeted for studies to investigate its potential function.

1.7 Aims of this research

The Abyssovirus sequence represents a novel addition to the nidovirales but lacks experiential verification of any sort. To provide a level of verification, that is, that some of the novel features described for it are biologically active, genetic constructs were designed using the Abyssovirus sequence to enable the test of biological activity. The hypothesis is that formal proof of the Abyssovirus features as biologically functional would support the case for a novel extant virus circulating in Aplasya. Accordingly, the Abyssovirus main protease (M^{Pro}), structural protease related to alphavirus (S^{Pro}) and the readthrough region at the end of ORF1a were cloned into reporter systems to be expressed with an amino-terminal herpes simplex virus epitope (HSV) tag, and a carboxyl-terminal poly-histidine (HIS) tag. This design allows for any translated products to be detected by western blot for either tag but importantly also allows for any cleavage product to be recognized by western blot for one or other of the epitope tags. To maximize the possibility of biological activity it was also considered beneficial to consider expression in a variety of cell backgrounds. Therefore, a universal expression vector was used with the potential for expression in *Spodoptera frugiperda (Sf9)* insect cells, *E. coli* (*Escherichia coli*) and mammalian cells.

Specifically, to investigate readthrough at the pp1a-pp1b region, the sequence was cloned into pTriEx1.1 with amino-terminal HSV and carboxyl-terminal HIS tags in a construct that would allow detection and quantification of a ~25 kDa proteins that stopped at the natural UGA stop codon and ~32 kDa readthrough products if they occurred. Once established this assay was used further to investigate the nature of the readthrough stop codon, UGA-C or UAA-A, and the role of a 42 nucleotide

sequence downstream of the readthrough site predicted to form an RNA stem-loop structure.

In the case of the main and structural proteases protein expression using the Abyssovirus sequences were assayed for full-length synthesis and also for their presumed catalytic activity. Once established these assays too were modified by mutagenesis of putative catalytic and cleavage site residues to assess their contribution to any activity detected.

The overall goal of this study was to contribute biological data to the discovery of new nido-like viruses and support the contention that additional lineages of nido-like viruses might be present in diverse hosts.

Chapter 2

2. Materials and methods

2.1. Plasmid construction and cloning of desired DNA fragments

2.1.1 Primers

The oligonucleotide primers listed in each chapter and used in this study were designed using Snapgene software using the sequence of Abyssovirus (GenBank accession number GBBW01007738). Primers were purchased from Integrated DNA Technology (IDT). Forward and reverse primers were reconstituted as 50 μM solutions in nuclease-free water and stored at -20°C.

2.1.2 PTriEx1.1 vector map

The pTriEx1.1 vector (Novagen) was kindly provided by Professor Ian Jones. This vector contains mammalian, bacterial and insect promoters upstream of the multiple cloning site. This vector contains an ampicillin resistance gene for positive colony selection, a HSV tag sequence upstream of the cloning site and a 8x HIS tag downstream of the cloning site to enable construction of N-terminal HSV-tagged and/or C-terminal HIS-tagged fusion proteins if desired - **Figure 2.1**. All Abyssovirus proteins in this study were cloned as N- terminal HSV-tagged and C-terminal HIS-tagged fusion proteins by insertion between one or two restriction sites, as described in each chapter.

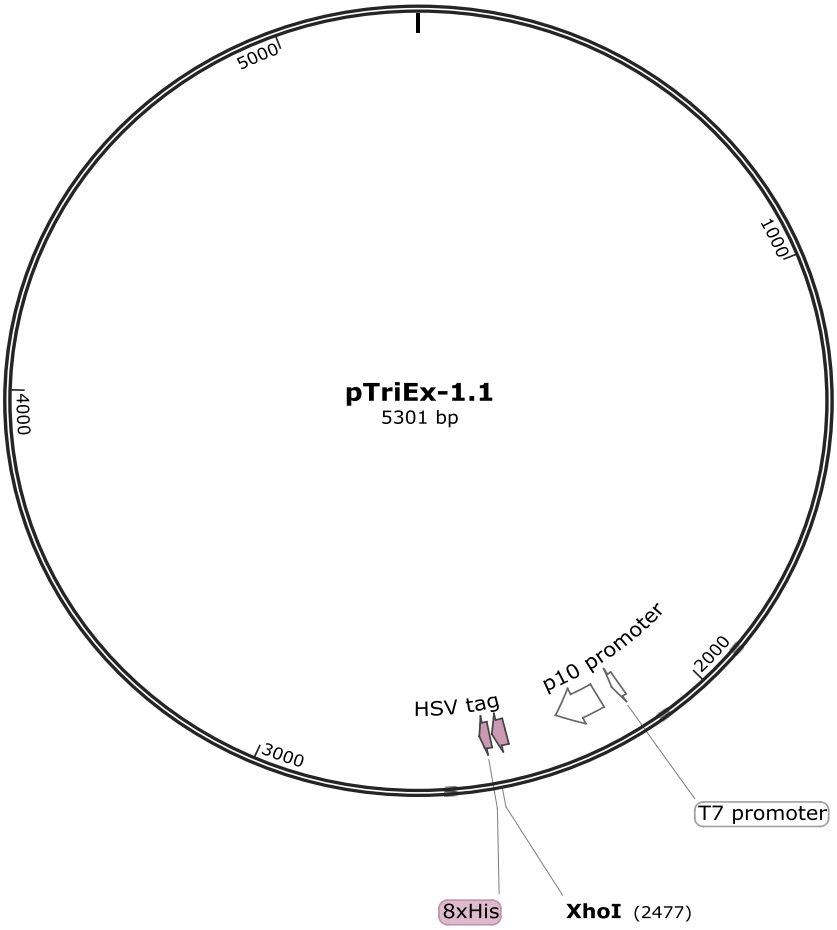

Figure 2.1. Map of pTriEx1.1 vector showing the restriction site (*Xho*I) used for cloning as well as the HSV and HIS tags and the P10 and T7 promoters, which were used for protein expression.

2.1.3 Source and amplification of Abyssovirus DNA fragments

The coding sequences of the wild type Abyssovirus main protease, readthrough sequence and structural protease and some mutants were purchased as synthetic DNA fragments from IDT (see Appendix 3). If necessary they were modified by being amplified by the polymerase chain reaction (PCR) as described in each

chapter. The amplification reaction contained the following components recommended by the manufacturer's protocol: Phusion® High-Fidelity PCR Master mix (BioLabs), 0.3 µM forward and reverse primers, >100 ng DNA template and volume completed to 50 µl with dH_2O. Assembled reactions were placed in a thermal cycler and subjected to the following thermal cycling protocol:

Step	Temperature	Time	No. of cycles
Initial denaturation	95℃	3:00	1X
Denaturation	98℃	0:10	30X
Annealing	53℃	0:10	
Extension	72℃	0:30	
Final extension	72℃	5:00	1X

PCR products were purified using a cleanup kit according to the manufacturer's protocol (Thermo Scientific GeneJET PCR Purification Kit) and eluted in ultra-pure water. The concentrations were determined by ND-1000 Nanodrop spectrophotometer and stored at -20°C.

2.1.4 Double digest and gel extraction of pTriEx1.1 vector

The pTriEx1.1 vector was single or double digested with one or two restriction enzymes (Thermo Fisher SCIENTIFIC) according to the following protocol: 10x Green buffer, template (>150 ng), 1microlitre (10 U/µl) restriction enzyme/s and volume completed to 100 µl, with incubation for 30 min at 37°C. After incubation,

100 µl was subjected to agarose gel electrophoresis, extracted and purified by cleanup kit (Thermo Scientific GeneJET PCR Purification Kit) following the manufacturer's instructions. The linearized vector was used in either infusion or T4 ligase dependent cloning as described.

2.1.5 In-Fusion cloning of fragments into the pTriEx1.1 vector

Different genes were amplified using primers designed for In-Fusion cloning, as listed in each chapter. In a total volume of 10 µl, purified DNA (10-200 ng) was mixed with linearized vector (50-200 ng), 2 µl 5X In-Fusion HD Enzyme Premix (Clontech) and dH_2O following the manufacturer's protocol. The incubation was for 15 min at 50°C.

2.1.6 Generating mutant Abyssovirus genes using overlap PCR

In nearly all mutation cases, the wide type of each DNA was used as a template and amplified by PCR with forward and reverse primers for each mutant sample in reaction PCR1. Another PCR was carried out with the forward primer of each mutant sample and a reverse primer in reaction PCR2. The PCR products were analyzed by agarose electrophoresis and the correct fragments were obtained for both PCR1 and PCR2. The DNA products were purified using a PCR purification kit (Thermo Scientific GeneJET PCR Purification Kit) and quantified by Nanodrop spectrophotometer. Then, 0.1ng/bp from each PCR product were mixed and subjected to a PCR reaction for five cycles without primers. Then, forward and reverse primers were added and another thirty cycles were completed. The correct size of overlap PCR product was analyzed by agarose gel electrophoresis and extracted by gel extraction kit (Thermo Scientific GeneJET PCR Purification Kit).

The purified product was quantified by Nanodrop spectrophotometer. The PCR product was double digested with restriction enzyme/s (Thermo Fisher SCIENTIFIC) for 30 minutes at 37°C and then a clean up kit (Thermo Scientific GeneJET PCR Purification Kit) was used to purify the product. The newly digested product was extracted by gel extraction kit after agarose gel electrophoresis and quantified by Nanodrop spectrophotometer.

2.1.7 Ligation of PCR fragments with pTriEx1.1

Ligation reactions were prepared using an approximate 3:1 molar ratio of the insert to vector with 1x ligation buffer (Thermo Fisher SCIENTIFIC), and 1 unit T4 DNA ligase in a final volume of 10 µl. The reaction was incubated at 22°C for 30 minutes. The ligation mix was then ready for transformation into Stellar competent cells (Clontech).

2.1.8 Transformation of *E. coli* competent cells

For transformation, 50 µl of Stellar competent cells (Clontech) were thawed on ice and 5-50 ng of DNA was added and gently mixed. The DNA mixture was incubated for 30 min on ice followed by a heat shock at 42°C for 45 sec, which opens pores in the cell membrane to allow DNA entry. The transformation reaction was placed back on ice for 1-2 min and 450 µl of pre-warmed SOC (super optimal broth) recovery media (Clontech) was added to allow the cells to recover from heat shock and to express the antibiotic resistance gene. After 1 hr incubation in the

shaker (225 rpm) at 37°C, 100 µl of each transformation reaction were plated onto Luria- Bertani (LB) agar supplemented with 100 µg/ml of ampicillin and incubated at 37°C overnight.

2.2 DNA agarose electrophoresis

Gel electrophoresis was used to visualize and purify DNA fragments. Agarose was dissolved in 1x TAE (Fisher Thermo SCIENTIFIC) buffer by heating in a microwave until boiling. After cooling Gel red (Cambridge Bioscience) was added to 0.01% (v/v). Electrophoresis was carried out at 100v for 60 min. Gels were imaged using a G:BOX Chemi XL imager (Syngene). The approximate size of the DNA bands visualized in the agarose gel was determined using a 1kb Hyper Ladder (Bioline).

2.3 Colony PCR screening of transformant bacteria

2.3.1 Colony PCR analysis

Colonies from transformation reactions were selected randomly and then analyzed by colony PCR. The colony was patched to a master plate and the remainder resuspended in 100 °C of water and heated at 100 °C for 5 min. The sample was then spun at 13,000 rpm in a microfuge and 10 µl was used as template for PCR. The PCR reaction contained the following components recommended by the manufacturer's protocol: Phusion® High-Fidelity PCR Master mix (BioLabs), 0.5

μM forward (TriExUP) and reverse (TriExDown) primers, < 250 ng DNA template and volume completed to 50 μl with dH_2O. The PCR protocol was 1 cycle at 95°C for 2 min, 35 cycles with sequential incubation for 30 sec at 95°C, 30 sec at 52°C and 1 min at 72°C. Finally, an extension cycle for 10 min at 72°C was used to ensure complete extension of all PCR products.

2.3.2 Plasmid DNA purification and sequencing

Colonies identified by colony PCR as containing the correct size insert were selected from LB agar plates and inoculated into LB broth containing 100 μg/ml ampicillin and incubated overnight in the shaker (225 rpm) at 37°C. A Thermo Fisher Scientific Miniprep kit was used for isolation of the plasmid DNA from the 10 ml cultures according to the manufacturer's guidelines for plasmid DNA purification. The sample concentrations were measured using a NanoDrop ND_1000 spectrophotometer. Later, purified plasmid DNA was sent to Source BioScience for sequencing. A cell suspension was used to create a stock for each clone in 50% glycerol and stored at -80°C.

2.4 Transformation of different bacterial strains

2.4.1 Preparing competent cells for the expression system BL21 (DE3)- pLysS

BL21 (DE3)-pLysS cells were streaked on an LB plate for overnight incubation at 37℃. The next day, one colony from the plate was placed in 10ml LB media for overnight growth and 1ml from the overnight BL21 (DE3)-pLysS culture was transferred into 100ml LB media in a 250ml flask to allow cells to grow at 37℃ until OD_{600} =0.5 (~3 hours). Cells were transferred to 50ml cold falcon tubes and placed on ice for 20-30 minutes. The cells were centrifuged at 4℃, 4000rpm for 15 minutes. The media was discarded and cells suspended with 5ml of cold of 0.1M $CaCl_2$ and placed on ice for 20 minutes. Cells were centrifuged for 15 minutes at 4℃ at 4000rpm, the supernatant poured off and the cells re-suspended in 1ml cold 0.1M $CaCl_2$ containing 10% glycerol. 100µl aliquots were transferred into cold (1.5 ml) Eppendorf and frozen in liquid nitrogen. Cells stored at -80℃.

2.4.2 Other *E. coli* strains

Plasmids were transferred into different *E. coli* strains including BL21 (DE3)-pLysS, Origami, Rosetta and Tuner (Invitrogen) (**Table 2.1**) by transformation. 0.1 µg plasmid DNA was mixed gently with 50 µl competent cells, incubated on ice for 30 min, heat shocked for 45 sec at 42°C water bath, replaced on ice for 1-2 min and then 450 µl of pre-warmed SOC medium was added to the plasmid-cell mixture and incubated in the shaker at 37°C for 1hour (225 rpm). After incubation, 100 µl from

the total mixture was plated on LB plates containing 100 μg/ml ampicillin and incubated overnight at 37°C.

Table 2.1. *E.coli* strains used for plasmid transformation and amplification and protein expression.

Strain/Name	Genotype	Source
Stellar competent cells	*F-, endA1, supE44, thi-1, recA1, relA1, gyrA96, phoA, Φ80d lacZΔ M15, Δ(lacZYA-argF) U169, Δ(mrr-hsdRMS-mcrBC), ΔmcrA, λ-*	Clontech
BL21 *(*DE3*)* pLysS	*F– ompT hsdS$_B$(r$_B$ – m$_B$ –) gal dcm (DE3) pLysS (Cam^R)*	Novagen
Origami	*Δ(ara–leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsL F'[lac+ lacI)pro] gor522::Tn10 trxB (Kan^R , Str^R , Tet^R)*	Novagen
Rosetta	*F– ompT hsdS$_B$(r$_B$ – m$_B$ –) gal dcm pRARE^2 (Cam^R)*	Novagen
Tuner	*F–ompT hsdS$_B$(r$_B$–m$_B$–) gal dcm lacY1*	Novagen

2.5 Induction of protein expression in *E. coli* T7 compliant strains

An overnight culture was inoculated (200 µl) into a 10 ml fresh LB broth with ampicillin (100 µg/ml) and incubated in the shaker at 37°C for 1-3 hr. The expression of Abyssovirus proteins was induced when cultures had reached an OD_{600} of 0.6 by adding IPTG (isopropyl-β- D-thiogalactopyranoside) to a final concentration of 1 mM. For determining optimal expression, 1 ml was taken at time points of 30min, 1, 2, 3 and 4 hr, centrifuged and the cell pellets were subjected to western blotting.

2.6 Transfection of Mammalian cells (17clone-1)

Transfection is a method used to introduce nucleic acids into cells in culture. 17-clone1 cells were cultured in Dulbecco's modified Eagle medium (DMEM) (Sigma Aldrich) supplemented with 10% fetal bovine serum (FBS) (GE Healthcare), and antibiotics penicillin/streptomycin (penicillin 100 U/ml, streptomycin 0.1 mg/ml; Gibco/Invitrogen) in a 6 well plate and transfected with plasmid DNA for the expression of viral proteins by using different transfecting reagents to compare their efficiency, Lipofectin and Lipofectamine3000 (Invitrogen), following the manufacturer's instructions for each reagent. After incubation for 24 hour, cell culture plates were placed on ice and a plastic cell scraper used to scrape adherent cells from each well gently. The suspensions were transferred into pre-cooled tubes, centrifuged for 4 min (2000 rpm) at 4°C. Finally, the supernatants were aspirated and the pellets re-suspended with 50 µl ice-cold phosphate-buffered saline (PBS) and stored at - 20°C.

2.7 Transfection of insect cells (Sf9)

The flashBAC™ GOLD (FBG) baculovirus Expression System (Mirus Bio) was used to produce recombinant baculovirus using transfer vector plasmids and transfection to produce recombinant baculoviruses. A cell density of 1 x 10^6 Sf9 insect cells was seeded in a 6 well tissue culture plate and the cells were left to settle for 20 minutes. Cells then washed twice with 2ml fresh media and a transfection mixture applied dropwise to the well. The transfection mixture consisted of 2µl of the recombinant plasmid (100ng/µl) mixed with 2µl of the FBG and 8µl of nuclease free water in one sterile Eppendorf tube. In a second Eppendorf tube 8µl of lipofectin transfection reagent was mixed with 4µl of nuclease free water. Then the two tubes (12µl each) were mixed and incubated for 15 min at room temperature. Following transfection, cells were incubated at 27 C° for 5-6 days and examined under microscopy to visualize the cytopathic effect caused by the recombinant virus. Cells were harvested and centrifuged at 2000 rpm/ 5min and the supernatant was collected as passage 0 (P0) and stored at 4 C°. Later a six well tissue culture plate was seeded with 1 x 10^6 Sf9 cells to obtain passage 1 by using 0.1ml of P0 and incubated at 27℃ for 3-4 days. This process was repeated to obtain passages 2 and 3 which had increasingly rapid cpe reflecting a higher titer of virus.

2.8 Protein purification from insect cells

Sf9 culture with a cell density of 1.5x10^6 /ml in a total volume of 300ml with 2% FCS media were infected with 10% v/v of recombinant virus stock giving an estimated MOI of 1 and incubated at 27 ℃ for 3 days in a rotary shaker. Before harvest the number of dead cells was estimated using 0.4% trypan blue exclusion

using the Countess® automated cell counter. The cells were harvested by centrifugation at 4000rpm for 20 min at 4 ℃. The cell pellets were collected on ice and lysed using lysis buffer (50 mM PBS, 500mM NaCl, 1% NP-40, pH 7.4) and a cocktail of protein inhibitor (Roche) was added to the lysate. Then sonication was performed on ice for 10 min / 20 seconds on/off pulse and 80% amplitude using a probe sonicator (Sonics). The cell lysate was centrifuged 15,000 rpm / 20 minutes/ 4 C˚ using a Sorval RC5B centrifuge. An integrated low pressure purification apparatus (BioRad) was used to purify the protein of interest. A 5ml IMAC column (GE Lifesciences) was washed with binding buffer (20 mM PBS, 500mM NaCl, 20 mM Imidazole, 0.1% Tween-20, pH 7.4) at 5ml/min after which the cell lysate was loaded into the column. HIS-tagged proteins were eluted using a gradient of 0-100% elution buffer (20 mM PBS, 500mM NaCl, 500 mM Imidazole, 0.1% Tween-20, pH 7.4) at a flow rate of 5ml/ min. The eluted fractions containing the protein of interest were collected separately and analyzed by SDS=PAGE and by western blotting. The fractions containing the protein of interest were pooled and concentrated using the viva spin column 10K MWCO (Sartorius) by spinning at 4000 rpm for 30 minutes at 4 C˚.

2.9 Cell culture

2.9.1 17clone-1 mouse fibroblast cells

Cells were grown at 37°C in Dulbecco's modified Eagle medium (DMEM; Gibco/Invitrogen) supplemented with 10% heat-inactivated fetal bovine serum (FBS) and penicillin/streptomycin (penicillin 100 U/ml, streptomycin 0.1 mg/ml;

Gibco/Invitrogen). For maintenance, cells were split once they reached ~90% confluence. The media was removed and the cells were washed twice with sterile 1x PBS. The cells were detached from the plastic using 0.2% Trypsin-EDTA (PAA; Sigma). The cells were then suspended in fresh DMEM media with serum to inactivate the trypsin. An aliquot of the cell suspension was added to a new flask containing fresh DMEM media and incubated at 37°C in 5% CO_2 as before.

2.9.2 Sf9 insect cells

Sf9 insect cell lines were cultured in EX-CELL® 420 Serum-Free Medium supplemented with 3% fetal bovine serum (FBS) and penicillin/streptomycin (penicillin 100 U/ml, streptomycin 0.1 mg/ml; Gibco/Invitrogen). Cells were split once they reached ~90% confluency. Following cell counting a new culture was inoculated in a plastic disposable flask with 0.3 x 10^6 cells /ml and incubated at 27℃ with shaking for 3 days until confluency again reached 90%.

2.10 Protein analysis

2.10.1 SDS Polyacrylamide gel electrophoresis (PAGE)

Proteins were analyzed by SDS-PAGE. Cells were resuspended in 20 µl chilled PBS and mixed with 20 µl of LDS sample buffer (980 µl 4x LDS sample buffer (Novex) and 20 µl β-mercaptoethanol) and incubated at 100°C for 10 min before analysis on a 4-12% SDS-PAGE gel (Life Technologies) assembled in a dual mini gel tank (Life Technologies). Samples were centrifuged for 2 min and 10 µl of each was loaded per well. Samples were separated at a constant voltage of 170v for 30 min in 1x MES (50 mM MES, 50 mM Tris, 0.1% SDS, 1% EDTA, pH 7.25)

running buffer (Life Technologies). A sample (10 µl) of Sharp pre-stained protein standard or See Blue Plus2 Prestained Standard (Life Technologies) was used for molecular size estimation. Following electrophoresis, the gel was stained or transferred onto PVDF membranes for Western blotting analysis.

2.10.2 Protein staining with Coomassie brilliant blue

Coomassie blue dye was used to visualize proteins separated by SDS-PAGE. After electrophoresis, the gel was incubated at room temperature on a rocking platform (25 rpm) in a solution of Coomassie brilliant blue stain (0.025% Coomassie Brillant Blue R-250, 45% methanol and 10% glacial acetic acid) for 30 min. The gel was then washed with a destaining solution (10% methanol, 10% glacial acetic acid) at room temperature on the rocker for 10 minutes. The washing step was repeated three times and finally overnight.

2.10.3 Western blot

Following electrophoresis, SDS-PAGE gels were incubated in transfer buffer (400 ml 1.5x Tris/glycine (Tris base 25 mM, Glycine 190 mM pH 8.3) buffer, 20% methanol for 3 min. The SDS-PAGE gels were then transferred onto PVDF membranes (Merck) by being electroblotted for 1 hour and 20 minutes/ at 150mA in transfer buffer using a semi-dry Western blotting apparatus (ATTO). Following transfer, the membranes were incubated in blocking buffer (5% non-fat milk in TBST buffer (500 ml 1xTBS (50 mM Tris-Cl, 150 mM NaCl, pH 7.6) and 2% Tween20 overnight. The next day, membranes were washed with TBST buffer 3 times for 5 min on a rocking platform (25-30 rpm) (Stuart). The primary HSV and 6x HIS tag antibodies (Abcam) were diluted 1/10,000 in blocking buffer, and added to the membrane in a plastic pouch.

Membranes were incubated for 1 hr at room temperature on a rocking platform (25-30 rpm). After incubation, membranes were washed with TBST buffer 3 times for 5 min. Membranes were then incubated for 1 hr at RT with the secondary antibody (polyclonal Goat Anti-mouse Immunoglobulins /HRP) dissolved in 1% TBST-milk on rocking platform (25-30 rpm), followed by 3 washing steps in TBST as previously described. The protein bands were detected using equal volumes of ECL reagents A (0.25 ml) and B (0.25 ml) enhanced chemiluminescent substrate (GE Lifesciences) and luminescence from the HSV and HIS tag labeled proteins was visualized using a G: BOX Chemi XL (Syngene).

Chapter 3

3. Cloning and expression of Abyssovirus main protease fragments in different expression systems.

3.1 Introduction

In a typical and well characterized coronavirus such as Mouse Hepatitis Virus (MHV) the ORF 1a portion of gene 1 encodes a 3C-like proteinase ($3CL^{Pro}$), nonstructural protein 5, also called the main protease (M^{Pro}), which has been experimentally confirmed to cleave at least 11 sites in the pp1a and pp1b polyproteins (**Figure 3.1**), 10 of which have a dipeptide consisting of Gln at position 1 (P1) and Ser, Asp, Gly, or Cys at P1' (the amino acids flanking the protease cleavage sites are labeled from the N to the C terminus as follows: –P3–P2–P1 ↓ P1'–P2'–P3'–) (Baker *et al*., 1993, Denison *et al*., 1992, Lee *et al*., 1991).

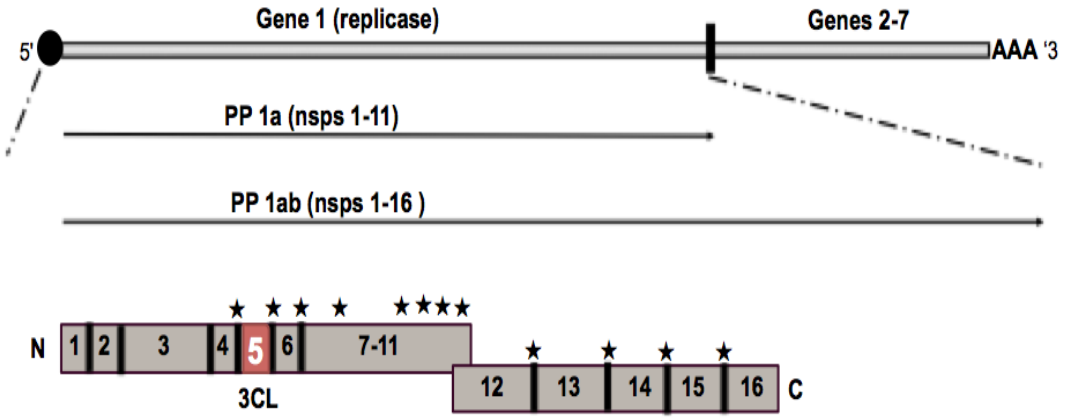

Figure 3.1. The genome of Murine hepatitis virus (MHV) as a typical coronavirus with nsp5 location shown. The genome encodes 7 genes in all. Gene 1, the replicase gene encodes 16 non-structural proteins that are processed by viral proteases including the nsp5 (*aka* 3CL or Main) protease. Nsp5 protease is responsible for 11 cleavage events between nsp 4 and 16 as marked by asterisks (adapted from Stobart *et al*., 2013).

Proteolytic enzymes are widespread and generally have a positive or negative effect on biological processes including protein degradation or controlling physiological events (Neurath, 1984) and thanks to numerous studies a wealth of information is available. The role of coronavirus main protease is clearly degradative, releasing the final replicase component proteins from pp1a and pp1ab.

Proteolytic enzymes have been classified into five classes based on their catalytic mechanism: aspartic, cysteine, metallo-, threonine and serine peptidases (Barrett *et al*., 2003) and over one third of all proteolytic enzymes are serine peptidases. The family name stems from the nucleophilic Ser amino acid residue at the active site of the enzyme which attacks the carbonyl moiety of the substrate peptide bond to form an acyl-enzyme intermediate (Hedstrom, 2002). Serine proteases have been well

studied and are classified into four groups including chymotrypsin, trypsin, elastase, and subtilisin. Based on the structure, Coronavirus main protease belong to the cysteine proteinases and feature a serine proteinase (chymotrypsin)-like two-βbarrel fold, but with a Cys–His catalytic dyad instead of the classical Ser–His–Asp triad (Thiel *et al*., 2003).

All coronaviruses have an equivalent main protease, several of which have been characterized in detail, and the reported three dimensional structures of those coronavirus main proteases that have been solved are similar (Huang *et al*., 2004). Crystal structures have been reported for the 3CL (or M^{Pro}) proteins of Transmissible Gastroenteritis Virus (TGEV), Human coronavirus 229E (HCoV 229E), SARS-CoV and HKU1 (Anand *et al*., 2002, Anand *et al*., 2003 and Yang *et al*., 2003, Zhao *et al*., 2008). The common structure of these proteins consists of two chymotrypsin-like anti-parallel beta barrels, with a third domain consisting of an arrangement of alpha-helices which constitutes a dimerization domain. The His and Cys residues located between the first two domains have been reported as the catalytic dyad and mutagenesis of these residues in SARS-CoV, Hokovirus (HoV), Avian Infectious Bronchitis (IBV) and MHV M^{Pro} have proved them essential for protease activity (Liu & Brown, 1995; Lu *et al*., 1995; Ziebuhr *et al*., 1995, 1997; Tibbles *et al*., 1996; Seybert *et al*., 1997, Shi *et al*., 2004, Fan *et al*., 2004).

Unlike classical 3C proteases however the substrate specificity can be different. Whereas in the coronaviruses the M^{Pro} cleavage site is canonical as described above, that is Q followed by S, A, G or C, in the Arteriviruses the cleavage site is E followed by S, A, G or sometimes N.

The conservation of coronavirus main proteases is such that it makes an ideal probe for the discovery of distantly related virus family members. Indeed its use to probe

the TSA and EST databases using tBLASTn (Altschul *et al.,* 1990) was used for the discovery, in the transcriptome of the California sea hare, *Aplysia californica,* of the *Abyssovirus* described here (Bukhari *et al*., 2018). A second distant related sequence was found in the ornate chorus frog *Microhyla fissipes* (Bukhari *et al*., 2018). Although the Abyssovirus M^{Pro} was identified strongly by homology with the more well characterized coronavirus M^{Pro} proteins (93% probability by HHPrep) and was located at an appropriate position in the translated pp1a (residues 4401-4618) it was nonetheless very divergent (**Figure 3.2**) and, given the nature of the sequences deposited in the TSA database, there was no functional data to support its identification. Moreover, when identifying viral proteins through bioinformatics there can be no guarantee that the putative function has not been falsely assigned from sequence mis-assembly, sequence errors or other artifacts of the sequence assembly processes. It was therefore unclear from sequence analysis alone if enzyme activity was to be expected or not.

Figure 3.2. HHprep output showing the identification of the Abyssovirus M^{Pro} sequence within the submitted translated 1a region of the TSA sequence GBBW01007739.1 as a member of the main coronavirus polyprotein processing enzymes. Maximum homology, as shown, was with the M^{Pro} of HKU1 coronavirus. The nomenclature used is standard, H for helix, C for Coil and E for beta sheet.

The exact boundaries of the protein are unclear, however Q/S and E/X codons do occur suggesting the use of either the coronavirus Glutamine (Q) or Glutamic acid (E) in the P1 position. A further possibility is that, as has been suggested in MHV, the processing of pp1a may involve M^{Pro} cleavage at a dibasic site (RK) which lies upstream of the preferred M^{Pro} Q/S substrate (Lee *et al*., 1991). A dibasic site also occurs in the Abyssovirus sequence (**Figure 3.3**) allowing for the possibility that this too could be recognized and cleaved. Whether any of these suggested substrates are *bona fide* processing sites was addressed in this chapter. Overall, the purpose of the chapter was to generate constructs for the expression of the Abyssovirus M^{Pro} in a system that could allow detection of the cleavage products. Ideally, the potential protease activity would be tested in different cell backgrounds to ensure that a suitable cellular environment for activity was found.

3.2 Results

3.2.1 Design and construction of tagged Abyssovirus main protease

The Abyssovirus main protease sequence, as defined in the TSA database sequence obtained from the sea hare *Aplasya californica* (accession no: GBBW01007739.1) was ordered as a synthetic DNA fragment of 1251bp and was provided by Dr. Geraldine Mulley. The annotated amino acid sequence of the fragment is shown in **Figure 3.3**.

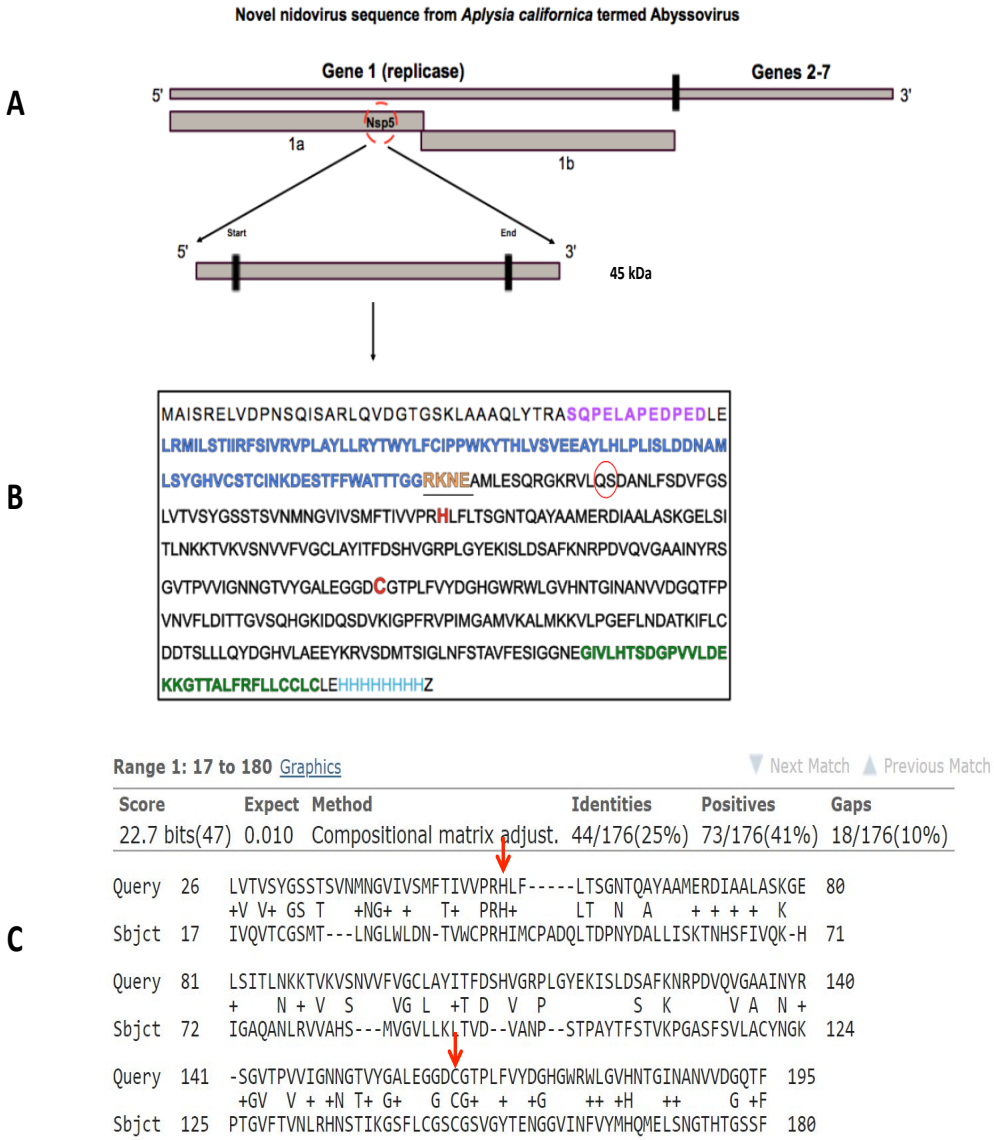

Figure 3.3 Location of nsp5 protease reading frame on the Abyssovirus genome and choice of mutation. **A:** A cartoon of the nsp5 (M^{Pro}) sequence which is located towards the 3' end of the predicted 1a translation product of gene 1. **B:** The Abyssovirus main protease (M^{Pro}) construct, designed as described in the text, has the following order: HSV tag – pre-protease sequence (blue) - Main proteinase (black) – post protease sequence (green) – HIS tag. The protease is autocatalytic and the residues at the possible cleavage junctions are also highlighted (underlined or circled). **C:** Identification of the catalytic dyad residues shown in Red identified by alignment with a well characterized HKU4-CoV main protease (John *et al.,* 2015). The arrowed conserved His and Cys residues are those in the catalytic site mutagenized as part of this study.

To provide tags for the facile detection of the translated product the synthetic fragment was cloned by in-Fusion technology into the vector pTriEx1.1 which is able to express proteins in a number of different phyla by virtue of the multiple promoters upstream of the inserted sequence. These include the T7 promoter for expression in *E. coli,* the p10 promoter for expression in the baculovirus/insect cell system and the hybrid CAG promoter for expression in mammalian cells (Jarvis, 2009). The in-fusion methodology allows directional cloning of one or more DNA fragments into any vector with high cloning efficiency. The strategy adds specific 15 bp overlap sequences to the termini of the DNA fragment to be cloned which are homologous to specific sequences in the linearized vector. When mixed together these overlap sequences recombine, catalyzed by enzymes able to carry out sequence specific recombination *in vitro* (Jacobus & Gross, 2015). To enable the cloning, the vector was linearized with restriction enzyme *Xho*I and mixed with the synthetic DNA fragment flanked with sequences that, once recombined, would allow expression of M^{Pro} appended to an N-terminal HSV and a C-terminal HIS tag, both already present in the pTriEx1.1 vector. The 15 bp overlap sequences were added to the synthetic DNA by PCR amplification and are shown with vector and M^{Pro} homologies identified (**Table 3.1**).

Table 3.1. Oligonucleotides used for in-Fusion cloning of wild Abyssovirus main protease in pTriEx1.1 plasmid. Yellow color represents vector sequence.

Primer	Sequence	bp
Fw_Main protease_Infu	5'CCCCGAGGATCTCGAGTTGCGAATGATTTTGTCTA CC ' 3	37
Rv_main protease_ Infu	5'GATGGTGGTGCTCGAGACACAGACAACACAACAA AAA '3	37

After transformation of the infusion mix into Stellar *E. coli* competent cells and overnight incubation on selective agar, 6 colonies were selected at random, transferred onto LB with ampicillin and grown overnight at 37°C with shaking. Plasmid DNA was extracted from the 6 candidates and analyzed by PCR using the primers used for the generation of the infusion fragment (**Table 3.1**). DNA which amplified to give the correct size of insert was digested with restriction enzyme *Xho*I to confirm its presence and the positives sent to Source BioScience for DNA sequencing to confirm the identity of the cloned sequence. All sequences obtained were as designed with no evidence of mutations and the reading frame was open through the sequence and encoded each of the elements described above. The predicted molecular weight of the complete translated product by use of the Compute pI/MW tool at Expasy (https://web.expasy.org/compute_pi/) is 46.4kDa and the predicted cleavage products are ~13kDa for the HSV tagged N terminal fragment and ~6 kDa for the HIS tagged C-terminal fragment. The assembled plasmid map with key features annotated is shown in **Figure 3.4**.

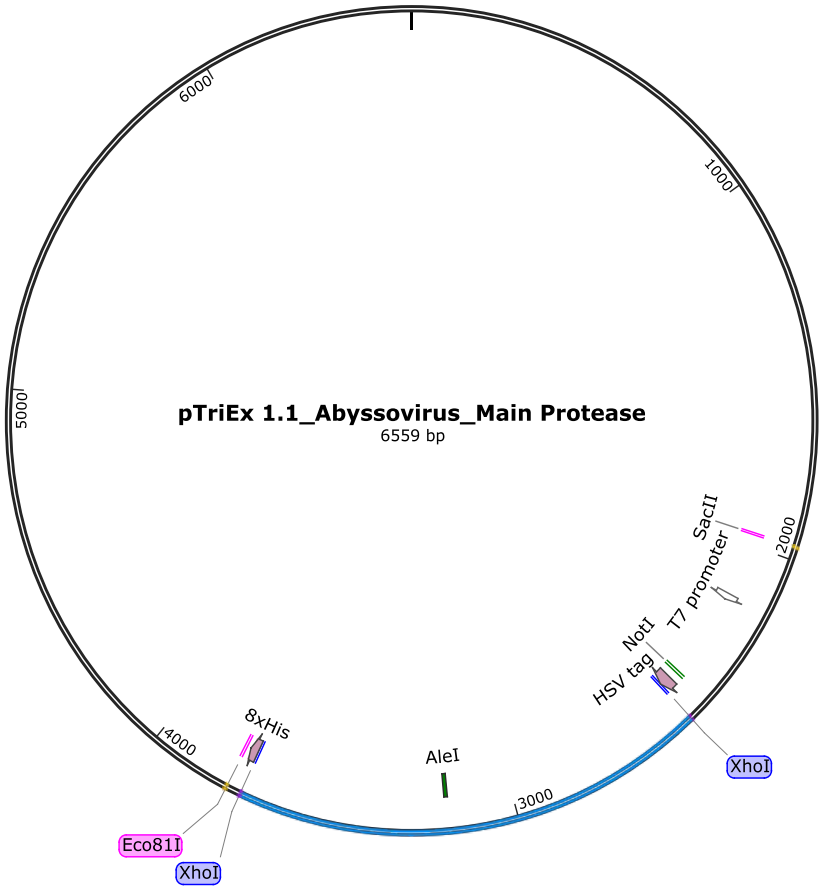

Figure 3.4. Plasmid map showing the complete Abyssovirus M^{Pro} sequence assembled into the pTriEx1.1 vector. Note the *Xho*I sites flank the wild type main protease. The *Sac*II *and Eco*81I restrictions sites and *Not*I and *Ale*I sites used for cloning the mutated proteases as well as the sequences encoding the HSV and HIS tags used for protein detection are also shown.

3.2.1.1 Generation of mutants in the Abyssovirus main proteases by overlap PCR

In SARS-CoV M^{Pro}, His41 and Cys145 were identified as essential for catalysis (Huang *et al*., 2004). The equivalent residues in the Abyssovirus main protease were identified by alignment (**Figure 3.2**) and targeted for mutagenesis to test the role of this catalytic dyad in any activity of the Abyssovirus main protease. To determine their relative impact each of the active site residues was mutated to Alanine, the active site 1 mutation was Cys to Ala at position 302 and the active site 2 mutation a combined mutant His to Ala at position 193 in addition to active site 1. To accommodate the several possibilities for the N terminal cleavage junction, as discussed above, residues predicted as essential for the auto-cleavage reaction, all located between the HSV-tag and the N-terminus of the M^{Pro} sequence, were mutated at each position. Four of the resulting mutants were made singly by changing the resident amino acid to Alanine using the overlap PCR method (**Figure 3.5**) using designing additional forward (Fw_) and reverse (Rv_) primers as shown in **Table 3.2**. A later mutant at the possible Q/S site was made by gene synthesis to include the change of Q to A at position 153 and the mutant fragment exchanged for the resident wild type fragment by use of unique flanking restriction sites *Not*I and *Ale*I.

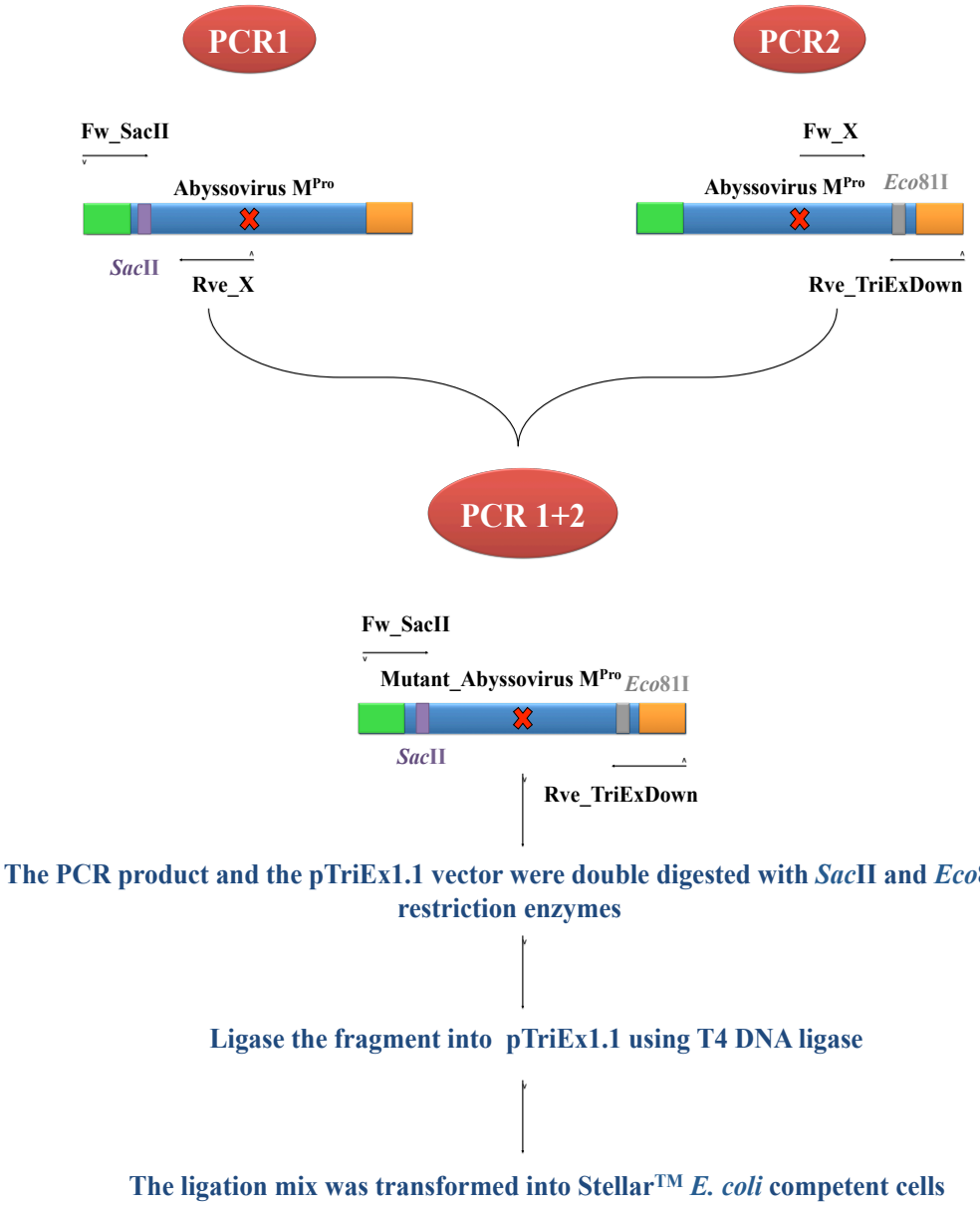

Figure 3.5. Generation of mutant Abyssovirus main protease fragments by overlap extension PCR. The final amplified DNA fragments were digested and inserted between the *Sac*II *and Eco*81I restrictions sites in the pTriEx1.1 vector.

Table 3.2. Oligonucleotides used for cloning in pTriEx1.1 plasmid. Highlighted sequence represents mutation sites.

Primer	Sample	Mutant	Sequence	bp
Fw_A1	First active site (A1)	TG→GC	5'-GCTTTAGAGGGTGGCGATGCTGGTACGCCGTTGT-3'	34
Rve_A1			5'-ACAACGGCGTACCAGCATCGCCACCCTCTAAAGC-3'	34
Fw_A2	Second active site (A2)	CA→GC	5'-CACTATTGTTGTTCCACGTGCTCTGTTTTTAACGTCTGGC-'3	40
Rve_A2			5'-GCCAGACGTTAAAAACAGAGCACGTGGAACAACAATAGTG-'3	40
Fw_1	First cleavage site (1)	CG→GC	5'GGGCAACTACAACGGGGGGCGCTAAGAATGAAGC'3	34
Rve_1			5'GCTTCATTCTTAGCGCCCCCCGTTGTAGTTGCCC'3	34
Fw_2	Second cleavage site (2)	AA→GC	5'CAACGGGGGGCCGTGCGAATGAAGCAATGC'3	30
Rve_2			5'GCATTGCTTCATTCGCACGGCCCCCCGTTG'3	30
Fw_3	Third cleavage site (3)	AA→GC	5'CGGGGGGCCGTAAGGCTGAAGCAATGCTTG'3	30
Rve_3			5'CAAGCATTGCTTCAGCCTTACGGCCCCCCG'3	30
Fw_4	Forth cleavage site (4)	GA→GC	5'GCCGTAAGAATGCAGCAATGCTTGAATCTCAGC'3	33
Rve_4			5'GCTGAGATTCAAGCATTGCTGCATTCTTACGGC'3	33

For the mutations made by overlap PCR, the wild type main protease DNA, cloned as described above, was used as template and amplification by PCR used a forward primer (Fw_SacII, **Table 3.3**) which primed 5' to the M^{Pro} sequence and a reverse primer which included the mutation to be introduced in each case (Rv_X see **Table 3.2**) in reaction PCR1. A second PCR reaction was carried out with a forward primer which included the mutation to be introduced (Fw_X see **Table 3.2**) and a reverse primer (Rve_ TriExDown, **Table 3.3**) which hybridized to a sequence 3' to the cloned M^{Pro} sequence, reaction PCR2.

The PCR products of these first round PCR reactions for the two active site and the four cleavage sites mutations were analyzed by agarose electrophoresis, and the correct sizes (PCR1: 1344, 1017, 846, 849, 852 & 855bp and PCR2: 617, 944, 1115, 1112, 1109 and 1106bp) were obtained for both PCR1 and PCR2 respectively **Figures 3.6, 3.7 and 3.8.**

Table 3.3. Flanking oligonucleotides used to generate mutant main protease in all cases.

Primer	**Sequence**	**bp**
Fw_SacII	5'- TGGCTGCGTGAAAGCCTTG-'3	19
Rve_TriExDown	5'-TCGATCTCAGTGGTATTTGTG-'3	21

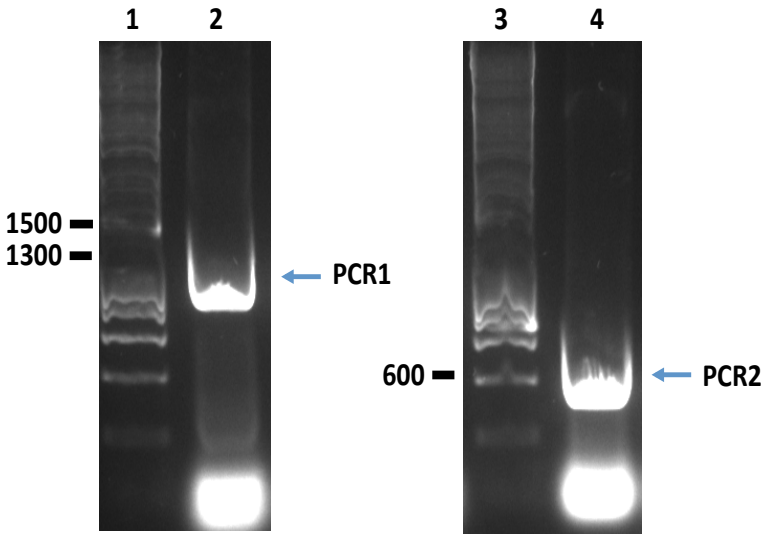

Figure 3.6. Gel electrophoresis of PCR products following amplification of Abyssovirus M^{Pro} first active site mutation using specific primers. Lane 1 and 3: Marker (1kb) DNA ladder, Lane 2: PCR1 (1344bp), Lane 4: PCR2 (617bp).

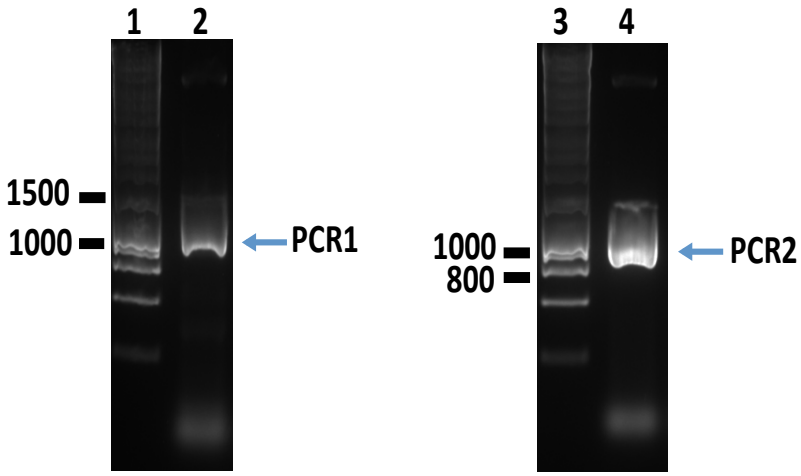

Figure 3.7. Gel electrophoresis of PCR products following amplification of the Abyssovirus M^{Pro} second active site mutation using specific primers. Lane 1 and 3: Marker (1kb) DNA ladder, Lane 2: PCR1 (1017bp), Lane 4: PCR2 (944bp).

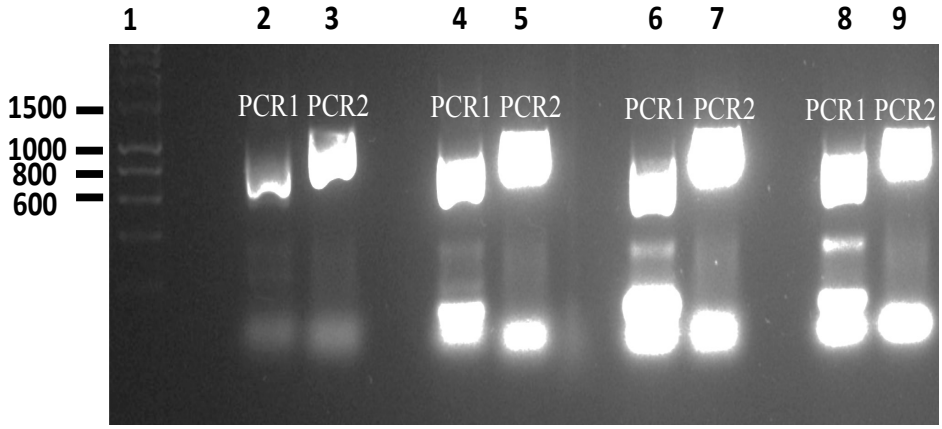

Figure 3.8. Gel electrophoresis of PCR products following amplification of four individual cleavage site mutations using mutation-specific primers. Marker (1kb) DNA ladder in the left side, Lanes 2, 4, 6 & 8: PCR1 (846, 849, 852 & 855bp), Lane 3, 5, 7 & 9: PCR2 (1115, 1112, 1109 and 1106bp) for the four mutant cleavage sites respectively.

The amplified products were purified by gel elution kit and quantified by Nanodrop spectrophotometer. Then 0.1ng/bp of each PCR product were mixed and subjected to amplification in a PCR reaction for five cycles without any primer addition. The forward primer (Fw_SacII) and the reverse primer (Rve_TriExDown) were then added and a further thirty cycles were completed. The presence of the correctly sized 1958bp overlap PCR product was assessed by agarose gel electrophoresis as before and the desired band extracted by gel extraction kit **Figure 3.9.**

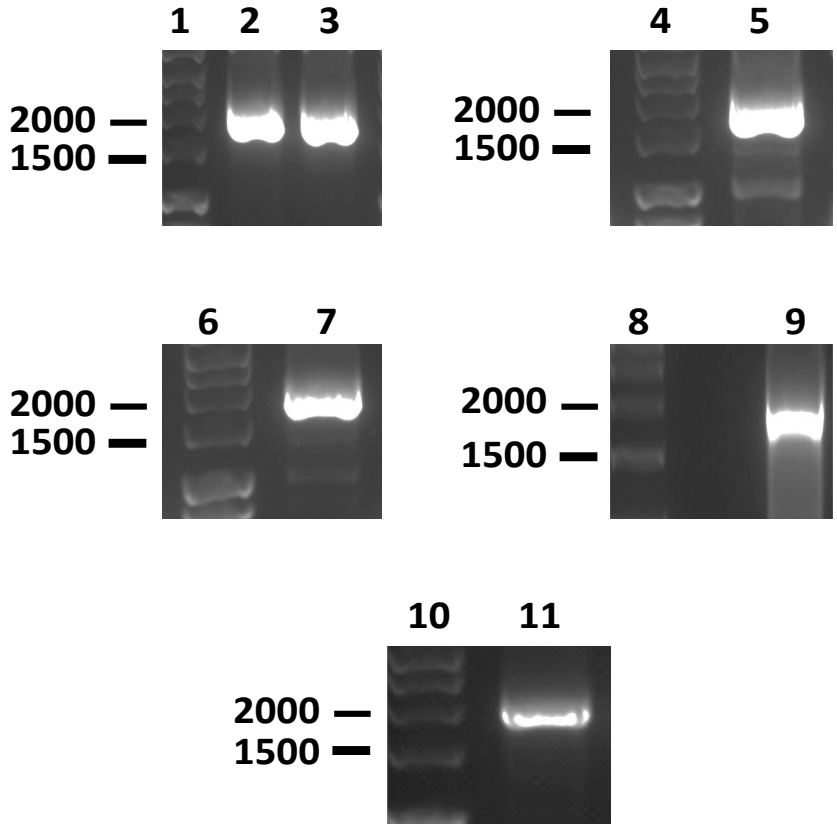

Figure 3.9. Agarose gel electrophoresis of overlap product of Abyssovirus two active sites (lanes 2&3) and four-cleavage sites mutants (lanes 5, 7, 9 & 11) amplified by PCR. Lanes 1, 4, 6, 8, 10: Marker (1kb) DNA ladder. Lanes 2, 3, 5, 7, 9, and 11: PCR1+PCR2 overlap product for all Abyssovirus main protease mutations (1958bp).

The purified product was quantified by Nanodrop spectrophotometer and double digested with *Sac*II and *Eco*81I (Bsu36I) restriction enzymes for 30 minutes at 37°C and the digested DNA recovered by use of a clean up kit. Unmodified pTriEx1.1 plasmid was also digested with *Sac*II and *Eco*81I restriction enzymes (for site locations see **Figure 3.4**) and the large vector band purified by gel extraction kit after agarose gel electrophoresis and similarly quantified by

Nanodrop spectrophotometer. The linearized vector was mixed with each overlap mutant fragment and a T4 ligation kit was used for their ligation according to the manufacturer's instruction. The ligation mix was transformed into Stellar^{TM} *E. coli* competent cells and a separate transformation with a small amount of non-digested plasmid acted as a transformation control.

3.2.2 Colony PCR screening of transformed bacteria

Colony PCR was performed using forward and reverse primers **(Table 3.3)** to check for the presence of the fragment from each of the main protease mutation cloning experiments and the positive colonies with bands of the expected size were selected **Figure 3.10.**

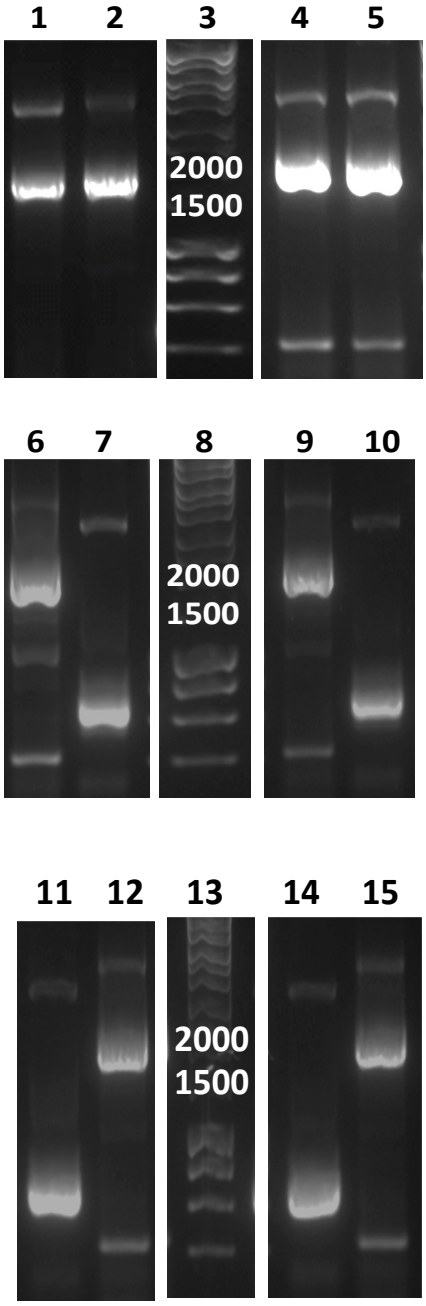

Figure 3.10. Main protease mutations screened by colony PCR. Lane 3, 8 and 13: Marker (1kb) DNA ladder, Lanes 1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 14 and 15: two colonies of first active site mutation, second active site mutation and the four cleavage sites mutations respectively were screened for the presence of pTriEx-M^{Pro} mutant insert (1958bp).

Colonies with the correct size of insert by colony PCR were grown at 37°C overnight in 10 ml LB broth with ampicillin and plasmid DNA extracted. The isolated DNA was sent for sequencing to confirm the relevant mutations in all cases and, once confirmed, a glycerol stock was prepared and the new strains stored at -80°C.

As a result of the cloning experiments, plasmids capable of the expression of wild type Abyssovirus main protease (M^{Pro}) and seven mutants, which change residues predicted to be at the active site or at a possible cleavage junction within the chosen sequence, were confirmed. In all cases the main protease sequence was flanked by epitopes for widely available antibodies, recognizing the HSV and HIS tags, so that the complete translation product (both epitopes) and cleavage products representing either end of the fragment (one or other epitope) could be detected by western blot. These plasmids represented the resource for an investigation of possible enzyme activity in the Abyssovirus M^{Pro} protein.

3.2.3 Expression and proteolytic activity of Abyssovirus main protease

As noted, the base vector in use, pTriEx1.1, allows the expression of target proteins in multiple expression systems (*E. coli*, insect and vertebrate). Expression in *E. coli* strains, including *E. coli* BL21 (DE3)- pLysS, Origami, Rosetta and Tuner, is enabled by these strains having been engineered to express the T7 polymerase under control of the lac promoter. IPTG induction turns on T7 polymerase expression which in turn transcribes from the T7 promoter located on the plasmid 5' to the inserted sequence (Studier and Maizel, 1969). Expression in insect cells is enabled by the presence of the p10 promoter, one of the very late promoters widely

used in the baculovirus expression system (Kelly *et al*., 2016) while expression in mammalian cells is mediated by a hybrid promoter composed of the cytomegalovirus (CMV) immediate early enhancer fused to the chicken β-actin promoter, termed a CAG promoter (Jarvis, 2009). None of these hosts is expected to contain endogenous proteins that will react with either the HSV or HIS tag antibodies meaning that only the plasmid encoded product, along with any processed intermediates, should be detected.

3.2.3.1 Main protease expression in *E. coli*

E. coli was screened first as an expression host following transformation of the verified Abyssovirus M^{Pro} plasmids into T7 polymerase encoding bacterial strains, typified by *E. coli* BL21 (DE3)-pLysS. All strains encode T7 polymerase, allowing transcription from the plasmid encoded T7 promoter, and T7 lysozyme which is a natural inhibitor of T7 RNA polymerase activity and minimizes "leaky" expression prior to induction. This can be important if the encoded product is detrimental to *E. coli* growth, as indeed a protease might be. Variant expression strains include *Origami* which has mutations in both the thioredoxin reductase (*trx*B) and glutathione reductase (*gor*) genes, which enhance disulfide bond formation in the cytoplasm. *Rosetta* is an engineered strain which is designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli* and *Tuner* is a strain which carries a *lacZY* deletion which enables adjustable levels of protein expression via varying levels of the lactose analogue isopropyl-β-D-1 thiogalactopyranoside (IPTG) used as inducer. For these tests, following transformation of the wild type construct into the relevant strain, a colony was

selected randomly and grown in 10 ml LB broth with ampicillin (100 μg/ml) at 37°C overnight in a shaking incubator. From these cultures, 200 μl was transferred to a fresh 10 ml LB broth with ampicillin and grown to OD_{600} = 0.4-0.6 and protein expression induced by the addition of (IPTG) and allowed to continue for 4 hours for optimal expression. Before and after IPTG induction, aliquots of the bacterial culture were harvested for western blot analysis with only the HSV antibody as an initial screen. Induction was successful for all strains except *Origami* with a clear band at around 46kDa, the predicted size of the complete translation product, after one-hour induction with IPTG, which became more intense with time **Figure 3.11**. However, no reactivity of the HSV antibody with a fragment of ~13kDa was apparent suggesting that although the wild type protease sequence was expressed no enzyme activity was present. Based on this data, the strains, BL21 (DE3)-pLysS and *Rosetta* were chosen as expression positive hosts for the analysis of the protease active site and cleavage site mutations already described.

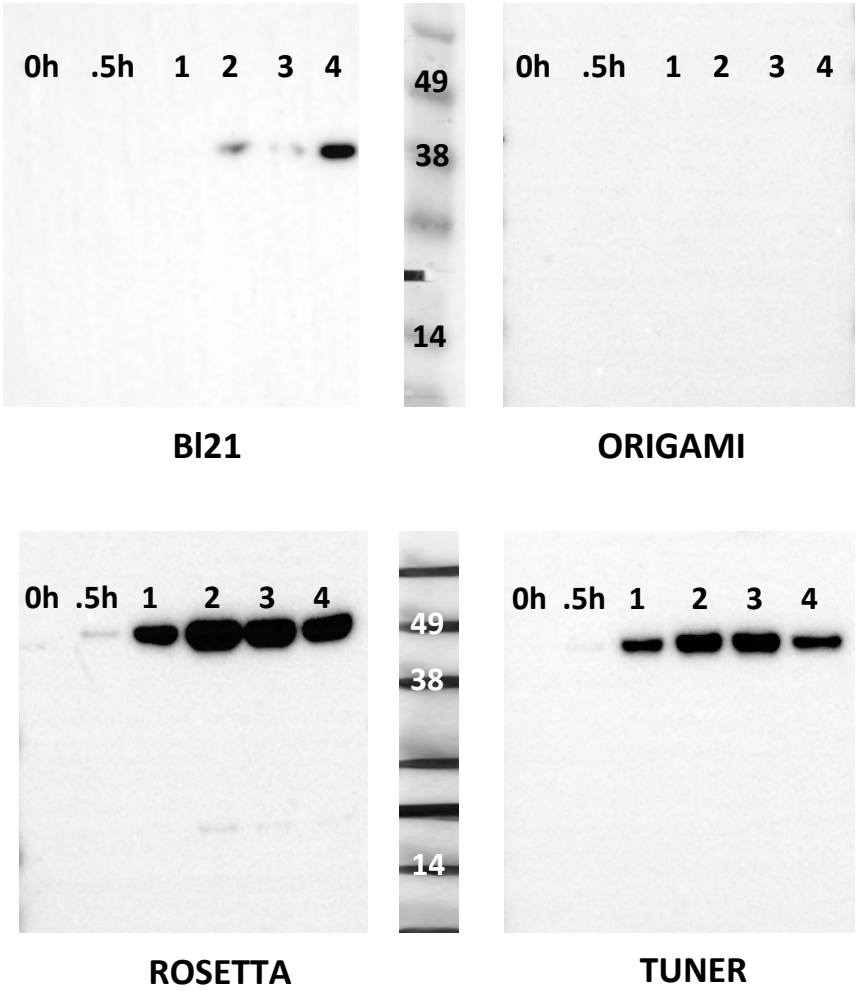

Figure 3.11. Western blot analysis of the wild type Abyssovirus M^{Pro} expression in *E. coli.* The strains BL21 (DE3)-pLysS, Origami, Rosetta and Tuner are indicated. Total protein samples were prepared prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C hourly to 4 hours post induction. The blots were probed with an anti HSV tag antibody only. Markers in the centre are in kilodaltons, kDa.

3.2.3.1.1 Abyssovirus M^{Pro} wild type and mutant expression in *E. coli* BL21 (DE3)-pLysS

All Abyssovirus M^{Pro} constructs, as described above, were transformed into *E. coli* BL21 (DE3)-pLysS and one colony from each construct was selected randomly for expression as before, that is growth to OD_{600}=0.4-0.6 before induction with IPTG and sampling for up to 4 hours post induction. Total protein extracts were resolved by SDS-PAGE, transferred to Immobilon membranes and Western blot analysis performed using both HSV and HIS antibodies. As for the original expression screen, Western blots with the HSV antibody detected a ~46kDa band in the wild type sample consistent with the full length translated product which became more intense with time. In addition, this pattern of expression was maintained for all constructs irrespective of the sequence expressed and there was no indication of smaller fragments consistent with proteolysis of the precursor in any of the mutant samples - **Figure 3.12**.

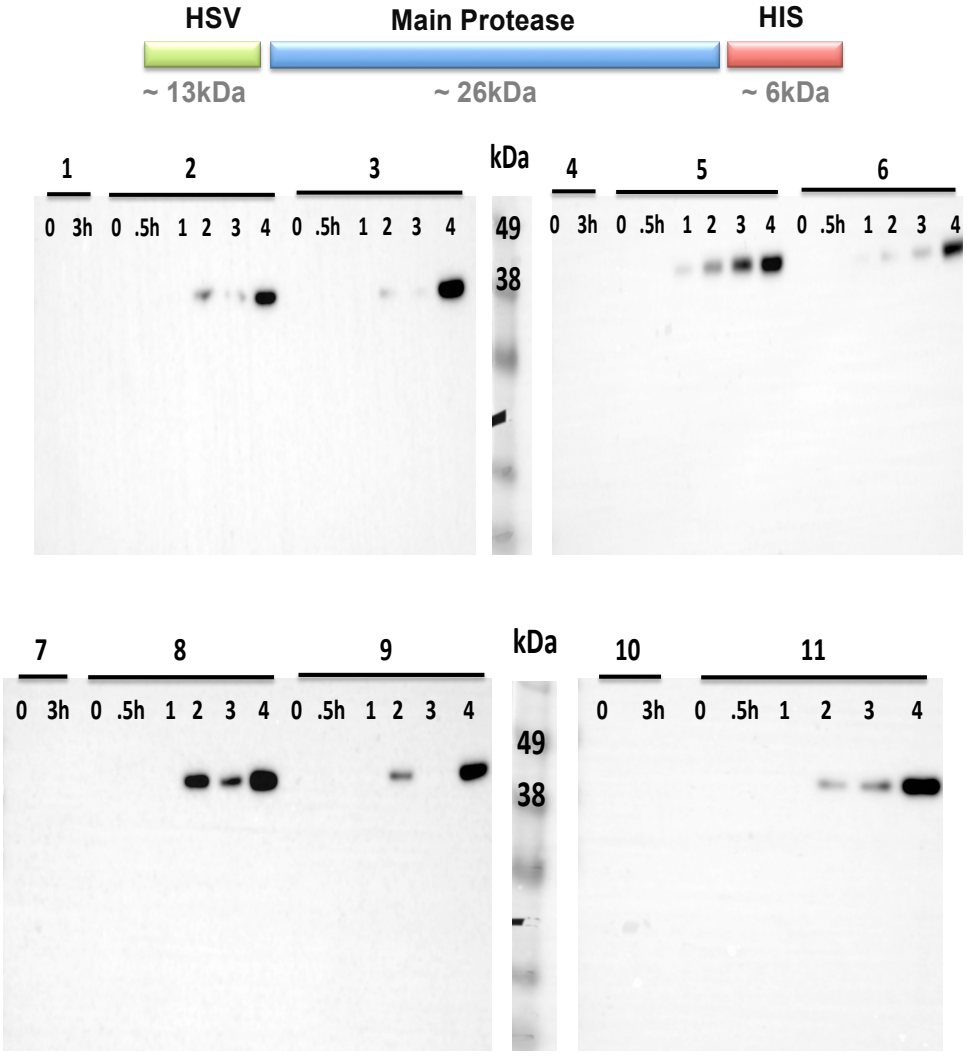

Figure 3.12. Western blot analysis of wild and mutant Abyssovirus M^{Pro} in *E. coli* BL21 (DE3)-pLysS with HSV tag antibody. Protein samples were extracted from *E. coli* prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. The construct design and expected molecular weights are shown above. Samples 1, 4, 7 & 10: non transformed cells acted as control. Sample 2: wild main protease. Sample 3: first active site mutation (C302A). Sample 5: Both active sites mutation (C302A and H193A). Samples 6, 8, 9 & 11: mutant main protease cleavage sites (R136A, K137A, N138A and E139A) respectively. The occasional lack of a band, e.g. 3rd cleavage mutant at 3hr post induction, is due to misloading. Markers in the centre are in kilodaltons, kDa.

In addition to Western blots using the HSV antibody the same set of samples were probed using the anti HIS -tag antibody for the detection of any C-terminal proteolytic fragments. Irrespective of the sequence being expressed the HIS pattern was consistent for all blots with a predominant band at ~30kDa and a variety of bands with higher and lower molecular weights. One of these could be consistent with the primary translation product (~46kDa) but it was not the predominant reactive band and its relative level did not alter among the mutants tested - **Figure 3.13**.

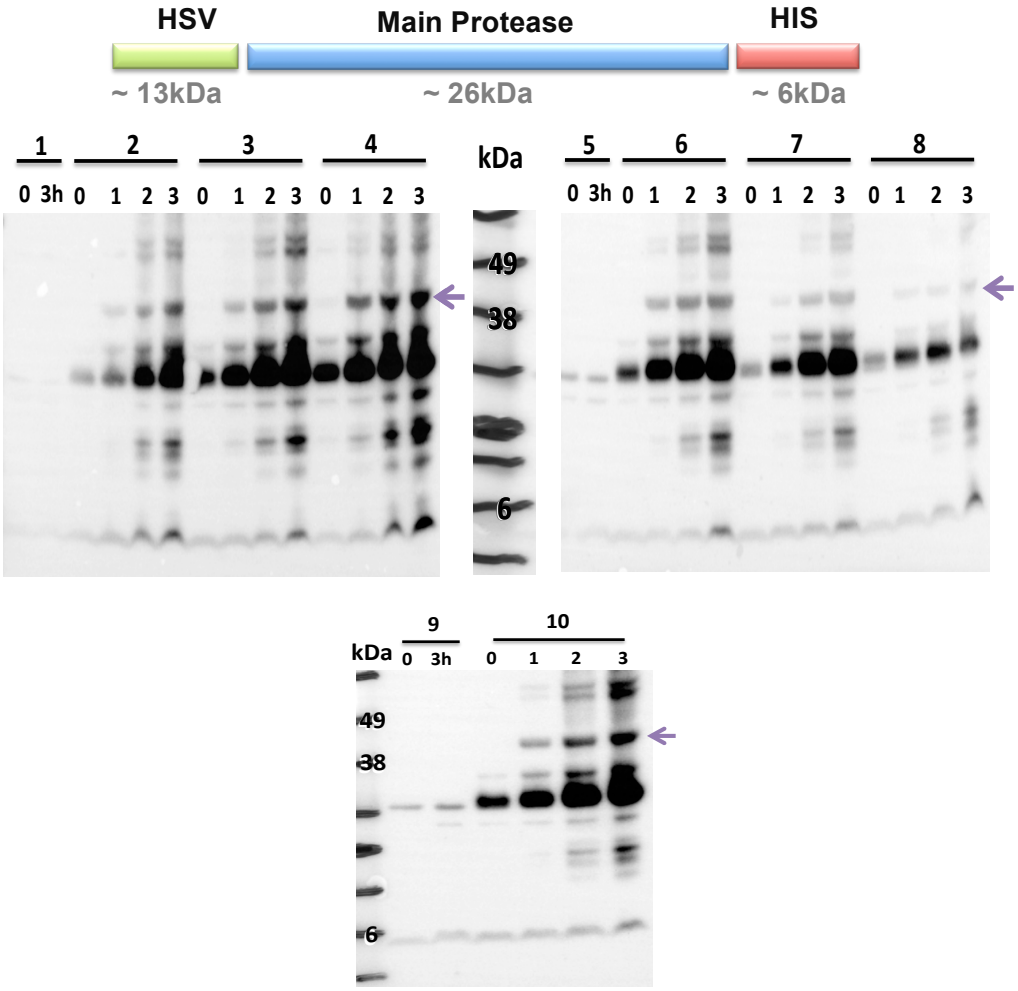

Figure 3.13. Western blot analysis of wild type and mutant Abyssovirus M^{Pro} protease in *E. coli* BL21 (DE3)-pLysS with HIS tag antibody. Protein samples were extracted from *E. coli* prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. Samples 1, 5 & 9 non transformed cells acted as control. The construct design and anticipated molecular weights are shown above. Sample 2: wild main protease. Sample 3: first active site mutation (C302A). Sample 4: both active sites mutation (C302A and H193A). Samples 6, 7, 8 & 10 are mutant main protease cleavage sites (R136A, K137A, N138A and E139A) respectively. A possible primary translation product of ~46kDa is indicated. Molecular weight markers are indicated in the center of the top blots and in the left of bottom blot and are in kilodaltons.

3.2.3.1.2 Abyssovirus M^{Pro} wild type and mutant expression in *E. coli* Rosetta

The data for expression of the Abyssovirus M^{Pro} wild type and mutant proteases in *E. coli* BL21 (DE3)-pLysS were disappointing in that while expression was detected no differences among the constructs was observed. However, it could not be excluded that this was a property of this particular strain and as many other T7 compliant strains are available the experiment was repeated after transformation of the *E. coli Rosetta* strain, a BL21 derivative designed to enhance the expression of eukaryotic proteins that contain rare codons. As before the transformants were grown and induced by the addition of IPTG at an OD_{600} of 0.4-0.6 and samples taken at regular intervals (1-4h). Following SDS-PAGE, Western blotting was performed with both HSV and HIS antibodies. After the addition of IPTG, expression of a band of~46kDa was apparent following the HSV-tag antibody, which peaked in intensity at around 3 hr post induction - **Figure 3.14**. As for expression in BL21 however, no differences were observed among the constructs tested.

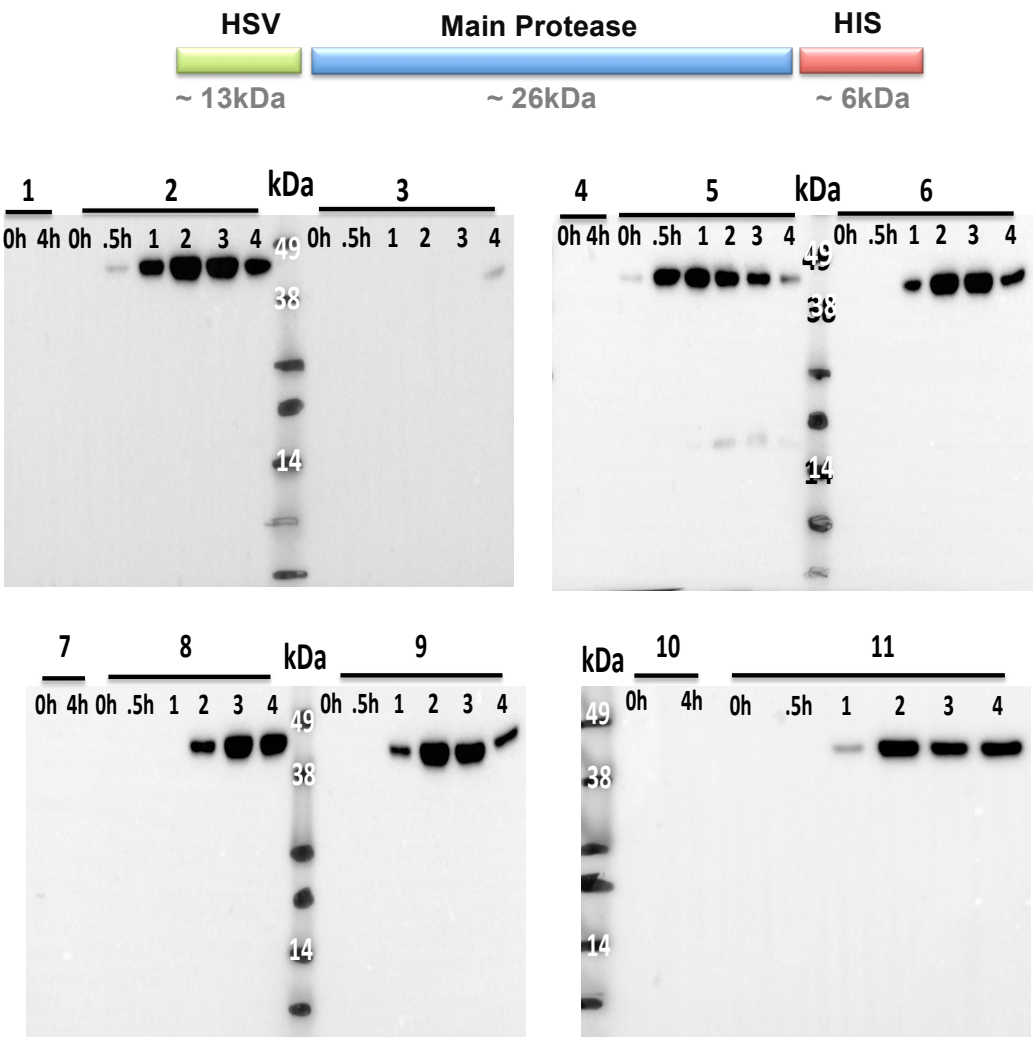

Figure 3.14. Western blot analysis by HSV tag antibody of wild type and mutant Abyssovirus M^{Pro} in *E. coli* Rosetta. Protein samples were extracted from *E. coli* prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. The construct design and anticipated molecular weights are shown above. Samples 1, 4, 7 & 10 non transformed cells acted as control. Sample 2: wild main protease. Sample 3: first active site mutation (C302A). Sample 5: Both active sites mutation (C302A and H193A). Samples 6, 8, 9 & 11 are mutant main protease cleavage sites (R136A, K137A, N138A and E139A) respectively. Molecular weight markers are indicated in the center or left of the blots and are in kilodaltons.

Similarly, when the same samples were processed for Western blot analysis using the anti HIS -tag antibody a predominant band of ~30kDa was highlighted in all constructs with a minor more slowly migrating bands at ~46 kDa which could represent the primary translation product - **Figure 3.15**.

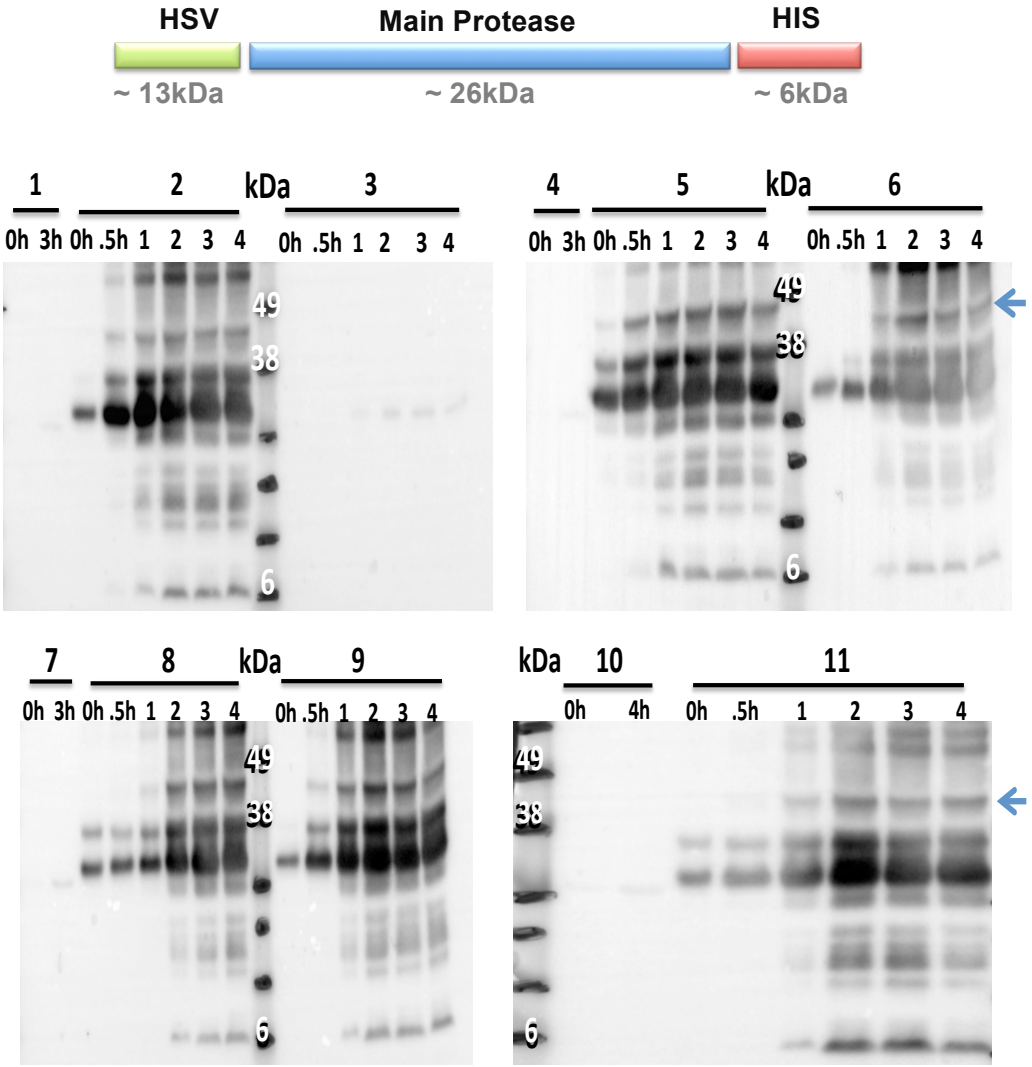

Figure 3.15. Western blot analysis by HIS tag antibody of wild type and mutant Abyssovirus M^{Pro} in *E. coli* Rosetta. Protein samples were extracted from *E. coli* prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. The construct design and anticipated molecular weights are shown above. Samples: 1, 4, 7 and 10 are non transformed cells acted as control. Sample 2: main protease wild type. Sample 3: first active site mutant (C302A). Sample 5: both active site mutants (C302A and H193A). Sample 6: first cleavage site mutant (R136A). Sample 8: second cleavage site mutant (K137A). Sample 9: third cleavage site mutant (N138A). Sample 11: forth cleavage site mutant (E139A). Molecular weight markers are indicated in the center or left of the blots and are in kilodaltons.

The overall outcome of expression screening in *E. coli* was confirmation of the translatability of the constructs designed and detection of the primary translation product by both HSV and HIS tag antibodies. However while the complete protein was detected at the predicted molecular weight and with very little degradation by the HSV tag antibody there was considerable non-specific cleavage apparent when the blots were probed with the HIS tag antibody. Moreover no differences were observed among the constructs tested; all gave the same expression profile in two different *E. coli* strains. While this could indicate that the putative Abyssovirus M^{Pro} protease was not active it was also possible that expression in *E. coli* was incompatible with functional activity, for example because the protein was not soluble or did not fold correctly. A eukaryotic expression system was therefore tested for the expression of the Abyssovirus M^{Pro} wild type and its mutant derivatives.

3.2.3.2 Protein expression in the insect Sf9 cells

The Abyssovirus M^{Pro} encoding plasmids, parental and mutant, were used to generate recombinant baculoviruses as described in Materials and Methods. Briefly, in addition to the promoter elements described, the pTriEx1.1 vector also contains short sequences of the *Autographa californica* baculovirus genome from either side of the polyhedrin locus, a site in the genome which can be altered without inactivating virus replication. When transfer vector and baculovirus DNA are transfected into insect *Spodoptera frugiperda* cells using a standard transfection agent recombination occurs between the sequences on the transfer vector and those of the virus genome resulting in a recombinant virus in which the sequence between

the recombination sequences, in this case the Abyssovirus M^{Pro} sequence or its mutated variants, is inserted into the virus genome. The recombinant genome then initiates a full infection and the recombinant virus expresses the target protein as part of the virus life cycle. The baculovirus genome used in the transfection has been modified to ensure that this process is 100% efficient and the only virus to grow is the desired recombinant (Zhao *et al*., 2003).

Sf9 insect cell lines were cultured in EX-CELL® 420 Serum-Free medium and transfection done in 6 well dishes seeded at ~50% confluence with 1 x 10^6 cells per well. In addition to the M^{Pro} plasmids a control vector pTriEx1.1- GFP, which expresses Green Fluorescent Protein (GFP), was used to visualize the efficiency of transfection (O'Flynn *et al.,* 2012). Recombinant viruses were amplified by passage on Sf9 cells until cytopathic effect was evident within 2 days of infection, usually by passage 3. To assess protein expression by the recombinant viruses 6 well dishes were seeded at near 100% confluence and infected with the P3 virus stocks at an estimated MOI of 3 or greater under which conditions every cell will receive at least one virus. The plates were incubated at 28 ^{O}C for three days and the monolayers collected by dislodging the cells with a pipette, transferred to Eppendorf tubes and pelleted at 13,000 rpm for 3 minutes (microfuge). The supernatant was discarded and the cell pellet lysed directly by resuspension in 1 x SDS-PAGE loading buffer and heated at 100^{O}C for 10 minutes. The equivalent of 5 x 10^4 cells was loaded per well of a standard 10 cm precast SDS polyacrylamide gel and electrophoresis was carried out for 30 minutes at 170V. Gels were then transferred to PVDF membranes by semi-dry blotter and the membranes processed for western blot as described in Material and Methods using antibodies to the HSV or HIS tag – **Figure 3.16**.

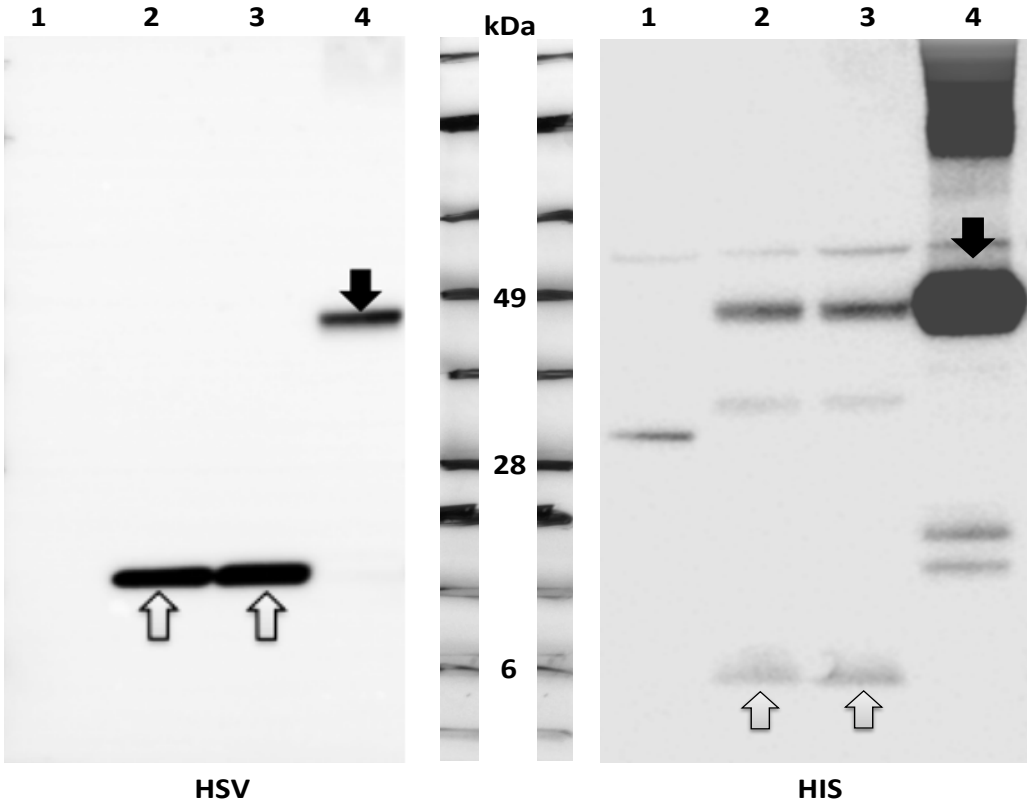

Figure 3.16. Abyssovirus main protease wild type and active site mutants expressed in insect cells. Lane 1: GFP control. Lane 2: Wild type M^{Pro}. Lane 3: active site 1 mutation (C302A). Lane 4: active site 2 mutation (C302A and H193A). Prominent bands detected by each antibody are indicated. Probing with the HSV or HIS tag antibodies is indicated. Molecular weight markers are indicated in the center of both HSV and HIS gels and are in kilodaltons.

In contrast to the data obtained from *E. coli* the results of the western blot analysis indicate the production of a 14kDa HSV related band and a 6kDa HIS related band (empty arrows) when the wild type M^{Pro} sequence is expressed in insect cells. No product the size of the calculated molecular mass of the primary translation product, ~46kDa, was detected in the wild type sample suggesting efficient auto processing by the encoded protease activity. This data is the first to indicate the

Abyssovirus M^{Pro} is enzymatically active. The same profile of antibody reactive bands was apparent when active site 1 mutation (C302A) was analyzed (Lane 3) but both the low molecular mass HSV and HIS reactive bands were lost when mutant active site 2 mutation (C302A and H193A) was analyzed. In their place both the HSV and HIS antibodies detected a single product (filled arrows) at the apparent ~46kDa MW of the primary translation product (Lane 4). Thus while a Cys to Ala change at position 302 did not inactivate enzyme activity change of both the His and Cys residues at position 193 and 302 abolished activity, consistent with them being key residues of the active site.

The same system of expression in insect cells with infection and processing carried out as described above was used to investigate the N terminal cleavage junctions used by Abyssovirus M^{Pro}. Mutation of either of the basic residues found in the RK sequence at positions 136 and137 did not alter the pattern of expression with both constructs showing cleavage at the N- and C-termini as for the wild type sequence. Mutation to Alanine at either N138 or E139 abrogated cleavage as probed with the HSV tag antibody which show only the ~46kDa primary translation product, consistent with these residues being part of an N-terminal cleavage site. However, when probed with the HIS antibody the ~46kDa primary product was also observed with no C-terminal cleavage fragments - **Figure 3.17**. These data suggest inactivation of protease activity rather than a site specific cleavage effect.

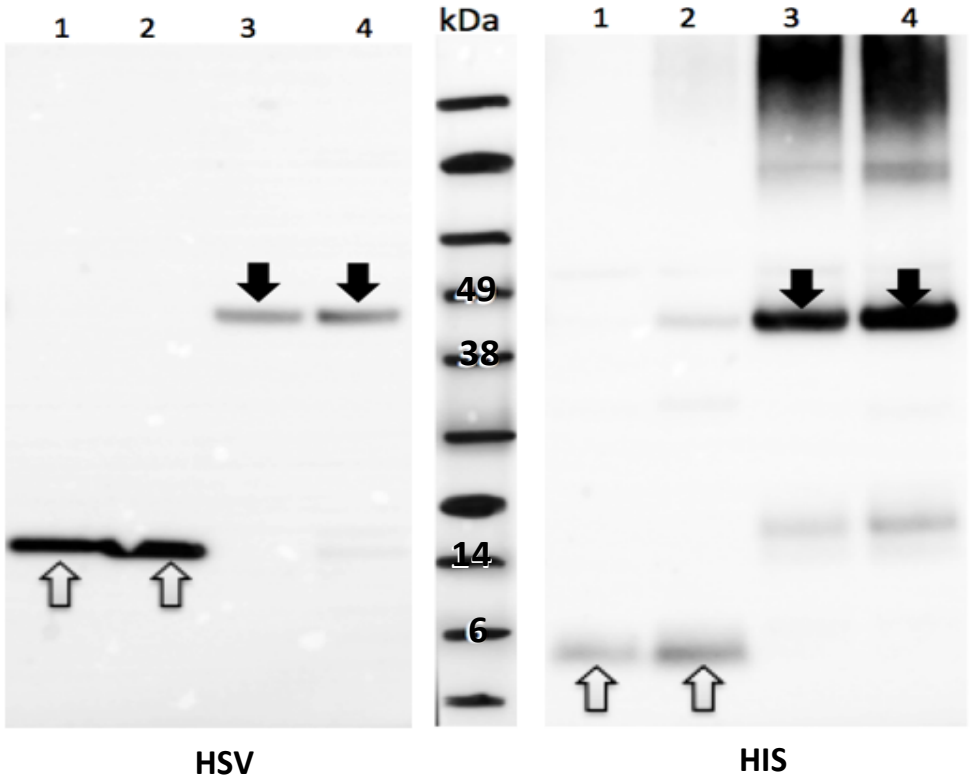

Figure 3.17. Abyssovirus main protease cleavage site mutants expressed in insect cells. The lanes are: 1-cleavage site 1(R136); 2-cleavage site 2 (K137); 3- cleavage site 3 (N138); 4- cleavage site 4 (E139). Prominent cleavage bands as detected by each antibody are indicated by open arrows while failed cleavage is shown by the closed arrows. Probing with the HSV or HIS tag antibodies is indicated. Molecular weight markers are indicated in the center of the blots and are in kilodaltons.

Mutation of Q at position 152 to Alanine stopped the cleavage as probed with the HSV tag antibody which show only the ~46kDa primary translation product, consistent with these residues being part of an N-terminal cleavage site. However, when probed with the HIS antibody the ~46kDa primary product was also observed with no C-terminal cleavage fragments in contrast with wild main protease and RK

sequence which showed evidence of cleavage in the C- terminus and with sizes 6kDa for the HIS-tag and 32kDa for the main protease + HIS-tag (**Figure 3.18**) which again suggests inactivation of the protease function rather than a site specific cleavage effect.

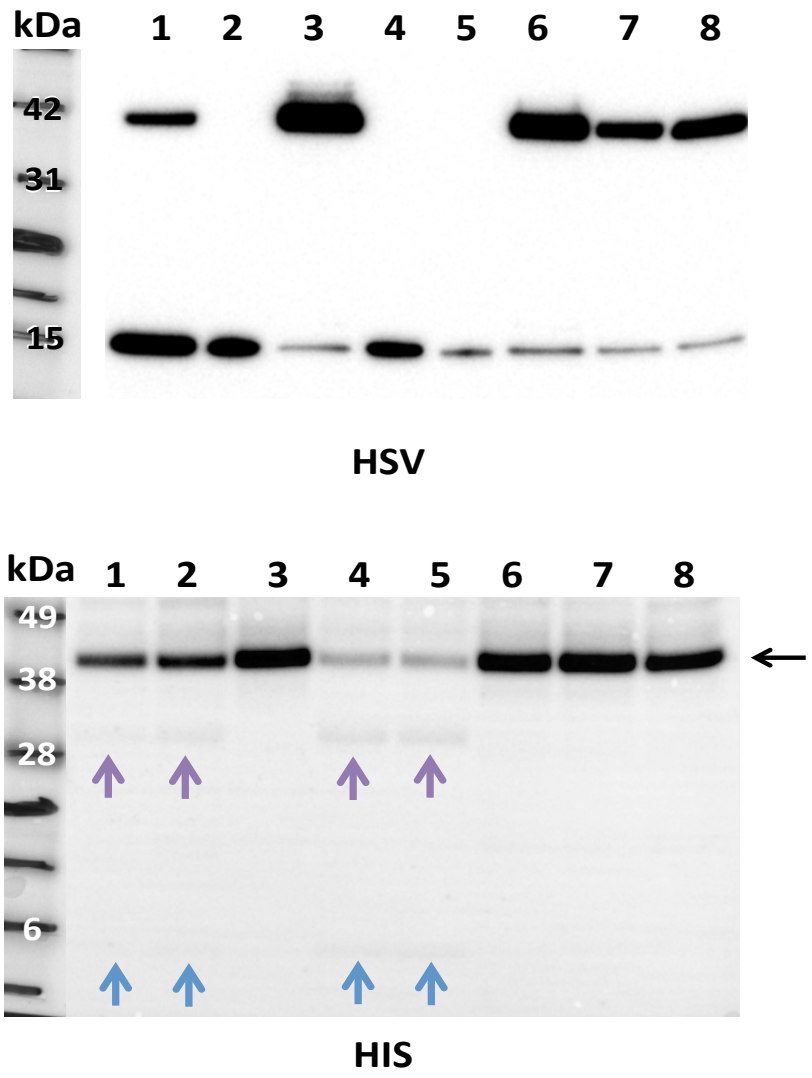

Figure 3.18. Expression of wild Abyssovirus main protease, mutant active sites and cleavage sites in insect cells. The lanes are: 1-Wild main protease. 2- First active site mutation (C302A). 3- Second active site mutation (H193A & C302A). 4- (R136A), 5- (K137A), 6- (N138A) & 7- (E139A) cleavage site mutations respectively. 8- Q↓S cleavage site mutation (Q152A). Probing with the HSV or HIS tag antibodies is indicated. Gray arrows as detected by HIS antibody indicate obvious cleavage bands, the blue arrows represent the protease+HIS-tag. The black arrow represents the complete proteins. Molecular weight markers are indicated in the left of the blots and are in kilodaltons.

Similar analysis of the cleavage site mutations revealed two residues where the pattern of tag detection did not change (mutants 1 and 2) but three residues (mutants 3, 4 and Q) where processing was again lost.

3.2.3.3 Expression and optimization of Abyssovirus main protease in mammalian cells using two transfection reagents

While the data obtained in insect cells was encouraging, it cannot be ruled out that even better data, particularly in relation to the effects of the cleavage junction mutations, might be obtained in other eukaryotic cells. Accordingly, mammalian cells were also transfected with the Abyssovirus M^{Pro} constructs made, taking advantage of the CAG promoter present on the pTriEx1.1 vector to drive expression of the encoded protease. To do this, the 17 clone-1 HEK (Human embryonic kidney) mammalian cell line, acknowledged to be easily transfected (Aricescu *et al*., 2006), were sourced and cultured in Dulbecco's modified Eagle's medium (DMEM). Transfections of the protein expressing constructs were carried out in 6 well plates using monolayers of cells and two different transfection reagents, Lipofectin and Lipofectamine3000, in accordance with the manufacturer's instructions, to determine the optimal conditions for plasmid transfection in this cell line. A control vector pTriEx1.1-GFP, which carries the Green Florescent Protein (GFP) gene inserted between the *Nco*I and *Bsu*361 sites in pTriEx1.1 was used to visualize the efficiency of transfection (O'Flynn *et al.,* 2012). Optimization was required to determine the optimal DNA concentration for both transfection reagents, recorded as the best level of GFP expression, and results showed that Lipofectamine3000 gave a better level, with high DNA concentration, while

Lipofectin gave a relatively low level of expression in all DNA concentrations **Figure 3.19**.

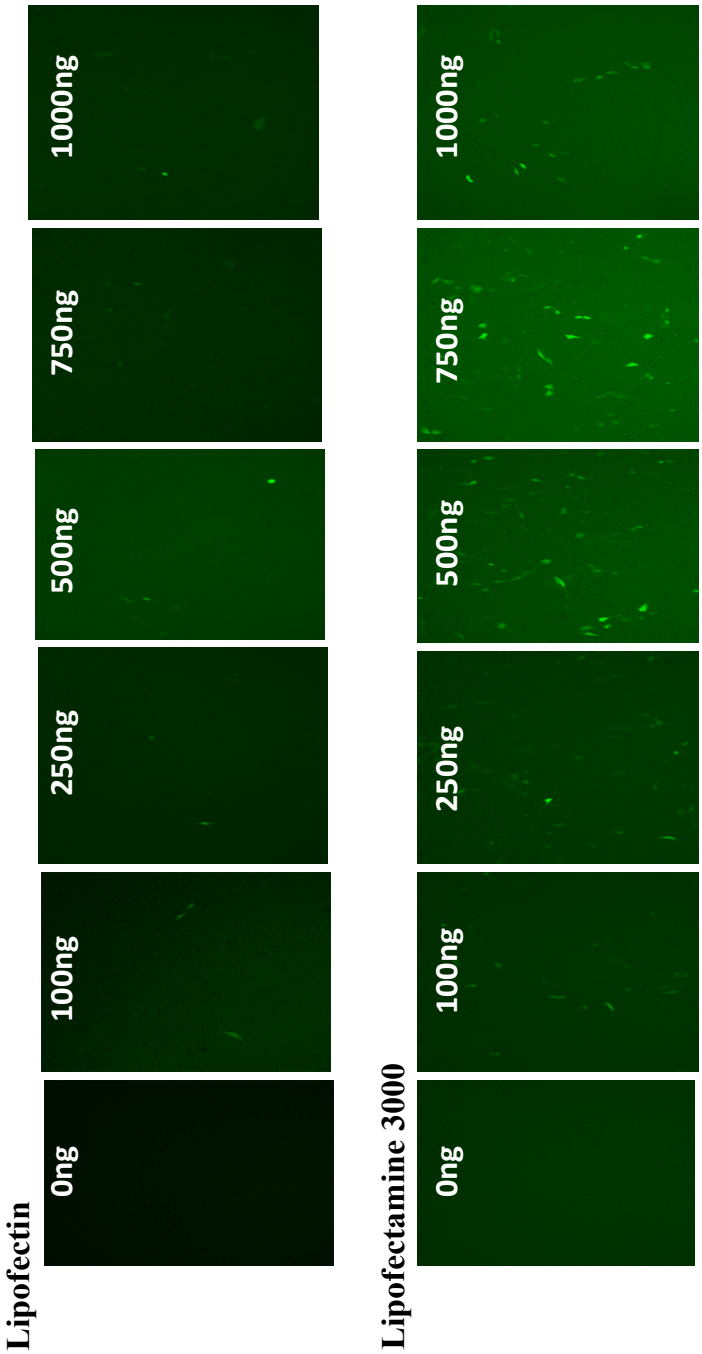

Figure 3.19. *In vitro* transfection of 17clone-1 cells with pTriEx1.1-GFP using Lipofectin and Lipofectamine 3000 transfection reagents. Cells were grown in equal conditions and transfected with different pTriEx1.1-GFP amounts using two different transfection reagents. The intensity of GFP was observed under fluorescence microscopy after 24 hours.

Disappointingly however, when seeded 6 well plates at the optimum density were transfected under optimal conditions as established using the GFP reporter plasmid, with the Abyssovirus M^{Pro} expressing constructs, already shown to function in other cell types, no wild main protease, first active site (C302A) and second active sites (C302A and H193A) mutants protein bands were detected for any of the clones tested by Western blot using both HSV and HIS tag antibodies. Although the HSV blot showed some bands, the band sizes were different from the size predicted for the translation product and also appeared in all tracks suggesting they are cross reaction of the HSV antibody with the cell background, **Figure 3.20**.

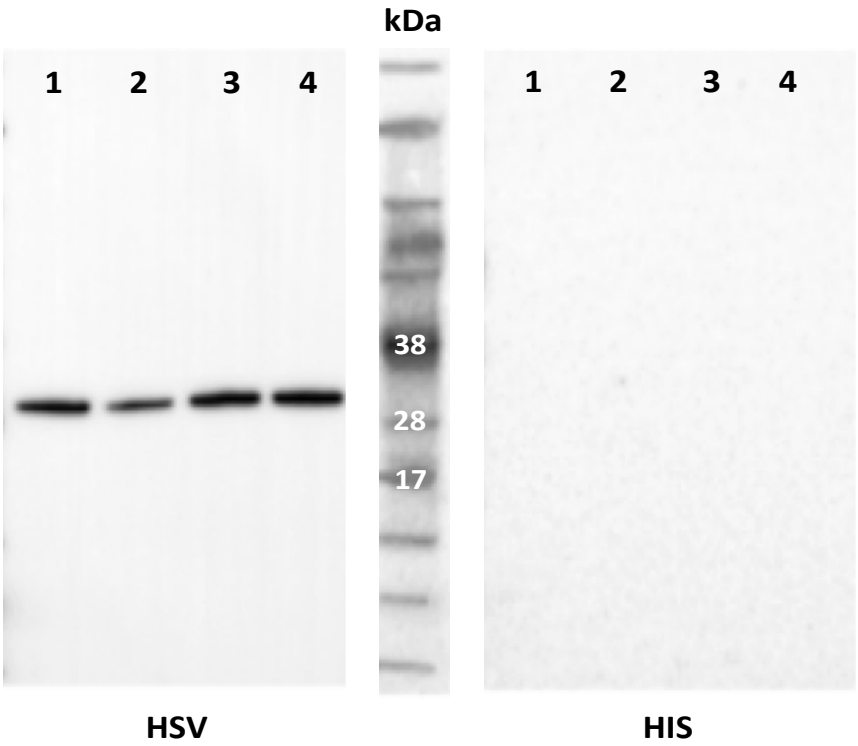

Figure 3.20. Main protease proteins expressed in mammalian cells. Cells were grown in equal conditions and transfected with the main protease protein constructs using Lipofectamine3000 reagent, after 24 hours cells were harvested and total cell extracts were prepared. Protein expression of total cell lysates was detected by Western blot using anti HSV and HIS tag antibodies. Molecular weight marker is in the center of blots and in kDa. Lane 1: GFP. Lane 2: wild Abyssovirus main protease. Lane 3: mutant first active site (C302A). Lane 4: mutant second active site (C302A and H193A).

The lack of products in the Western blot is likely due to poor expression levels, combined with only a subset of cells transfected, in contrast to the insect cell system where every cell is infected. A number of simple options were investigated to improve the outcome, for example the use of polyethylenimine (PEI) transfection

reagent and harvesting cell on ice to reduce protease enzyme activity prior to analysis. PEI condenses DNA into positively charged particles that bind to anionic cell surfaces and DNA:PEI complexes are endocytosed by the cells and the DNA released into the cytoplasm (Sonawane *et al*., 2003). However, as before, the results showed no bands for all the test proteins at the predicted molecular masses and any band visible is different in size from the desired protein. Although the expression of GFP as a control appeared positive on the HIS blot with a size typical of GFP (26kDa), the same band appeared in the HSV blot, suggesting that this band is cell background **Figure 3.21**. From these tests it was concluded that mammalian cell expression would require further optimization which would take time. Mammalian cell expression was therefore not pursued further for these constructs.

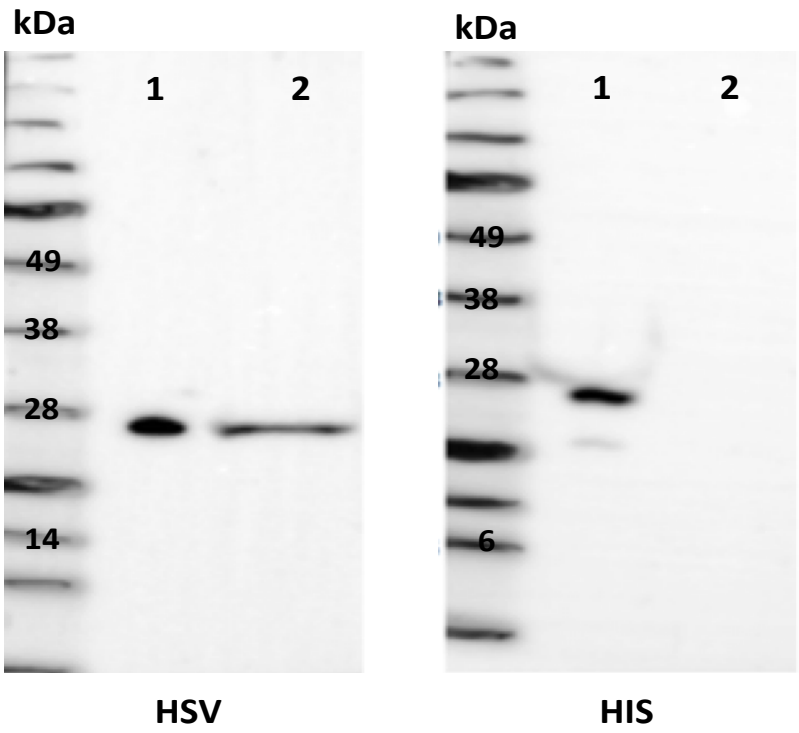

Figure 3.21. Western blot analysis of main protease proteins expressed in mammalian cells using Polyethylenimine (PEI) transfection reagent. Cells were grown in equal conditions and transfected with main protease proteins, after 24 hours cells were harvested then total cell extracts were prepared. Right membrane: Protein expression of total cell lysates was detected by Western blot using anti-HSV tag antibody. Left membrane: protein expression of total cell lysates was detected by Western blot using anti-HIS tag antibody. Molecular weight markers in the left of blots and in kDa, lane 1: pTriEx1.1-GFP was used as a control. Lane 2: wild main protease.

3.3 Discussion

In this chapter the functional features of the putative Abyssovirus non-structural protein M^{Pro} (*aka* the $3CL^{Pro}$) located in ORF1a were examined. The chymotrypsin-like protease shows similarity with other nidovirus main proteases as well as those in many other plus-strand RNA viruses where they play an essential role in the release of mature protein domains from a polyprotein precursor. The 3C designation refers the type member of the class, the 3C protease of picornaviruses which represents the third domain (the C domain) of the third initially translated polyprotein fragment (P3), hence 3C. Many such enzymes have had their structures solved (e.g. Tan *et al*, 2013), including examples from the coronaviridae (Needle *et al*., 2015) and compounds that inhibit proteolytic processing have been shown to have anti-viral activity (Park *et al*, 2016; Sun *et al*, 2016). The Abyssovirus M^{Pro} is assumed to function in the same way, releasing the non-structural proteins required for replicase formation from the 1a and 1ab polyproteins that are translated from the genomic RNA on cell entry (Ziebuhr *et al*., 2000, Dougherty and Semler 1993). The Abyssovirus main protease was identified by sequence analysis and the wild type sequence ordered as a synthetic DNA fragment. It was cloned into an expression vector, pTriEx1.1, using the In-Fusion cloning technique and further mutated variants (including two active site mutations and four suggested cleavage site mutations located in the N terminal region of the protease) were constructed using the overlapping PCR technique. In all cases the cloning design was such that it led to the construction of HSV and HIS tagged open reading frames which translate to tagged fusion proteins which could be identified by western blot with the appropriate antibodies.

As noted, the advantage of using the pTriEx1.1 expression vector as the base vector for activity tests is that expression is possible in three different phyla, bacterial, arthropod and mammalian, as a result of the triple promoters present on the plasmid backbone. As a result it was possible to test for Abyssovirus M^{Pro} activity in cells of each of these backgrounds, represented here by *E. coli* BL21 type strains, insect cells (Sf9) and mammalian cells (17clone1 HEK). Although transfection of mammalian cells reached a reasonable percentage of the culture as shown by transfection with a GFP expressing control plasmid it was not possible to reproducibly detect the M^{Pro} fragments by Western blot. This was likely a result of poor expression levels but it cannot be excluded that the Abyssovirus encoded proteins were particularly sensitive to proteolysis in these cells. Further optimization of transfection or transfer to mammalian cell specific vectors with stronger promoters might resolve this at a future date. By contrast the expression of the Abyssovirus encoded protein in *E. coli* was clear. However, the data on molecular weight and band identity were uncertain despite assessing expression in a number of strains that were all compliant with the T7 based expression system (including: BL21 (DE3)-pLysS, Origami, Rosetta and Tuner). IPTG induction of these strains led to the induction of intense tag reactive bands in whole protein lysates after only one hour of induction but the pattern was inconsistent. For example, probing the N terminus of the primary translation product with HSV antibody detected a protein of the anticipated molecular weight in *E. coli* strains BL21 (DE3)-pLysS, Rosetta and Tuner but no detectable band in *E. coli* strain Origami. Probing the C-terminus using a HIS antibody however revealed a strongly reactive band, in all strains, with an unexpected MW (32 kDa). This could have been concluded to be a processed intermediate, with the HSV tag cleaved, but that

was not supported by the HSV blot which showed only the full length 42kDa product. As with mammalian cell expression it is possible that proteins may be selectively expressed as a result of the action of host factors such as the stability and translational efficiency of the mRNA, protein folding, degradation by host cell proteases and protein solubility (Arun *et al*., 2009). However, these factors might be expected to vary with strain and yet, with minor exceptions, there was no change of pattern with any of the Abyssovirus M^{Pro} mutants. While further work may address the precise reasons for the bands observed the outcome of expression tests in *E. coli* did not allow an assessment of the potential Abyssovirus M^{Pro} activity in this system.

The third system tested was expression in insect cells (*Spodoptera frugiperda*) following infection with recombinant baculoviruses. In this system very clear translation products of the anticipated molecular mass were found in insect cells at two days post infection when the vector encoded p10 promoter is most active (Bleckmann *et al*, 2016). These proteins were detected by both HSV and HIS specific antibodies, depending on the mutant expressed. Histidine (H) and Cysteine (C) residues have been shown to be essential in the catalytic process for coronavirus main proteases (Shan *et al*., 2004). For example, both Cys and His have been mutated in the M^{Pro} of Transmissible Gastroenteritis Virus (TGEV) resulting in the complete loss of catalytic activity (Anand *et al.,* 2002: Huang *et al.,* 2004: Tan *et al.,* 2005: Chang *et al.,* 2007). Alignment of the Abyssovirus $3CL^{Pro}$ sequence with a known coronavirus structure, the $3CL^{Pro}$ of coronavirus HKU4 showed that both catalytic residues were conserved, identifying these residues in the Abyssovirus sequence for mutagenesis.

The aligned active site residues were mutated to Ala in all cases. As noted, the M^{Pro} of MHV has been suggested to process pp1a at dibasic sites (RK) in addition to the preferred M^{Pro} Q/S substrates (Lee *et al*., 1991). More generally, nidovirus main proteases recognize a glutamine or occasionally a glutamate-containing motif at the cleavage site (Ulferts *et al*., 2011) and the potential glutamate-alanine cleavage site near the beginning of the conserved main protease region was tested by alanine scanning mutagenesis. Four mutants, in which each residue was changed to Ala, were assembled at the dibasic sites to assess the possibility of similar cleavage. Expression analysis showed an obvious enzymatic activity with clear bands by both HSV and HIS probes of the correct size (HSV (13kDa) – main protease (26kDa) – HIS (6kDa)), The apparent molecular masses identified were consistent with the upper band representing the protease+tag and the lower band represent the tag only. This data proves that this main protease of this new virus is function. The two active site mutants (First active site mutant Cys to Ala and second active site mutant His and Cys to Ala) were expressed in insect cells (Sf9) and analyzed by western blot. Mutation of the putative catalytic cysteine C302 did not strongly reduce proteolytic processing, while mutation of both histidine H193 and cysteine C302 blocked protease activity. This data is consistent with the published data of other M^{Pro} proteases (Liu & Brown, 1995; Lu *et al*., 1995; Ziebuhr *et al*., 1995, 1997; Tibbles *et al*., 1996; Seybert *et al*., 1997; Shan *et al*., 2004).

As noted previously, four mutants were made (R, K, N and E to Ala) in the N-terminal region of main protease, all suggested from previous studies to be involved in the cleavage junctions, e.g. by MHV by Lee *et al*., 1991. The expectation of the loss of an N-terminal processing site would be for a HIS tag reaction at 6kDa but for an intermediate reaction with the HSV tag antibody at

6kDa less than the unprocessed primary translation product. Curiously however neither low MW band is present suggesting outright inactivation. Other bands of weak reactivity were detected by the HIS antibody but as their molecular mass did not fit with any expected processed product they are assumed to be the products of general cellular proteolysis. Two of the mutants (N (Position 138) and E (Position 139) to ALA) yielded only the higher apparent molecular mass (see **Figure 3.17**) suggesting that these two mutants did not show any enzyme activity. However, the pattern for the R (Position 136), K (Position 137) mutants, there were two bands visualized in western blot (see **Figure 3.17**) which matched those identified in the products of the wild type constructs, suggested no role a key residue of a cleavage junction. Thus, the RKNE sequence targeted as a possible N terminal processing site has a role in M^{Pro} activity but it is not clear what that role is. Neither basic residue affected processing, as might have been expected, but mutation of either the Asn or Glu residue abolished it. It cannot be ruled out that these residues affect the folding of the primary translation product so that M^{Pro} cannot liberate itself by *cis*-cleavage but more mutagenesis will be required to investigate this further.

Commonly, the cleavage sites of M^{Pro} contain a dipeptide (LQ) sequence, and the amino acid Gln is reported as an essential conservative residue (Yang *et al*., 2003). The cleavage site peptide of SARS-CoV main protease has been reported to be in (TSAVLQ/SGFRK) (Fan *et al*., 2004) and a QS motif occurs in the Abyssovirus M^{Pro} sequence just downstream of the RKNE motif. Mutagenesis of this motif failed to demonstrate that it acted as a cleavage junction. Instead, as for the NE mutants, mutation at either residue appeared to inactivate the enzyme.

Overall, the data obtained with the Abyssovirus M^{Pro} demonstrate an activity consistent with that described for related enzymes but the question of cleavage

junctions remains unresolved. Any mutant that appeared to affect the cleavage junction appeared to lose activity altogether. A feature of coronavirus replication is the reorganization of intracellular membranes to form replication centers where all viral enzymes are concentrated (van hemert *et al*., 2008). This is not recapitulated by the constructs described here as no other Abyssovirus protein is co-expressed with M^{Pro}. Therefore it remains possible that while activity is clear, the conformation required for authentic junction cleavage requires other virus proteins to be present at the same time. Alternatively, the Abyssovirus M^{Pro} is somewhat diverged from the mainstream coronavirus M^{Pro} and uses a novel set of boundaries when acting *in cis* and *in trans*.

Chapter 4

4 Stimulation and translation program of Abyssovirus 1a/1b junction sequence

4.1 Introduction

Two thirds of the 5' end of a canonical coronavirus genome encodes the nonstructural proteins which are produced initially as the polyproteins 1a and 1ab. These polypeptides are cleaved into the mature proteins necessary for virus replication by the main protease studied in the preceding chapter. The 1a and 1ab proteins are encoded directly by the incoming positive strand virus genome with a reported ratio of 3:1 for the 1a to 1ab products (Nakagawa *et al*., 2016). This ratio is achieved at the translational level by the ribosome recognizing specific signals in the RNA template at the junction of the 1a and 1b coding sequences. Classically, this signal is a programmed ribosomal frameshift mechanism which regulates the gene expression of ORF 1b relative to 1a.

Frameshifts are composed of two elements, a "slippery sequence" U_UUA_AAC followed by a folded RNA structure in the form of either a pseudoknot or kissing stem loops, both of which have a strong effect on frameshifting efficiencies (**Figure 4.1**) (Baranov *et al*., 2005).

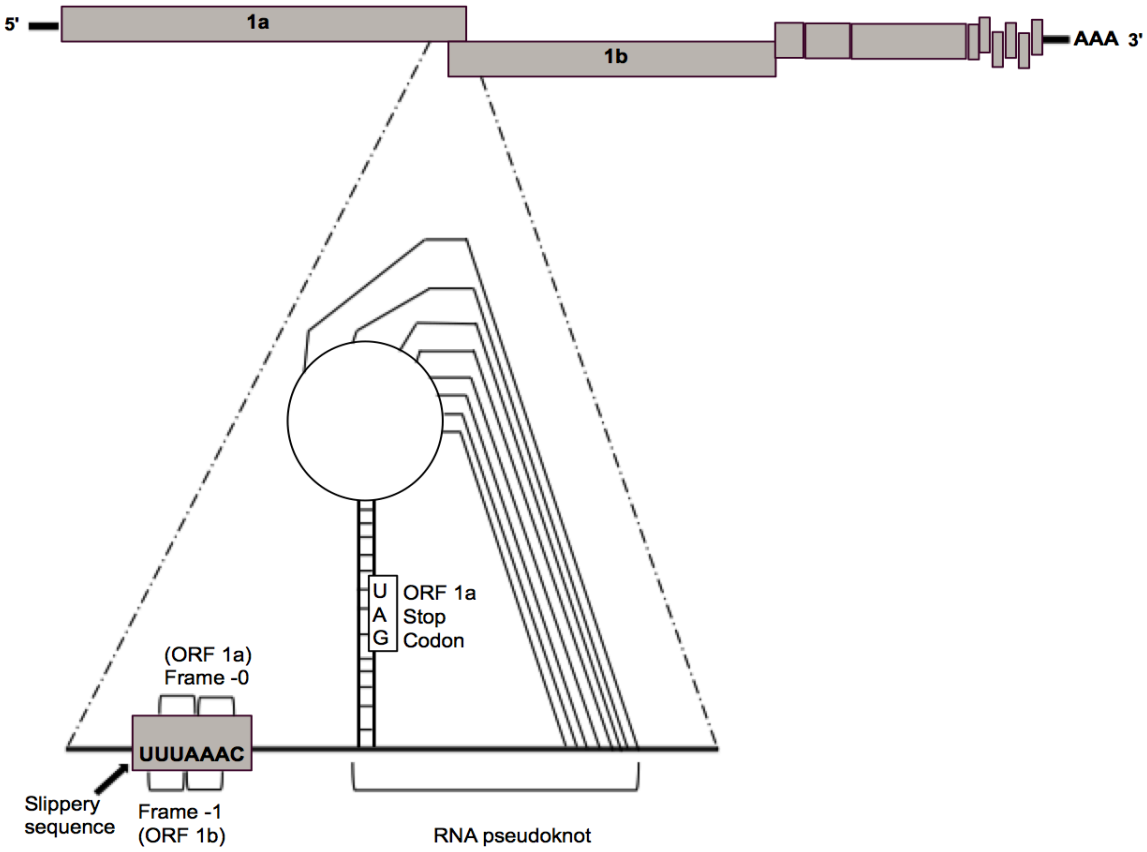

Figure 4.1. Elements required for ribosomal frameshifting in MHV (adapted from Masters, 2006).

Programmed ribosomal frameshifting is the directed switching of a proportion of translating ribosomes into a new reading frame, either shifted one nucleotide position forward (+1 frameshift) or one nucleotide position backwards (−1 frameshift). As a consequence of frameshifting the upstream reading frame become fused to that downstream, which was previously out of frame. Frameshifting occurs in all known organisms, from *E. coli* to mammals, and is an essential mechanism of translational control used by many viral and cellular genes (Baranov *et al*., 2003; Namy *et al*., 2004; Stahl *et al*., 2002; Dinman, 2012).

In coronaviruses the ribosomes translating ORF1a change the reading frame at the 1ab junction to translate information contained in ORF1b. This translational process is essential for the synthesis of the viral RNA-dependent RNA polymerase and other replication factors as ORF1b lacks its own independent translation initiation codon (Baranov *et al*., 2005). Coronavirus frameshifting was established first in Avian Infectious Bronchitis Virus (IBV) (Brierley *et al*., 1989) and extensive subsequent analysis by Brierley *et al*. (1991, 1992), Liphardt *et al*. (1999), and Napthine *et al*. (1999) showed that ribosomes shift translation frame at the slippery sequence U_UUA_AAC, which is invariant among known coronaviruses, stimulated by a downstream pseudoknot. As stated by Ten Dam *et al*., 1990, 3' RNA pseudoknots are commonly used for stimulating -1 programmed frameshifting in viruses.

As described previously, a wide variety of virus-like sequences including new members of the Nidovirales have been reported in the transcriptome shotgun assembly (TSA) and expressed sequence tag (EST) databases. One such sequence found in the transcriptome of *Aplysia californica* and tentatively named the *Abyssovirus*, had homology to the main proteinase, polymerase and helicase, or complete p1b regions of Infectious bronchitis virus, Gill-associated virus, White bream virus, Cavally virus and Wobbly possum disease virus, all related nidoviruses. In the Abyssovirus sequence no sequence typical of a frameshift was found at the predicted 1a/1b junction. Instead, the more unusual feature, an in-frame stop codon was found separating the pp1a and pp1b genes. Termination-suppression signals are used as alternate translational control elements to frameshifts in several other viruses including alphaviruses and retroviruses, and

typically consist of a UAG or UGA stop codon followed by an RNA secondary structure element. The efficiency of suppression normally depends on the stop codon, the nucleotides immediately following the stop codon, and the free energy of the RNA secondary structure element. This finding suggested that Abyssovirus may use translational termination-suppression signal as a way to control expression of the p1b region in relation to p1a.

The mechanism of translational termination-suppression is known. During protein translation, tRNAs occupy successively the universally conserved A (aminoacyl), P (peptidyl), and E (exit) sites of the ribosome (Frank *et al*., 2007). Generally, translation termination occurs when a stop codon enters the decoding center of the A-site. Here they are recognized by protein release factors which promote hydrolysis of the peptidyl-tRNA linkage by the large ribosomal subunit (Petry *et al*. 2008; Loh and Song 2010). Although the termination of protein synthesis is highly efficient it can be affected by the nature of the stop codon (UAA, UAG, or UGA) and the flanking nucleotides, particularly the immediately adjacent 39 bases (Harrell *et al*. 2002). The contexts of some termination codons are noticeably ‘‘leaky’’ (e.g., UGAC - McCaughan *et al*. 1995) and so allow ‘‘readthrough’’ of the stop at a level of 0.3% to 5% (Bertram *et al*. 2001). When a stop codon is readthrough it is decoded by a near relative or “suppressor” tRNA, and the translation process continues to the next stop codon. As with frameshifting, readthrough is used by numerous viral and cellular genes as a means of translational control (Namy *et al*. 2004; Dreher and Miller 2006; von der Haar and Tuite 2007; Namy and Rousset 2010). Analysis of readthrough in viral systems has shown that most readthrough signals use limited flanking information (Harrell *et al*. 2002). For example, in the alphavirus Sindbis virus, the polymerase (nsp4)

readthrough of a UGA codon requires only a single cytosine located immediately downstream from the termination codon (Li and Rice 1993). Similarly, in the tobamovirus tobacco mosaic virus (TMV), the expression of viral p183 RNA-dependent RNA polymerase by readthrough is reported to require only six bases following the stop codon (UAG CAA UUA) (Skuzeski *et al*. 1991).

In the Abyssovirus sequence the p1a gene ends in a UGA stop codon, and the region that follows it is predicted by Mfold (Zuker, 2003) to form several possible RNA secondary structures, including a potential pseudoknot. As with the function of M^{Pro} (Chapter 3) however the lack of experimental evidence for the replication of the Abyssovirus and the absence of any characterisation of its protein products means that the function of the 1a/1b junction sequence is untested. This lack of data is addressed in this chapter with the generation of constructs for the expression of the Abyssovirus readthrough junction in an assayable sequence assembly based on the vector pTriEx1.1 where the predicted unusual translational stop-start signal at the end of p1a can be tested following expression in different host backgrounds including insect cells (*Sf*9), *E. coli* (BL21 (DE3)- pLysS) and mammalian cells (17clone1).

4.2 Results

4.2.1 Construction of tagged Abyssovirus stop suppression vectors

The Abyssovirus stop suppression signal sequence (TGA) located at the end of the 1a reading frame as defined in the TSA database sequence obtained from the sea hare *Aplasya californica* (accession no: GBDA01003924.1) was ordered as a synthetic DNA fragments of 624bp and provided by Dr. Geraldine Mulley. A

second fragment containing mutations of the stop codon and the 3' nucleotide, TGA-C to TAA-A, were obtained in the same way, these changes being suggested from a previous study which showed them to be involved in the efficiency of readthrough (Li and Rice, 1993). Other variants of the stop and following nucleotide were not examined. To provide tags for detection of the translated product the synthetic fragments were cloned by in-Fusion technology into the vector pTriEx1.1 to allow expression of the test protein in *E. coli,* insect cells or mammalian cells as described in Chapter 3 (Jarvis, 2009). The annotated nucleotide sequence of the final assembled fragment and the position of the stop codon readthrough and downstream pseudoknot are shown in **Figure 4.2**. The 15bp overlap sequences enabling infusion cloning were added to the synthetic DNA fragment by PCR amplification and are shown with vector and readthrough homologies identified in **Table 4.1**.

Novel nidovirus sequence from *Aplysia californica* termed Abyssovirus

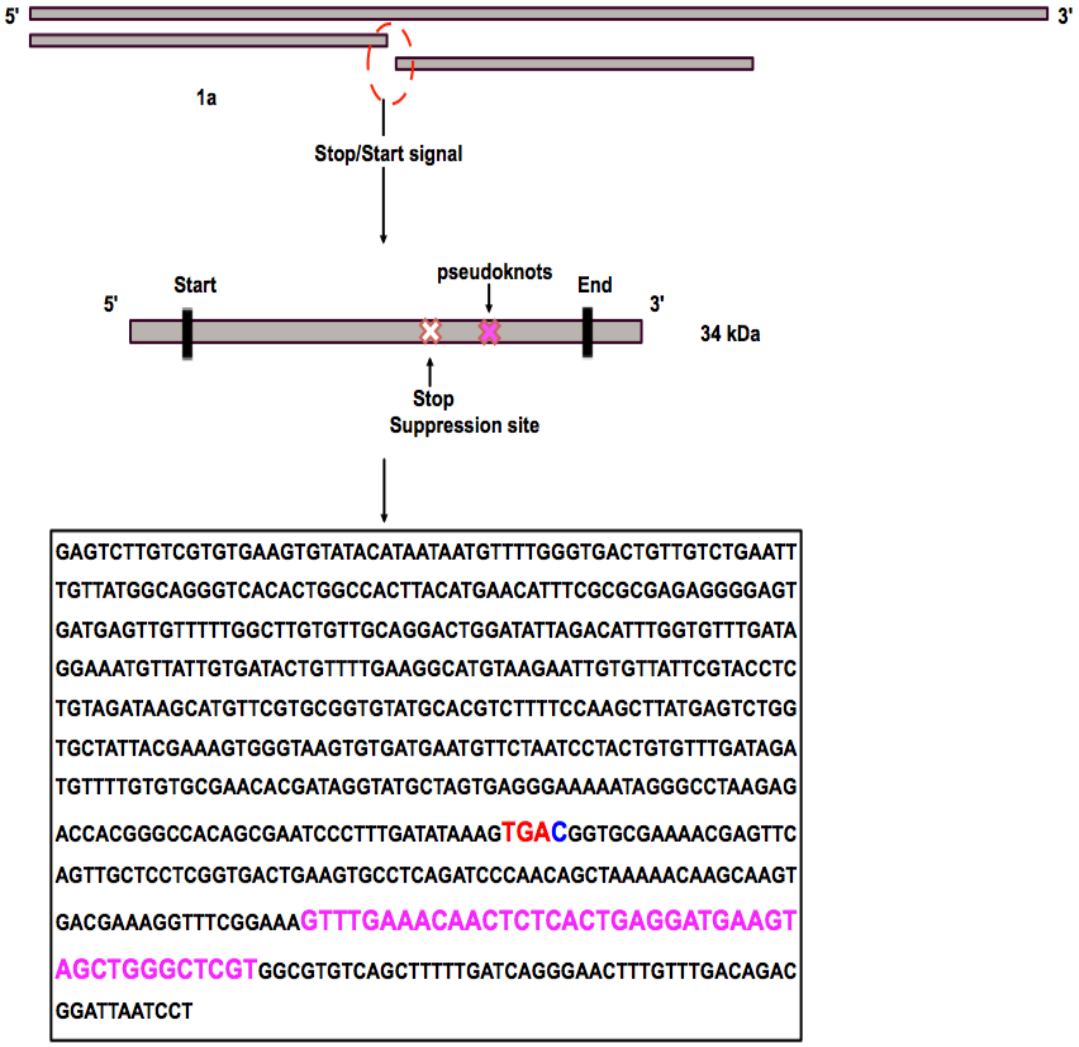

Figure 4.2. Sequence from the Abyssovirus genome showing stop suppression sequence (TGA) located in the frame at the junction of 1a and 1b. The stop codon is in red followed by a cytosine in blue. The base pairs in pink are the Abyssovirus pseudoknot, which is located ~100 bps downstream of the stop suppression sequence itself.

Table 4.1. Oligonucleotides used for in-Fusion of wild type and mutant Abyssovirus stop suppression fragments (TGA-C and TAA-A) into pTriEx1.1 plasmid. Yellow color represents vector sequence. Underlined sequence is the *Xho*I site used for preparation of the linear vector.

Primers	**Sequence**	**bp**
Fw_Inf_Readthrough	5'-CCCCGAGGATCTCGAGGAGTCTT GTCGTGTGAAGT-3'	35
Rv_Inf_Readthrough	5'- GATGGTGGTGCTCGAGAGGATTA ATCCGTCTGTCAA-3'	36

After transformation of the infusion mixes into Stellar^{TM} *E. coli* competent cells and overnight incubation on selective agar, 6 colonies from each infusion reaction were selected at random, transferred onto LB with ampicillin and grown overnight at 37°C with shaking. Plasmid DNA was extracted and analyzed by PCR using the primers used for the generation of the infusion fragments (**Table 4.1**) and DNA which amplified to give the correct size of insert (656 bp) was digested with restriction enzyme (*Xho*I) to confirm the junctions and sent to Source BioScience for DNA sequencing to confirm the identity of the sequences cloned. All sequences obtained were as designed and as a result of their insertion into pTriEx1.1 the test 1ab sequence was flanked with vector sequences encoding the HSV tag 5' to the Abyssovirus insert and the HIS tag 3'. Sequences confirmed that the reading frame was open through the Abyssovirus sequence and in fame with each tag.

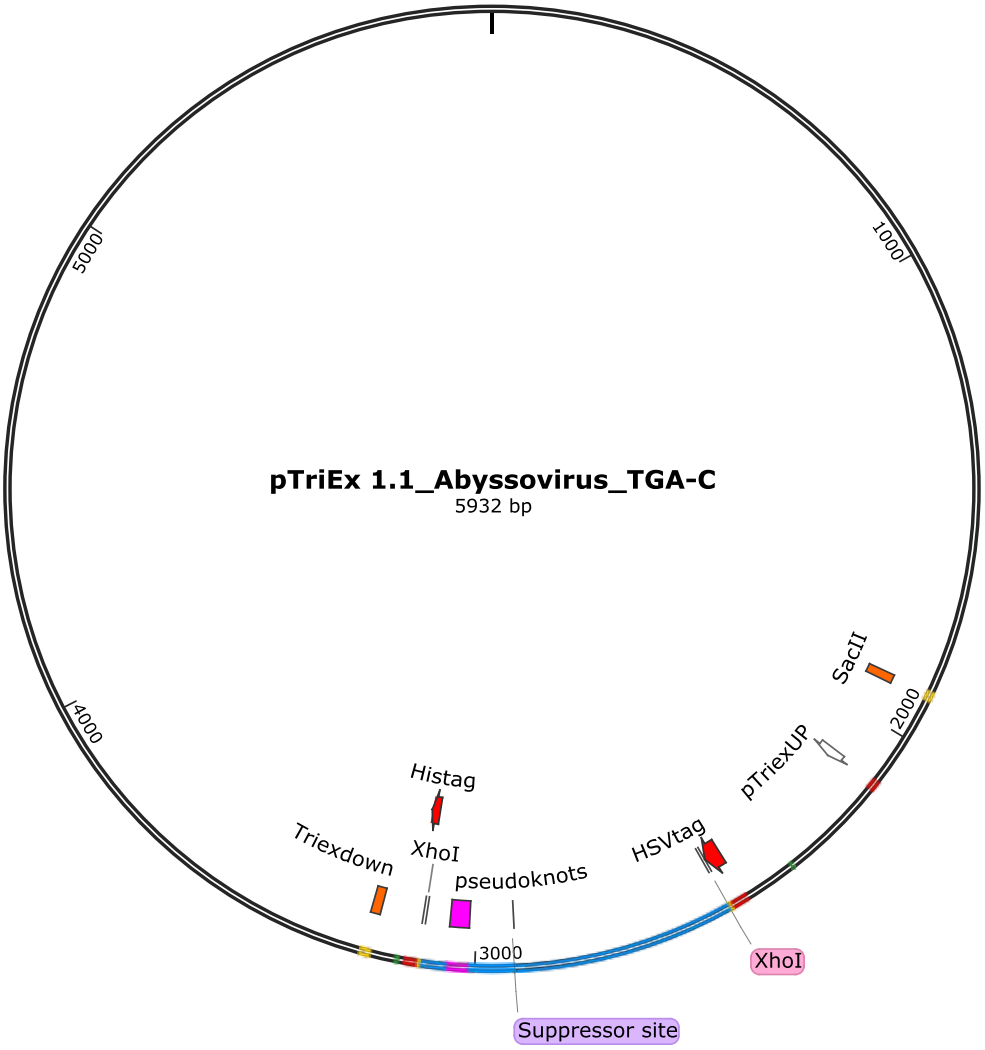

Figure 4.3. Plasmid map showing the complete Abyssovirus stop suppression site (or suppressor site) and pseudoknots (pink), located in the readthrough frame, sequence in the pTriEx1.1 vector. Note the *Xho*I site used for cloning as well as the HSV and HIS tags used for protein detection.

4.2.1.1 Generating of mutant Abyssovirus stop suppression (TGA-C) by using overlap PCR

As noted, it has been reported that pseudoknots play a role in stimulating the efficiency of a number of programmed translational recoding events (Baranov *et al*., 2002), including stop codon readthrough (Howard *et al*., 2005) and ribosomal frameshifting (Farabaugh, 1996; Giedroc *et al*., 2000). In order to determine if the pseudoknot downstream of stop codon at the end of the Abyssovirus 1a reading frame was affected by the presence of a pseudoknot downstream, two further mutants were designed to delete the pseudoknot from each of the original fragments (constructs TGA-C and TAA-A).

As before, the overlap PCR method was used to generate these deletions (See **Figure 3.5** in Abyssovirus main protease chapter), using designed additional forward (Fw_) and reverse (Rv_) primers for each case **Table 4.2.**

Table 4.2. Oligonucleotides used for cloning in pTriEx1.1 plasmid. Highlight sequence represents deletion of pseudoknot.

Primer	**Sequence**	**bp**
Fw_readthrough_new	5'-GCAAGTGACGAAAGGTTTCGGA AA(Δ42)GGCGTGTCAGCTTTTTGATCAGG-'3	47
Rv_readthrough_new	5'-CCTGATCAAAAAGCTGACACG CCTT(Δ42)TCCGAAACCTTTCGTCACTTGC-'3	47

The original cloned fragments (TGA-C and TAA-A), cloned as described above, were used as templates and amplification by PCR used a forward primer (Fw_SacII, **Table 4.3** for the TGAC and Fw_TriExUp**, Table 4.3** for the TAAA) which primed 5' to the inserted sequences and a reverse primer which included the mutation to be introduced (Rv_X see **Table 4.2**) in reaction PCR1. Another PCR was carried out with a forward primer which included the mutation to be introduced (Fw_X see **Table 4.2**) and a reverse primer (Rve_TriExDown, **Table 4.3**) which hybridized 3' to the inserted sequence, reaction PCR2. The PCR products of these first round PCR reactions were analyzed by agarose electrophoresis, and the correct sizes (PCR1: 1160 bp, TGA-C and 956 bp, TAA-A and PCR2: 213 bp) were obtained for both templates **Figure 4.4.**

Table 4.3. Oligonucleotides used to generate mutant Abyssovirus stop suppression site sequence cases.

Primer	Sequence	bp
Fw_TriExUp	5'-GGTTATTGTGCTGTCTCATCA-3'	21
Rv_TriExDown	5'-TCGATCTCAGTGGTATTTGTG-3'	21
Fw_SacII	5'-TGGCTGCGTGAAAGCCTTG-3'	19

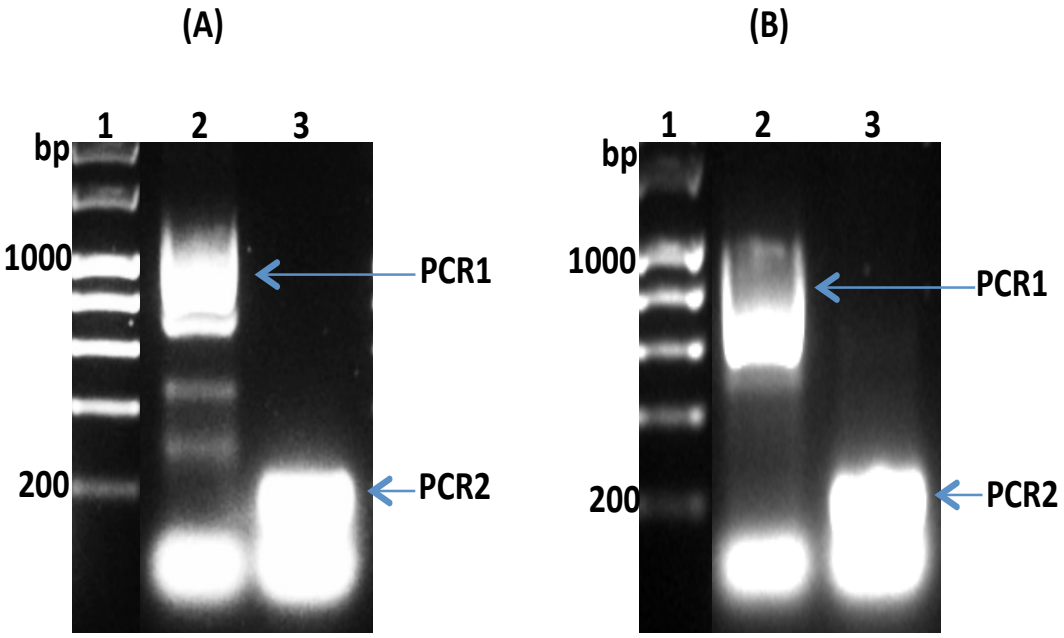

Figure 4.4. Gel electrophoresis of PCR products following amplification of the Abyssovirus stop suppression signal TGA-C (**A**) or TAA-A (**B**) to introduce deletions of the pseudoknot. In both gels, Lane 1: Marker (1kb) DNA ladder, Lane 2: PCR1 ((A)_1160 bp), ((B)_956 bp), Lane 3: PCR2 (A & B 213 bp).

The amplified products were purified from the gel using a PCR purification kit and quantified by Nanodrop spectrophotometer. Then 0.1ng/bp of each PCR product were mixed and subjected to amplification in a PCR reaction for five cycles without any primer addition. Subsequently, the forward primers (Fw_SacII – used for Abyssovirus stop suppression signal TGA-C and Fw_TriExUp used for mutant Abyssovirus stop suppression signal TAA-A with no pseudoknot) and the reverse primer (Rv_TriExDown- for both samples) were added and a further thirty cycles were completed. The correct size 1289 bp and 1085 bp of overlap PCR product of both TGA-C and TAA-A respectively were analyzed by agarose gel electrophoresis as before and the desired overlap bands were extracted by gel extraction kit **Figure 4.5.**

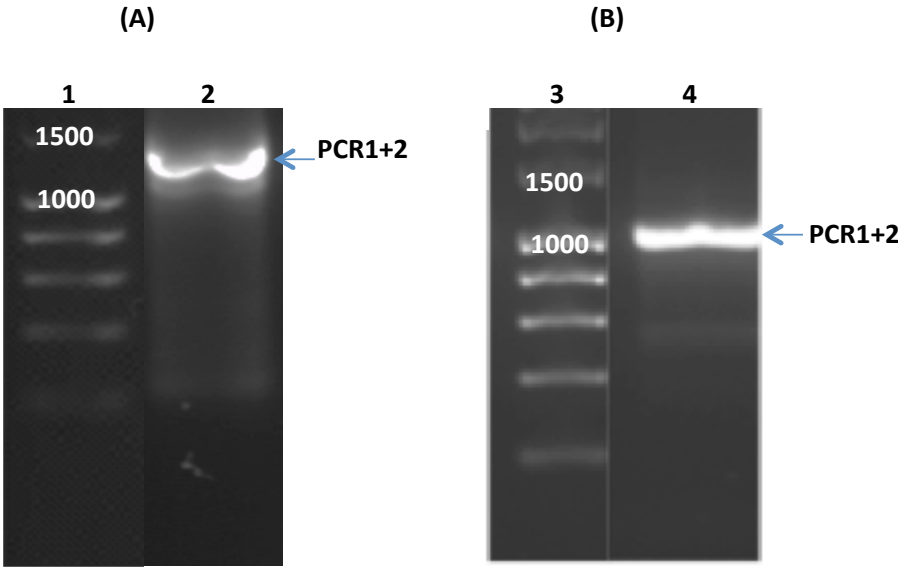

Figure 4.5. Agarose gel electrophoresis of overlap products of Abyssovirus stop suppression site TGA-C (A) and TAA-A (B) deletion mutations. Lane 1 and 3: Marker (1kb) DNA ladder, Lane 2 and 4: PCR1+PCR2 for both Abyssovirus stop suppression site TGA-C and TAA-A mutation cases (1289 bp and 1085 bp). (The image shown is a composite of individual lanes from a single gel where irrelevant lanes have been removed).

The purified products were quantified by Nanodrop spectrophotometer and digested with *Xho*I restriction enzyme for 30 minutes at 37°C followed by a cleanup kit to re-purify the products (**Figure 4.6**). Empty pTriEx1.1 plasmid was also digested with *Xho*I restriction enzyme (**Figure 4.7**) and the large vector band was extracted by gel extraction kit after agarose gel electrophoresis and also quantified by Nanodrop spectrophotometer. The linearized vector was mixed with each overlap mutant fragment and a T4 ligation kit was used for ligation. The ligation mix was transformed into Stellar^{TM} *E. coli* competent cells.

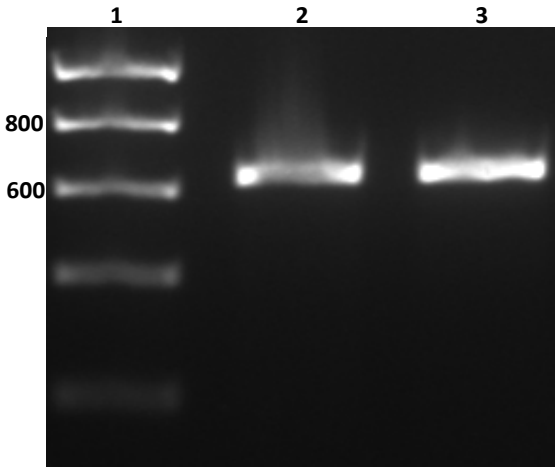

Figure 4.6. Gel electrophoresis of digested mutant Abyssovirus stop suppression site (TGA-C and TAA-A). Lane 1: Marker (1kb) DNA ladder, Lane 2 and 3: TGA-C and TAA-A digests with *Xho*l (594 bp) respectively.

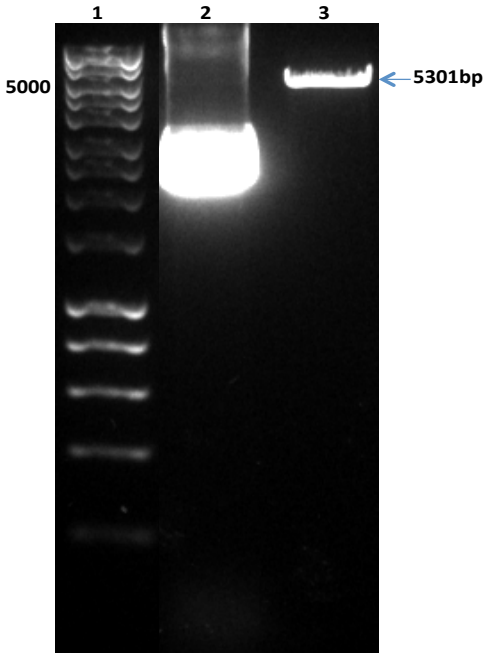

Figure 4.7. Double digest of pTriEx1.1. An agarose gel showing the digested pTriEx1.1. Lane 1: Marker (1kb) ladder. Lane 2: circle pTriEx1.1. Lane 3: pTriEx1.1 plasmid digested with *Xho*I restriction enzyme (5301bp).

4.2.2 Colony PCR screening of transformed bacteria

Colony PCR was performed using forward and reverse primers (Fw_TriexUp and Rv_TriexDown-**Table 4.3**) to check for the presence of the pseudoknot deleted versions of the Abyssovirus stop suppression sequences TGA-C and TAA-A and positive colonies with bands of the expected size were selected **Figure 4.8.**

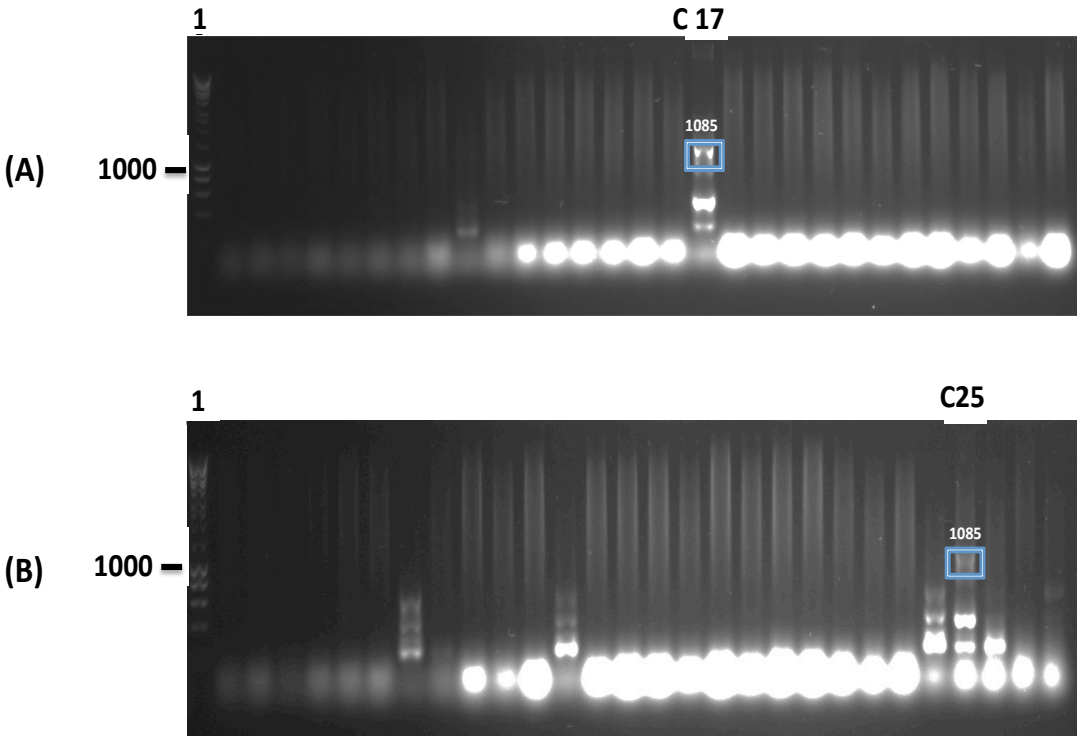

Figure 4.8. Mutants of Abyssovirus stop suppression sequences (TGA-C and TAA-A with no pseudoknot) screened by PCR. 27 colonies were screened for checking the presence of mutant stop suppression sequences (TGA-C (A) and TAA-A (B)) inserts. In both gels, Lane 1: Marker (1kb) DNA ladder, C17 in A and C25 in B: colonies showed positive results with correct sizes (1085 bp).

Positive colonies screened as above were grown at 37°C overnight in 10 ml LB broth with ampicillin and plasmid DNA extracted. The isolated DNA was sent for sequencing by SourceBioscience to confirm the relevant mutations in all cases and, once confirmed, a glycerol stock was prepared and the new strains stored at -80°C.

As a result of the cloning experiments, plasmids capable of the expression of wild type Abyssovirus stop suppression sequence and three mutants were isolated. The mutations were designed to investigate stop suppression function by changing the stop codon and 3' flanking nucleotide, TGA-C to TAA-A and, from each wild type and mutant, deleting the pseudoknot that occurs downstream. In all cases the fragments were flanked by epitopes for widely available antibodies, the HSV and HIS tags, so that the complete translation product representing either end of the fragment (one or other epitope) could be detected. These plasmids represent a resource for an investigation of the efficiently of both the stop codon identity and the pseudoknot on the proposed readthrough event.

4.2.3 Analysis of Abyssovirus stop codon readthrough signal

As described in Chapter 3, pTriEx1.1 was used as the vector for these studies as it allows the expression of target proteins in multiple hosts (*E. coli*, insect and vertebrate) and, as shown by the analysis of the Abyssovirus M^{Pro}, none of the hosts used appeared to have endogenous proteins that reacted with either the HSV or HIS tag antibodies, so the plasmid encoded product, along with any processed intermediates, should be clearly detected.

4.2.3.1 Abyssovirus TGA-C readthrough in *E. coli* BL21 (DE3)-pLysS strain

In chapter 3 successful expression of the Abyssovirus M^{Pro} was shown in *E. coli* but no activity was observed, in contrast to the results obtained in insect cells. However the stop suppression sequence was somewhat shorter and the encoded RNA was not expected to translate into any functional protein. Rather, RNA secondary structure signals were to be investigated. As discussed in the introduction to this chapter these are unlikely to be host specific and this could be confirmed if stop suppression activity was observed in *E. coli*. Accordingly in order to test the translation signals of the Abyssovirus readthrough stop suppression site, protein expression in the *E. coli* BL21 (DE3)- pLysS strain, was investigated. All plasmids characterized as correct by sequence analysis in this study were transformed into *E. coli* BL21 (DE3)- pLysS and cells plated on LB agar plates containing ampicillin. A colony selected randomly was grown in 10 ml LB broth with ampicillin (10 µg/ml) at 37°C overnight in a shaking incubator. From this culture, 200µl was transferred to a fresh 10 ml LB broth with ampicillin (100 µg/ml) and grown to OD600 0.4-0.6. Protein expression was induced by the addition of isopropyl-β-D-1 thiogalactopyranoside (IPTG) and allowed to continue for 4 hours post addition. Both before and after IPTG induction, aliquots of the bacterial culture were harvested for western blot analysis with the tag specific antibodies. The induction was successful in that new tag reactive proteins were clearly identified in the *E. coli* BL21 (DE3)- pLysS strain used following induction. The tag reactive bands were detected after one-hour induction with IPTG and became more intense with time **Figures 4.9 and 4.10**.

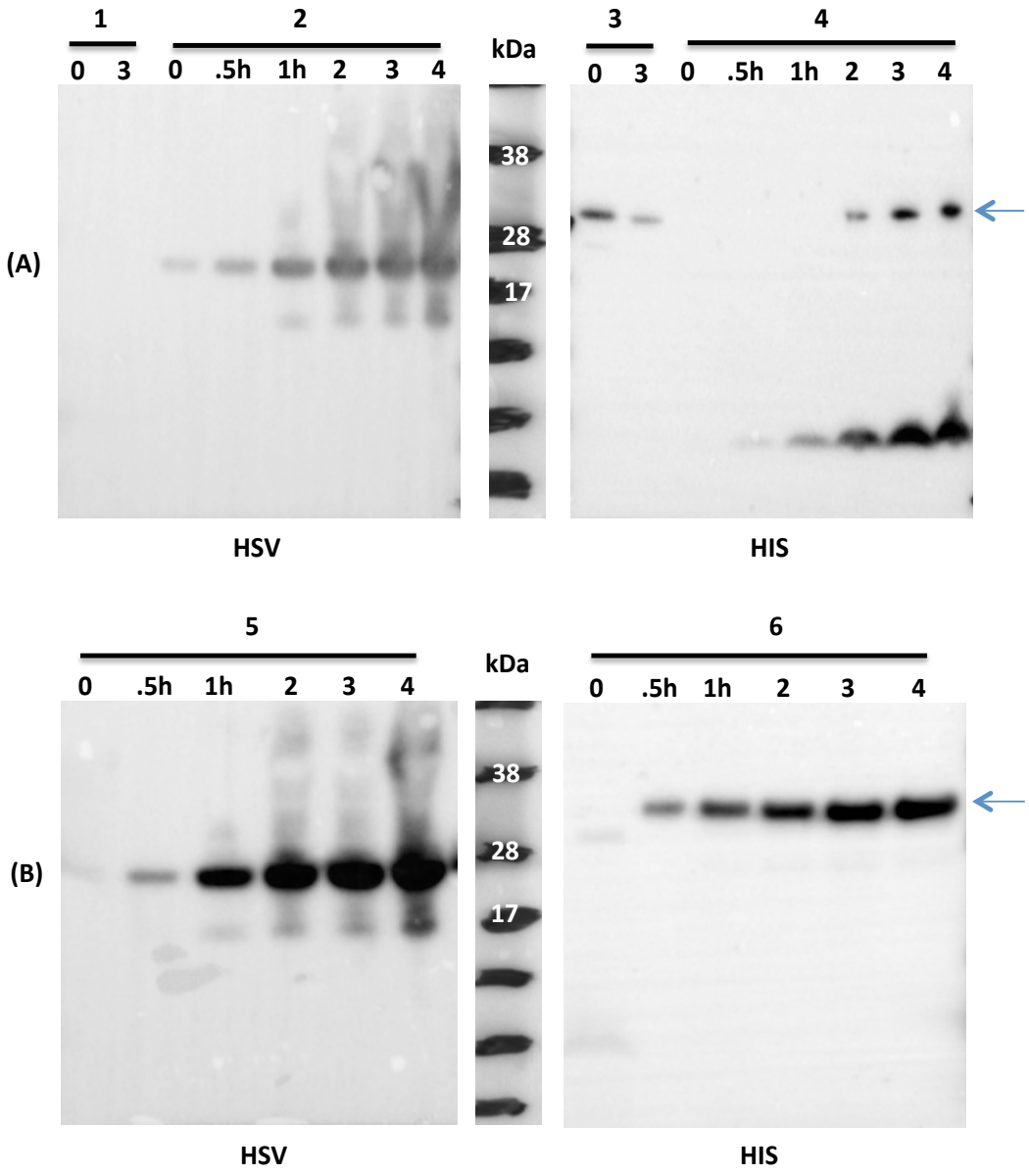

Figure 4.9. Western blot analysis of recombinant expression of the wild (TGA-C) and mutant (no pseudoknot) Abyssovirus readthrough stop suppression site signal proteins. Protein samples were extracted from *E. coli* BL21 (DE3)-pLysS prior to (non-induced; 0hr) and following induction with IPTG after 0.5 hour and hourly for 4 hours. Samples 1 & 3: Cells control. The two left membranes (Samples 2 and 5) were probed with the HSV antibody and the two right membranes (Samples 4 and 6) with the HIS tag antibody. The upper membranes are the wild type and the lower the mutant lacking the pseudoknot. The arrows represent the readthrough 32kDa.

In **Figure 4.9**, induction of expression from the original TGA-C and corresponding pseudoknot deleted mutant constructs gave a band with size ~25kDa using HSV antibody in both cases. When probed with the HIS antibody however the expressed TGA-C stop protein of the wild type construct showed a larger band of ~32 kDa which was reduced to ~30kDa for the no pseudoknot mutant.

Similarly, when the TAA-A mutant and its corresponding pseudoknot deleted variant were induced, harvested and probed in the same way, the same pattern and antibody binding and the same molecular weight of bands were apparent (**Figure 4.10**). In both figures it is assumed that the readthrough product is also present in the HSV blot. However, as the calculated readthrough efficiency is ~4% (1 in 25 molecules – see below) it fails to resolve adequately in exposures based on the primary, stopped, product. However, it is visualized unambiguously by the HIS antibody.

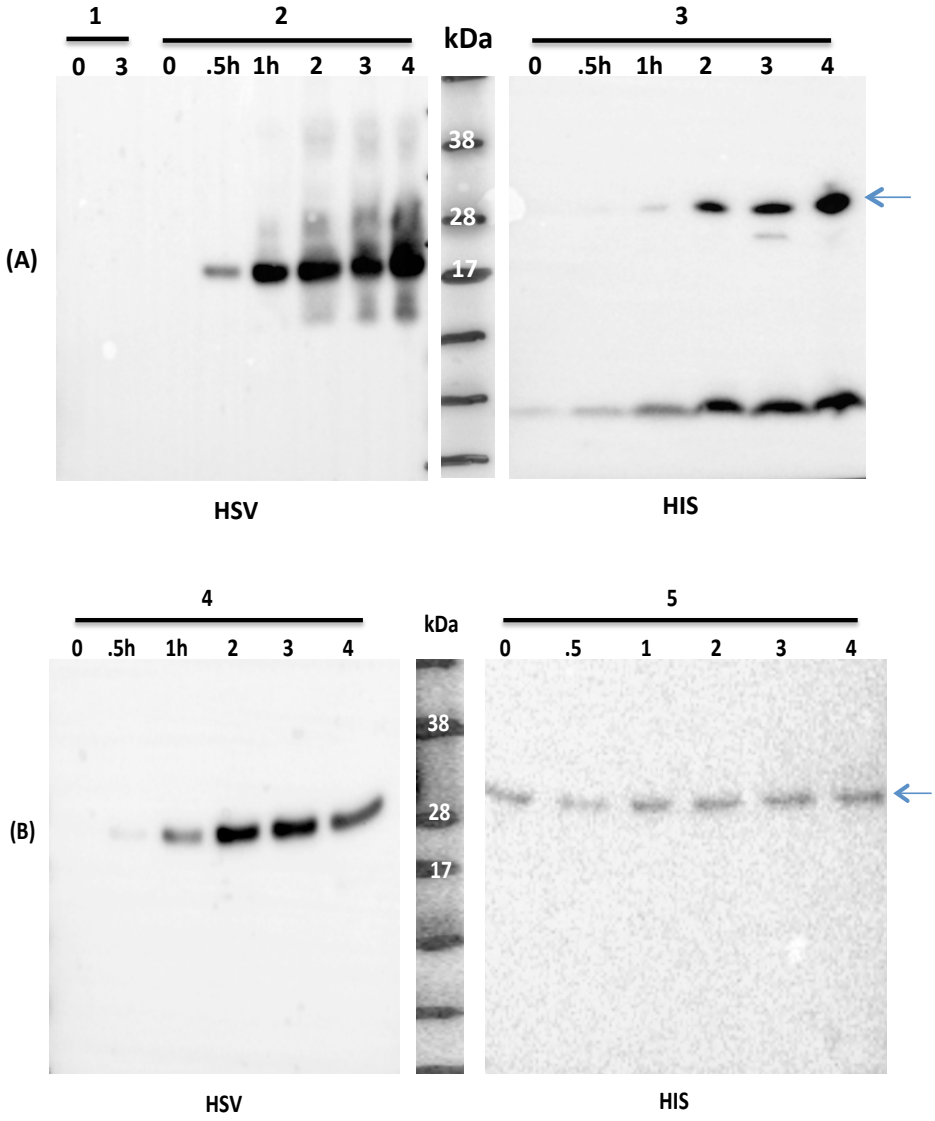

Figure 4.10. Western blot analysis of recombinant expression of the mutant stop suppression signal (TAA-A) and mutant TAA-A (with and without the downstream pseudoknot sequence) proteins. Protein samples were extracted from *E. coli* BL21 (DE3)-pLysS prior to (non-induced; 0hr) and following induction with IPTG after 0.5 hour and hourly for 4 hours. The two left membranes were probed with the HSV antibody and the two right membranes with the HIS tag antibody. The upper membranes are the TAA-A parent and the lower the mutant lacking the pseudoknot. The arrows represent the readthrough 32kDa.

In contrast to the results following expression in *E. coli* in chapter 3, the results obtained following expression of tagged reporters in the case of the Abyssovirus stop readthrough were encouraging in that clear bands at the predicted molecular weights were apparent for each of the constructs tested. Nevertheless, as it was interesting to assess if the patterns obtained were specific to one expression system or were universally observed, as might expected of signals that operate at the level of the mRNA. Therefore recombinant baculoviruses were also constructed with the same vectors for an assessment of stop-readthrough signals in insect cells.

4.2.3.2 Abyssovirus TGA-C readthrough expression in the insect Sf9 cells

The readthrough encoding plasmids, wild type and mutant, plus and minus the pseudoknot, were used to generate recombinant baculoviruses as described in Materials and Methods and Sf9 insect cells were infected at high multiplicity with the isolated viruses as described in Chapter 3. Briefly, 6 well dishes were seeded at near 100% confluence and infected with P3 virus stocks at an estimated MOI of 3 or greater. The plates were incubated at 28^{O}C for three days and the monolayers collected by dislodging the cells with a pipette, transferred to Eppendorf tubes and pelleted at 13,000 rpm for 3 minutes (microfuge). The supernatant was aspirated and the cell pellet lysed directly by re-suspension in 1 x SDS-PAGE loading buffer followed by heating at 100^{O}C for 10 minutes. The equivalent of 5 x 10^4 cells was loaded per well of a standard 10cm precast SDS polyacrylamide gel and electrophoresis was carried out for 30 minutes at 170V. Gels were then transferred to PVDF membranes by semi-dry blotter and the membranes processed for western

blot as described in Material and Methods using antibodies to the HSV or HIS tag. The tagged proteins expressed by each readthrough construct were detected by imaging the chemiluminescent substrate used for the anti-tag antibody conjugates on a Syngene XL 1.14 imager **Figure 4.11.**

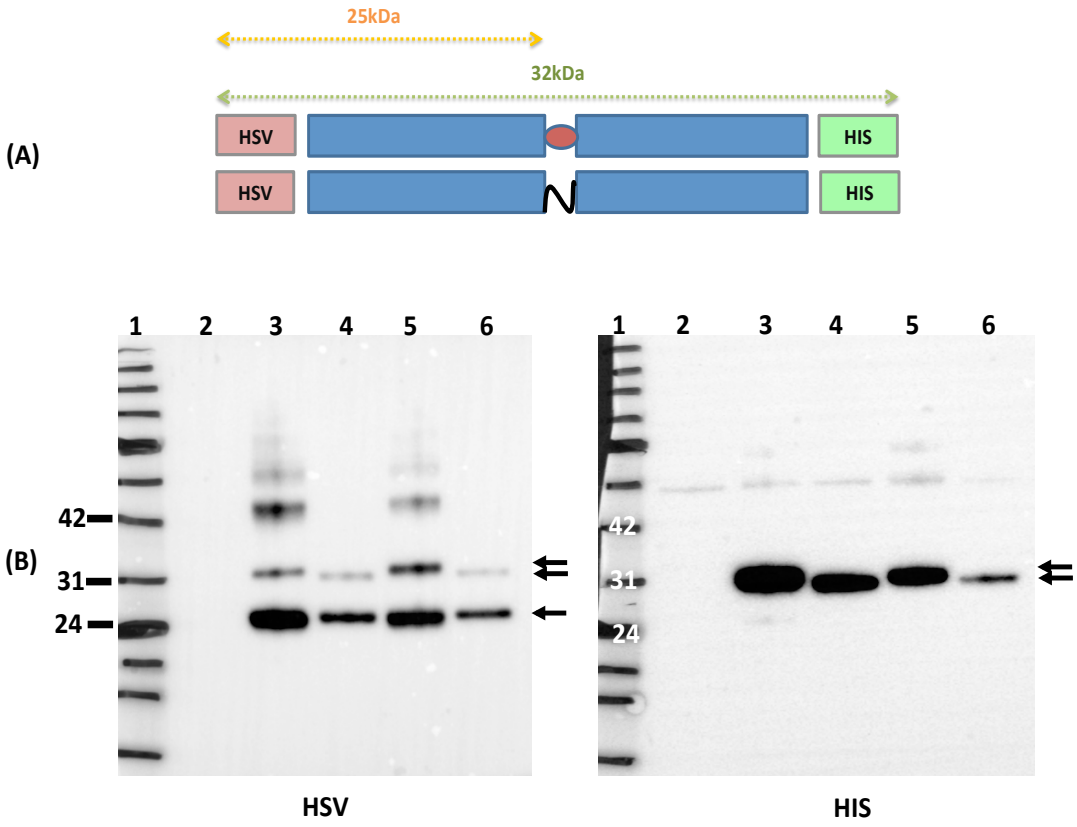

Figure 4.11. The Abyssovirus readthrough stop suppression signal wild type (TGA-C) and mutant proteins (mutant stop suppression signal (TAA-A) and mutant TGA-C and TAA-A (with no pseudoknot) expressed in Sf9 insect cells. **A:** Cartoon of the wild Abyssovirus and mutant constructions. **B:** Western blot of wild type and mutants readthrough stop-suppression signal, Molecular weight marker is indicated on the left of both HSV (left) and HIS (right) gels. In both gels, lane 2: GFP (26 kDa), lane 3: Abyssovirus stop suppression signal wild type (TGA-C), lane 5: mutant stop suppression signal (TAA-A) and the size of these samples (HSV size = 25kDa and HIS size = 32 kDa), lanes 4 and 6 are mutant TGA-C and TAA-A (with no pseudoknots) the sizes of both proteins are HSV size= 25 kDa and HIS size 30.46kDa.

For the wild type readthrough stop-suppression fragment, TGA-C, and the mutated codon and 3' nucleotide, TAA-A, the results of the western blot analysis were the same and indicated the production of a predominant band with apparent molecular weight of 25kDa when blotted with the HSV-antibody (left gel - lanes: 3 and 5) and a more minor band of 32 kDa along with some higher molecular weight products. When blotted with the HIS-antibody (right gel - lanes: 3 and 5) only the 32kDa band was visualized. For the pseudoknot deleted variants, the HSV blot showed a predominant band at 25kDa whose migration was slightly altered by the introduced deletions but a lower molecular weight band which now migrated at ~30.5 kDa as a result of the deletions made (left gel -lanes: 4 and 6 respectively). When blotted with the HIS-antibody (left gel -lanes: 4 and 6) only the larger product was detected at 30.5 kDa.

4.2.3.3 Expression of Abyssovirus readthrough stop suppression site signaled in mammalian cells

As in chapter 3 an attempt was made to also express the stop–readthrough constructs in mammalian cells. To do this HEK 293T 17 clone-1 mammalian cell lines were used and cultured in Dulbecco's modified Eagle's medium (DMEM). Transfections of the protein expressing constructs were carried out in 6 well plates using monolayers of cells and Lipofectamine3000 in accordance with the manufacturer's instructions. To determine the optimal conditions for plasmid transfection in this cell line, a control vector pTriEx1.1-GFP, which carries the Green Florescent Protein (GFP) gene inserted between the *Nco*I and *Bsu*361 sites in pTriEx1.1 was used to visualize the efficiency of transfection (O'Flynn, 2012). As for the data reported in chapter 3 however no western blot positive bands were

obtained at the sizes predicted for the encoded products. Reaction with the HSV antibody was observed with a band of ~30 kDa but as this was apparent in all lanes it was clearly reaction of the antibody to a non-specific cellular protein and not related to the samples tested.

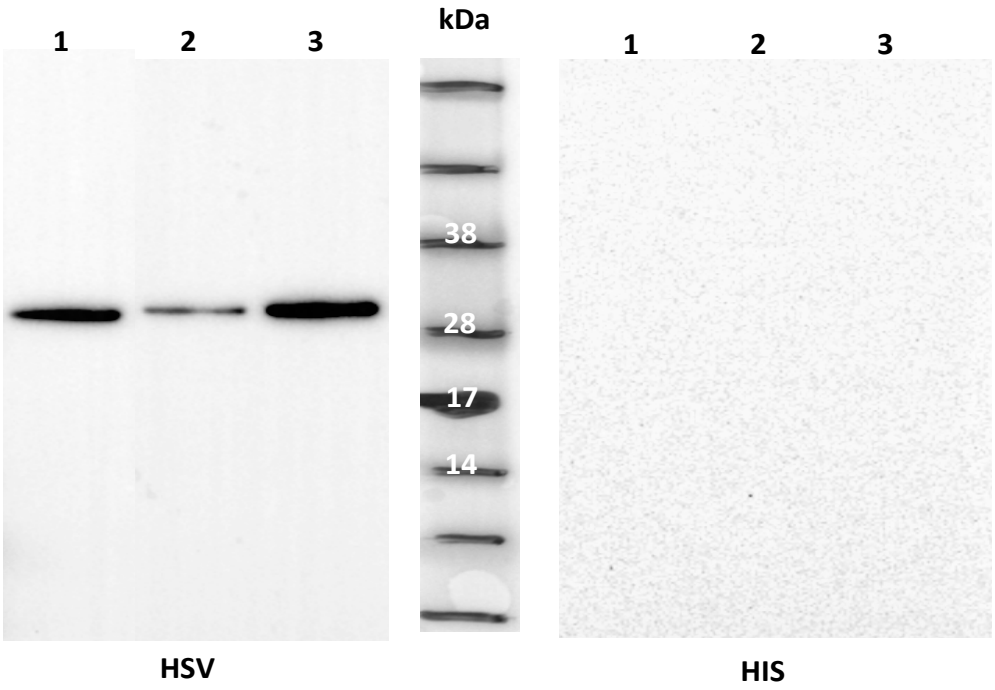

Figure 4.12. Abyssovirus stop suppression site signal proteins expressed in mammalian cells. Transfected cells were harvested after 24 hours cells and total cell extracts were prepared. Western blots used anti HSV (left side) and HIS tag (right side). Molecular weight marker in kDa is indicated on the middle. In both blots, lanes 1 - pTriEx1.1-GFP as a control. Lanes 2 -wild readthrough (TGA-C). Lanes 3 - mutant readthrough (TAA-A).

The lack of product in Western blot was undoubtedly due to poor expression as a result of limited transfection efficiencies coupled with the use of a non-replicative vector which limited the gene dosage, very different from the baculovirus system for example. Although a number of different transfection reagents such as Polyethylenimine (PEI) were assessed and post harvesting steps were modified, e.g. harvesting on ice to reduce protease enzyme activity prior to analysis, no bands for any of the test proteins were detected in subsequent transfections. In some cases, GFP expression was detected as green cells by UV microscopy and by Western blot but the levels were low and as a result analyses of expression in mammalian cells was not pursued further.

4.3 Discussion

This chapter described the successful cloning of the Abyssovirus suggested stop-suppression signal sequence (TGA-C) located at the junction of the 1a and 1b polyproteins in a configuration that would allow readout of its function. To assess the role of the native sequence, a variant in which the stop sequence was mutated to a different stop-suppression sequence (TGA-C to TAA-A) was also constructed. In addition, the role of the pseudoknot, found downstream of the suppression site, was addressed with the cloning of two new stop-suppression mutant samples (TGA-C and TAA-A with no pseudoknot). All the Abyssovirus stop-readthrough constructs were made in the pTriEx1.1 vector using the In-Fusion cloning technique combined with the overlap PCR method. The general design used was to generate clones capable of simple detection in that the various sites to be tested were flanked by

expression tags, the HSV tag and HIS tag, allowing the detection of fusion proteins by western blot with the requisite antibodies. All the constructs designed were shown to be correct by sequence analysis and were subsequently introduced into different expression systems to assess stop-readthrough function in different genetic backgrounds. These included *E. coli* (BL21 (DE3)- pLysS strain), insect cells (Sf9) and mammalian cells (17clone1). Through the use of these constructs the function of the predicted unusual translational stop-readthrough signal located at the 1a/1b junction in the Abyssovirus replicase gene was assessed for the role of the stop codon and immediate downstream nucleotide and the effect of the pseudoknot on the translation process. The choice of stop codon and elements of the two codons that follow have been shown to affect the efficiency of translational termination in other cases (Cridge *et al.,* 2018; Skuzeski *et al.,* 1991).

The second step in coronavirus lifecycle is the translation of the replicase gene, from the virion genomic RNA, which encodes two large open reading frames ORFs 1a and 1b that produce two polyproteins pp1a and pp1ab. All coronaviruses studied to date use a frameshift mechanism to produce suitable ratios of these two important polyproteins which is a combination of elements including a slippery sequence (5′-UUUAAAC-3′) and an RNA pseudoknot that stimulates the ribosomal frameshifting event from the ORF1a reading frame into ORF 1b. Generally, the ribosome unwinds the pseudoknot structure, and continues translation until it reaches the ORF 1a stop codon. However, sometimes the pseudoknot prevents the ribosome from continuing and the pause causes the ribosome to slip on the slippery sequence, shifting the reading frame back one nucleotide, a 1- frameshift, so that continued translation occurs now in ORF 1b leading to the expression of pp1ab (Baranov *et al.,* 2005, Brierley *et al.,* 1989). Based on this precedent, it may be

suggested that the pseudoknot found in the Abyssovirus sequence would act in a similar way to stimulate the readthrough event. A second question to be addressed was if any readthrough observed was limited to any one expression background or if it occurred irrespective of the type of ribosome, prokaryotic or eukaryotic, used for the translation.

Following their introduction into the various expression systems, HSV and HIS-tagged proteins were detected at various level of expression. All proteins were expressed in *E. coli* BL21 (DE3)-pLysS successfully by inducing protein expression with IPTG. The predicted initial protein size is ~25kDa and the readthrough was ~32 kDa for the wild types TGA-C and TAA-A respectively and ~30 kDa for the mutant TGA-C and TAA-A proteins lacking the pseudoknot. Reaction with both tag antibodies was apparent and the HSV/HIS-tagged proteins were detected at the expected molecular weights by Western blot for all cloned constructs after one hour of induction. The levels thereafter were variable, perhaps indicating minor issues of protein stability but the data obtained were consistent with termination at either stop codon (TGA or TAA) to generate a ~25 kDa fragment which contains the HSV tag and a proportion of readthrough to generate the ~ 32kDa band which contains both HSV and HIS tags. Given the HIS antibody can only detect the readthrough product, the presence of the HIS reactive band confirms the readthrough event but cannot be used to quantify the rate of readthrough. However, the HSV tagged bands include the initial termination product as well as the readthrough product and present the opportunity to quantify the readthrough event. The clarity of the western blots from *E. coli* was generally less than those from insect cells, however, considering the relative intensity of the HSV detected bands in the *E. coli* lysates, the efficiency of the stop suppression

signal was poor, around 10% and appeared not to be affected by the nature of the codon or the following nucleotide. That is, the relative band intensities were similar for constructs TGA-C and TAA-A. When the pseudoknot deleted variants of each construct were assessed, a readthrough band was observed as ~30kDa consistent with the deletions made (pseudoknot = 42bp) and, as the pseudoknot is downstream of the stop codon, the change of mass is only relevant to the readthrough product. When compared to their parental sequences the rate of readthrough appeared little altered by the loss of the pseudoknot sequence (compare Samples 4 to 6 and 3 to 5 on the HSV blot **Figures 4.9 & 4.10** respectively) although formal quantification was not performed with the *E. coli* samples.

Following the same analysis in insect cells however, the data was confirmed and the signal intensity was such that an accurate quantitation could be made. The data presented in **Figure 4.11** showed all protein expressed in insect cells (Sf9) with the same bands identified as in the *E. coli* lysates. That is, the two observed translation products are present, the predicted termination and readthrough products at ~25 and ~32 kDa respectively. Lanes 3 and 5 present Abyssovirus wild stop suppression signal (TGA-C) sequence located in the reading frame and mutant stop suppression (TAA-A) respectively which showed similar results of both proteins with obvious predicted termination products in the HSV blot (25 kDa) and readthrough products with clear bands in the HIS blot (32kDa). Lanes 4 and 6 present in both Abyssovirus wild stop suppression signal (TGA-C) and mutant stop suppression (TAA-A) respectively but deleted for the pseudoknot located downstream of the stop suppresser. Both samples with deletions showed similar results compared with the original samples (TGA-C and TAA-A). In the HSV blot, the predicted termination product size is ~25 kDa and on the HIS blot the readthrough product

size was ~30.5kDa after deleting the pseudoknot. Together the data show readthrough of the stop codon at the 3' end of the Abyssovirus 1a coding region in insect cells that was not affected by the nature of the stop codon or the presence of the downstream pseudoknot. The western blot signals for insect cell expression were sufficiently clear to allow densitometric analyses of the lanes. To do this the blot image was imported into ImageJ software (Schneider *et al.,* 2012) and the bands identified as the primary and readthrough translation products were isolated and their pixel content determined. The values obtained were compared as a measure of readthrough rate. Following this analysis, the expression of wild type and mutant stop suppression Abyssovirus in the insect cells (TGA-C and TAA-A) was estimated that 4.65-5.27% of translation events resulted in readthrough and 9.42-12.58% for the same constructs when the pseudoknot was deleted. **Figure 4.13** shows the quantitation of readthrough protein expression in the wild types TGA-C and TAA-A and the mutant TGA-C and TAA-A proteins lacking the pseudoknot.

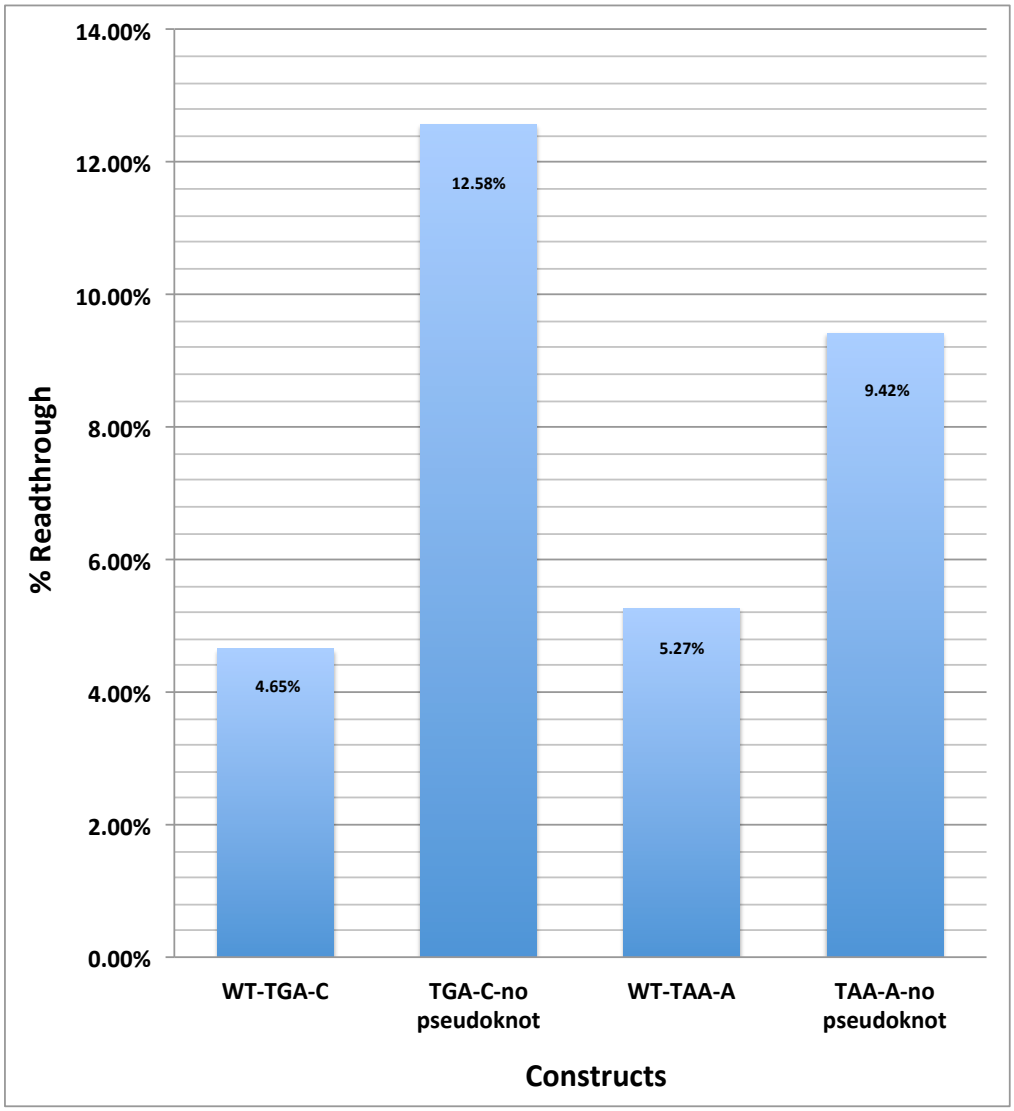

Figure 4.13. Percentage of Abyssovirus readthrough proteins expression in the insect cells Sf9. The data used for the ratio calculation was the average of two experiments.

By contrast with the data obtained in insect cells (Sf9) and *E. coli* BL21 (DE3)-pLysS, transfections of the same plasmids into the mammalian cell line 17clone-1 showed no bands for any of the proteins (**Figure 4.12**) consistent with low transfection frequencies. It was not pursued further as both the *E. coli* and insect cells had produced similar data.

Protein translation usually terminates at one of the three stop codons: UAG, UGA, and UAA. However, protein translation termination can sometimes be leaky, and the translation process may continue until a second stop codon is reached. Therefore, an extended polypeptide can be produced. This phenomenon is known as translational readthrough (Li and Rice, 1993) and has been reported as a mechanism for some viruses. For instance, the translation process of Sindbis virus (SIN) and other alphaviruses require readthrough of a UGA codon in order for synthesis of the viral RNA-dependent RNA polymerase (Firth *et al.,* 2011) and such readthrough of the UGA has been reported to required a single cytidine located immediately downstream of the stop codon. Furthermore, it has been demonstarted that in Murine Leukemia Virus (MuLV), belonging to the *retroviridae* family, 5-10% of the ribosomal translation processes readthrough the stop codon (UAG) (Csibra *et al.,* 2014), and readthrough results from the inclusion of an arginine, cysteine or tryptophan in place of the stop codon (Attoui *et al.,* 2002). A mutational study of Colorado Tick Fever Virus (CTFV), a member of the *Reoviridae* family, discovered that the readthrough process efficiency of a UGA codon is influenced by the sequence around the stop codon (Napthine *et al.,* 2012). In these published examples readthrough rates were between 1 and 10% broadly in keeping with the rates observed here (Li and Rice 1993; Napthine *et al*., 2012). In the data obtained here the relative readthrough rate was approximately 2 fold higher

in both TAA and TGA constructs when the pseudoknot was deleted, consistent with a role for the pseudoknot in attenuating pp1ab synthesis in Abyssovirus infections. This may be consistent with the pseudoknot pausing the ribosome long enough for it to ensure disengagement rather than hunt for a closely matching tRNA.

Together these results indicate that the pp1b region of Abyssovirus is plausibly expressed by readthrough of the stop codon UGA, mediated by a functional termination-suppression signal whose efficiency is partly dependent on a pseudoknot sequence following the stop codon. The readthrough event occurred in both prokaryotic and eukaryotic cells consistent with a purely *cis* effect not dependent on *trans* acting factors present in one cell type or another. It is possible this sequence could be used practically for the regulated expression of downstream sequences in recombinant systems where small amounts of an additional reading frame may be required although this application was not investigated here.

Chapter 5

5. The biological function of Abyssovirus Structural protease

5.1 Introduction

Work in previous chapters has demonstrated that novel features found in the Abyssovirus sequence, M^{Pro} like protease and a stop readthrough signal, were functional in some expression systems. This data supports the contention that Abyssovirus is an extant organism but the data also validate the approach taken to demonstrate biological activity. That is, cloning the sequences concerned into a relevant reporter construct followed by analysis in a variety of cellular backgrounds. In this chapter the same approach is taken to look for activity associated with a novel chymotrypsin-like serine protease found by bioinformatics analysis of pp2. Unusually, the protease has homology with the alphavirus capsid protease (Melancon and Garoff, 1987), but until now this type of protease has not been described as being present in any other nidovirus genome and, assuming a role similar to that of the alphavirus CP (see below), it has been termed the structural protease (S^{Pro}).

The alphavirus genomic RNA contains two open reading frames (ORFs) that encode both a virus non-structural polyprotein and a structural polyprotein. During replication, a 26S subgenomic RNA is responsible for producing the structural polyprotein which encodes the capsid protein (CP), E3, E2, 6K and E1 (**Figure 5.1)** (Aggarwal *et al.,* 2015).

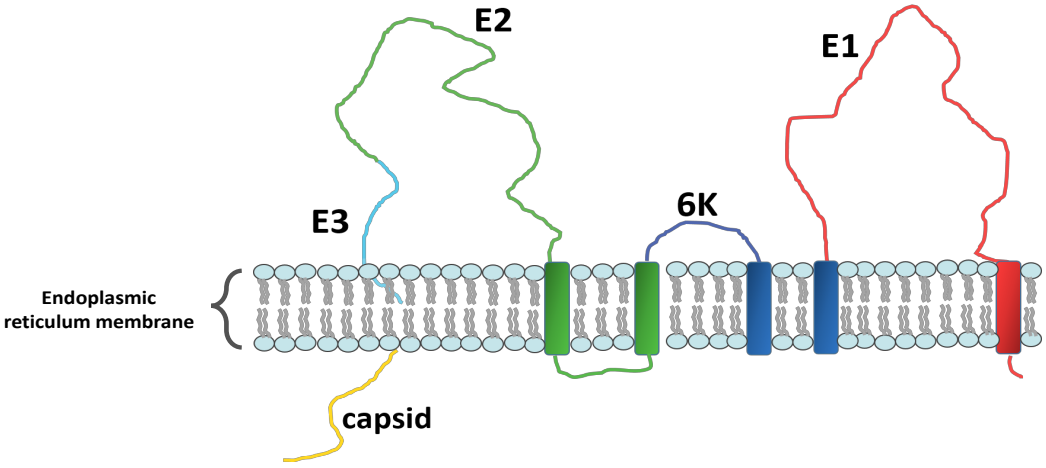

Figure 5.1. Sindbis Virus, one of alphavirus members, structural proteins (capsid (yellow), E3 (light blue), E2 (green), 6K (dark blue) and E1 (red)) and their positions through an endoplasmic reticulum membrane (adapted from Li *et al.,* 2010).

The alphavirus CP is essential for the viral life cycle and is released by a serine protease auto-cleavage activity inside the capsid (Choi *et al.,* 1996). The protease domain involves a β-barrel motif with a typical catalytic triad, serine-histidine-aspartic acid, and the active site that is split between the two subdomains (Choi *et al.,* 1997, Owen and Kuhn, 1997). Changing the serine in the catalytic triad to arginine in Semliki Forest Virus (SFV), one of the alphavirus members, stops capsid protease activity (Melancon and Garoff, 1987). CP activity is one of the first cleavage events in SFV polyprotein processing and is carried out by the CP serine protease domain which recognizes a tryptophan-serine cleavage site in the polyprotein and releases the capsid protein. As a result, the alphavirus structural protein has been the subject of a significant number of studies (Choi *et al.,* 1997,

Thomas *et al.,* 2010), the majority with SFV and all analysis of alphavirus CPs to date have confirmed the autocatalytic cleavage between tryptophan and serine residues although some activity can be retained by replacement of the Trp with Phe or Tyr (Skoging and Liljestrom, 1998). Although this activity has not previously been described for nidoviruses, inspection of the pp2 amino acid sequence by the EXPASY protein translation tool showed the presence of 32 tryptophan–serine di-residue motifs within the frame which would be consistent with a potential activity acting at these sites assuming the specificity remained unaltered. However, as the role of S^{Pro} in nidovirus processing is novel, the question of activity is debatable. That is to say, while assessment of the activity of the M^{Pro} and the stop readthrough in previous chapters were seeking to validate an accepted nidovirus feature, albeit not the norm in the case of the stop readthrough function, no expectation can be associated with S^{Pro} as there is no precedent for this activity in other family members. The aim of this chapter therefore was to generate constructs capable of the expression of the Abyssovirus S^{Pro} and assess its potential function as a protease. To confirm this, if positive, the wild type sequence was also mutated at the predicted catalytic triad serine-histidine-aspartic acid motif. As before, the pTriEx1.1 vector was used allowing expression of the potential protease in different cell backgrounds including insect cells (Sf9) and *E. coli,* to maximise the possibility of finding a background compatible with activity.

5.2 Results

5.2.1 Construction of tagged Abyssovirus structural protease vectors

The Abyssovirus S^{Pro} sequence as defined in the TSA database sequence from the sea hare *Aplasya californica* (accession no: GBDA01003924.1) was ordered as a synthetic DNA fragment of 1266bp and was provided by Dr. Geraldine Mulley. The annotated amino acid sequence of the fragment is shown in **Figure 5.2.** To provide tags for the facile detection of the translated product the synthetic fragment was cloned by in-Fusion technology into the vector pTriEx1.1 to allow the test of different cellular backgrounds on the activity of the translated product. In-fusion adds specific 15 bp overlap sequences to the termini of the DNA fragment to be cloned which are homologous to specific sequences in the linearized vector and when mixed with the correct enzymes are able to undergo sequence specific recombination *in vitro* (Jacobus & Gross, 2015). As previously, the vector was linearized with *Xho*I restriction enzyme and mixed with the flanked synthetic DNA fragment in a design that would allow expression of S^{Pro} with an N-terminal HSV and a C-terminal HIS tag. The 15 bp overlap sequences added to the synthetic DNA by PCR amplification and are shown with vector and structural protease homologies identified (**Table 5.1**).

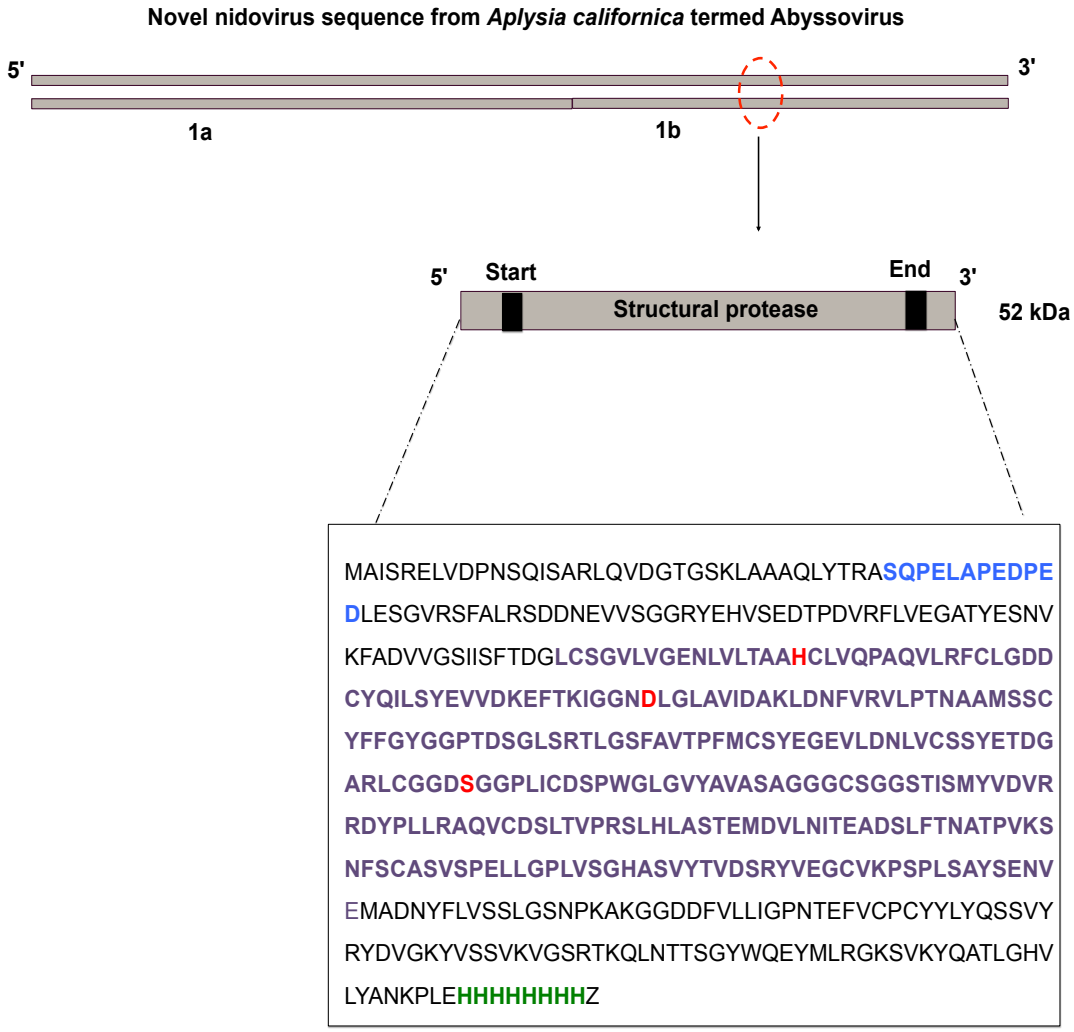

Figure 5.2. The genome of Abyssovirus and the location of the structural protease. The protease construct order is as following: HSV tag – pre-protease sequence - structural proteinase – post protease sequence – HIS tag. The protease triad residues are present in red.

Table 5.1. Oligonucleotides used for the in-Fusion cloning of the Abyssovirus structural protease in pTriEx1.1. Yellow color represents vector sequence.

Primer	Sequence	bp
Fw_ tructuralPro_Infu	5' CCCCGAGGATCTCGAGTCTGGTGTAAGG TCTTTTGC ' 3	36
Rv_ tructuralPro_Infu	5 ' GATGGTGGTGCTCGAGCGGTTTGTTC GCATACAGA ' 3	35

After transformation of the infusion mix into Stellar^{TM} *E. coli* competent cells and overnight incubation on selective agar, 6 colonies were selected at random, transferred onto LB with ampicillin and grown overnight at 37°C with shaking. Plasmid DNA was extracted from the 6 candidates and analyzed by PCR using the primers used for the generation of the infusion fragment (**Table 5.1**). DNA which amplified to give the correct size of insert (1298 bp) was sent to Source BioScience for DNA sequencing with the T7 forward primer to confirm the identity of the sequence cloned. All sequences obtained were as designed with no evidence of mutations and the reading frame was open throughout the sequence encoding each of the elements described above **Figure 5.2**.

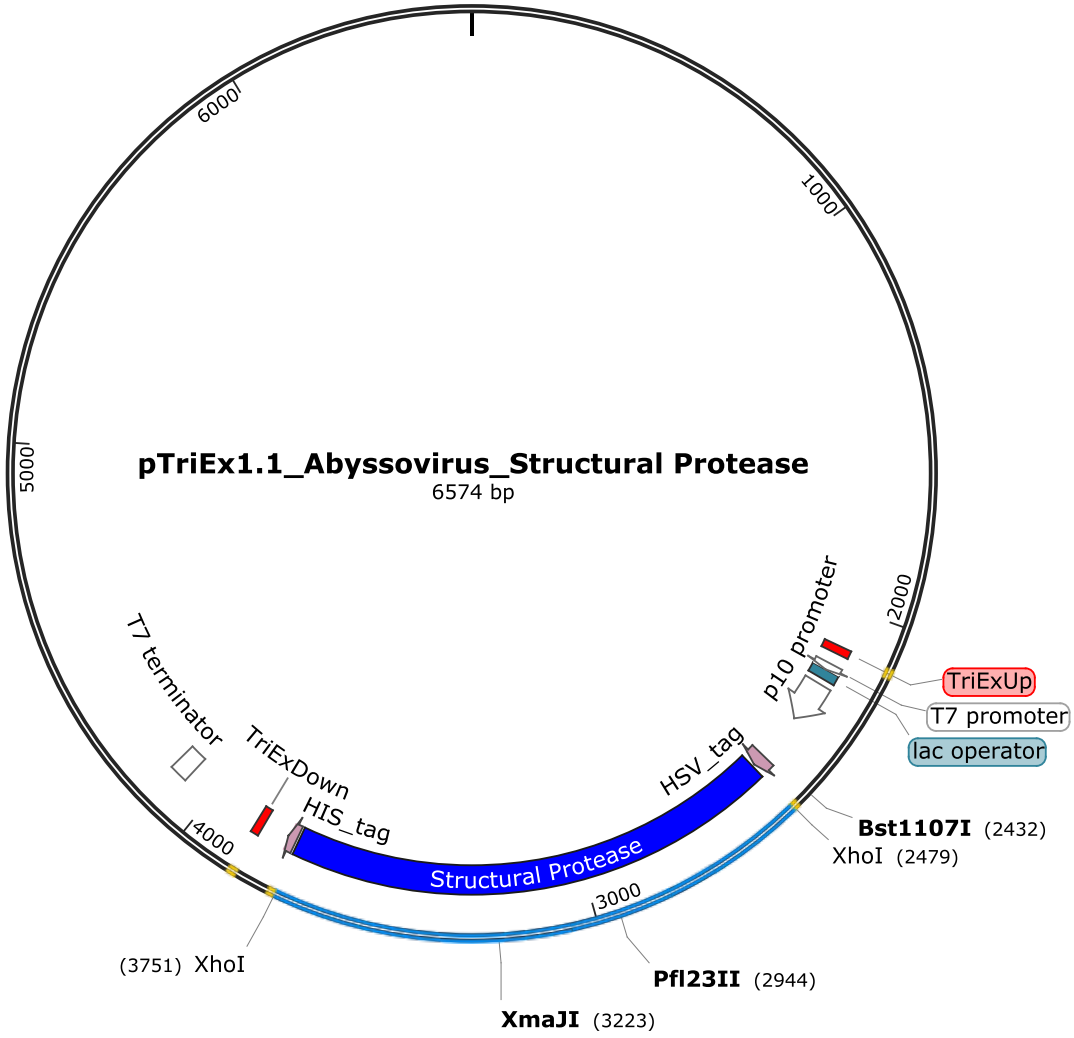

Figure 5.3. Plasmid map showing the complete Abyssovirus structural protease sequence (blue) in the pTriEx1.1 vector. Note the unique *Xho*I site which flank the sequence of the wild type Abyssovirus structural protease. *Bst*1107I, *BsiW*I and *XmaJ*I sites used for cloning of the mutant fragments, as well as the HSV and HIS tags used for protein detection.

5.2.2 Generating of mutant Abyssovirus structural protease constructs

Catalytic triads are well studied among all serine proteases including alphavirus CP (e.g. in Aura virus (AVCP)) and show similar spatial architectures in the active site (Choi *et al.,* 1991, Aggarwal *et al.,* 2012, Hahn and Strauss, 1990). Mutation of any of the capsid autoprotease active site residues leads to a block in the structural protease catalyzed processes (Steel and Geiss, 2015). The equivalent residues in the Abyssovirus structural protease were identified by alignment and targeted for mutagenesis to test the role of these resiudes in the activity of the Abyssovirus S^{Pro}. To determine their relative impact the three active site residues (Serine 243, Aspartic acid 163 and Histidine 126) were mutated to Alanine, each alone and in combination.

Two methods were used to create the various mutants, the exchange of a synthetic DNA fragment containing the mutations for the resident wild type sequence, and overlap PCR. The mix of methods arose as the cost of synthetic DNA became much lower over the course of the project and its use began to replace other PCR based methods when a strategy to replace the target region of the mutation was obvious, essentially if the region was flanked by unique restriction sites. For the synthetic exchange described here, three individual sequences were ordered as synthetic DNA fragments, each of which contained a mutated residue Histidine to Alanine, Aspartic acid to Alanine and all active sites mutants (Serine, Histidine and Aspartic acid to Alanine). Each synthetic DNA fragment was flanked with two unique enzymes, *Bst*1107I and *BsiW*I, for the Histidine and Aspartic acid mutant fragments and *Bst*1107I and *XmaJ*I for the mutation of all active site residues (see **Figure**

5.3). To enable the cloning, the vector was digested with the same restriction enzymes and the large vector band now lacking the region to be replaced was isolated. The gel eluted band was mixed with the synthetic DNA fragment and a T4 ligation kit was used for their ligation according to the manufacturer's instruction. The ligation mix was transformed into Stellar^{TM} *E. coli* competent cells and a separate transformation with a small amount of non-digested plasmid acted as a control to check the efficiency of transformation.

5.2.3 Screening of transformed bacteria using *Xho*I restriction enzyme

Six colonies from each transformation were selected randomly and grown at 37°C overnight in 10 ml LB broth with ampicillin and plasmid DNA extracted. The isolated DNAs from each sample was digested with *Xho*I enzyme to check for the presence of the mutant Abyssovirus structural protease gene and the positive colonies with bands of the expected size (1272bp) were selected (**Figure 5.4**) and sent for sequencing to confirm the relevant mutation and, once confirmed, a glycerol stock was prepared and the new strain stored at -80°C.

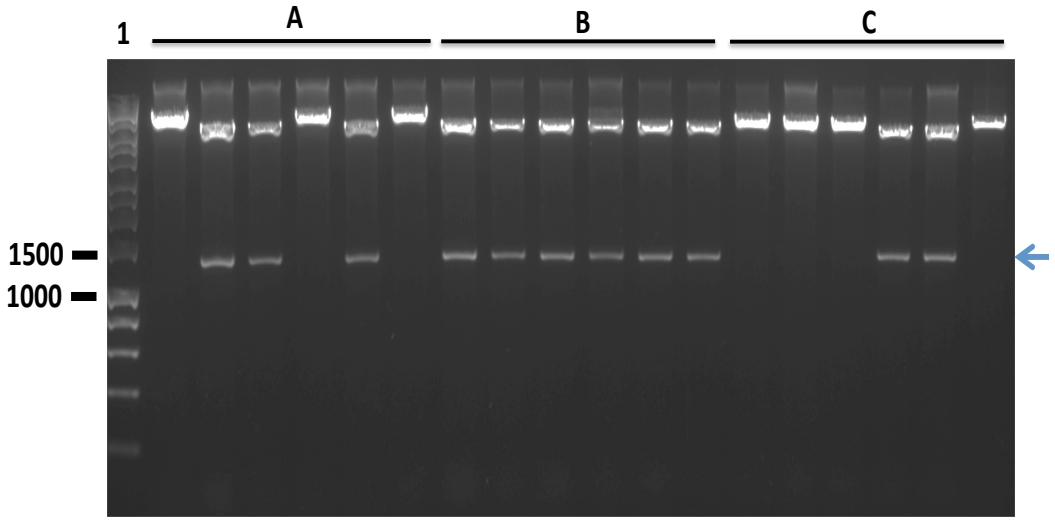

Figure 5.4. Gel electrophoresis of different mutants in pTriEx1.1 plasmid digested with *Xho*I enzyme. Lane 1: Hyperladder, **A:** Structural protease first active site mutation (H126A). **B:** Structural protease second active site mutation (D163A). **C:** Structural protease all active site mutations (Serine, Histidine and Aspartic acid to alanine (H126, D163 & S243 to A). The correct size of all mutant structural protease cases is 1272bp (blue arrow).

5.2.4 Generating of mutant Abyssovirus Structural protease by using overlap PCR method

The overlap PCR method (see Abyssovirus main protease chapter 3-**Figure 3.5**) was used to generate the third active site mutation of the Abyssovirus structural protease (Serine243 → Alanine), by designing additional forward (Fw_) and reverse (Rv_) primers **Table 5.2.**

Table 5.2. Oligonucleotides used for mutation of the S^{Pro} gene in pTriEx1.1 plasmid. Highlight sequence represents mutations.

Primer	Sample	Mutant	Sequence	bp
Fw_StruPro_ Third	Third active site	TCT→GCT	5'-CGTTTGTGTGGAGGTGA CGCTGGTGGCCCATTGA-3'	34
Rv_StruPro_ Third			5'-TCAATGGGCCACCAGCGT CACCTCCACACAAACG-3'	34

In the case of this mutation, the wide type structural protease DNA, cloned as described above, was used as template and amplification by PCR used a forward primer (Fw_TriExUp, **Table 5.3)** which primed 5' to the structural protease sequence and a reverse primer which included the mutation to be introduced in each case (Rv_ StruPro_Third see **Table 5.2**) in reaction PCR1. Another PCR was carried out with a forward primer which included the mutation to be introduced (Fw_StruPro_Third see **Table 5.2**) and a reverse primer (Rve_TriExDown, **Table 5.3**) which hybridized to a sequence 3' to the cloned structural protease sequence, reaction PCR2. The PCR products of these first round PCR reactions for the active site mutation (S243A) was analyzed by agarose gel electrophoresis, and the correct sizes (PCR1: 963bp and PCR2: 809bp) were obtained for both PCR1 and PCR2 respectively **Figure 5.5.**

Table 5.3. Oligonucleotides used to generate mutant structural protease case.

Primer	Sequence	bp
Fw_TriExUp	5'-GGTTATTGTGCTGTCTCATCA-3'	21
Rve_TriExDown	5'-TCGATCTCAGTGGTATTTGTG-'3	21

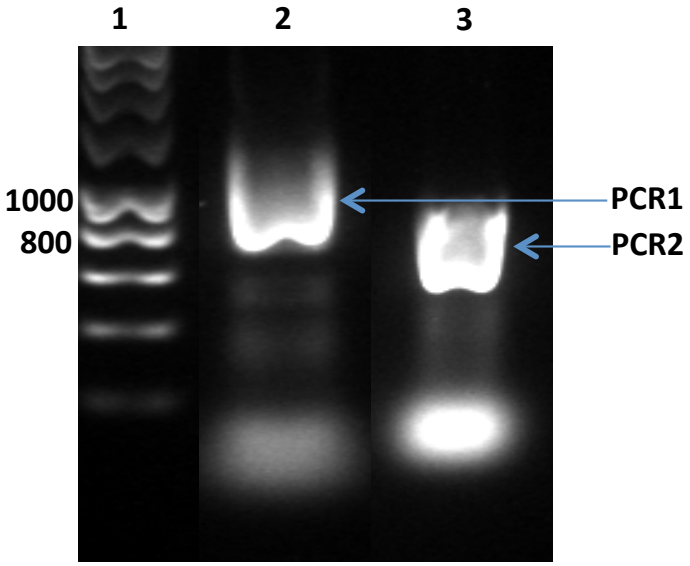

Figure 5.5. Gel electrophoresis of PCR products following amplification of the Abyssovirus structural protease gene to create active site mutation (S243A) using specific overlap primers. Lane 1: Marker (1kb) DNA ladder, Lane 2: PCR1 (963 bp), Lane 3: PCR2 (809 bp).

The amplified products were purified from the gel using a PCR purification kit and quantified by Nanodrop spectrophotometer. Then, 0.1ng/bp from of PCR product was mixed and subjected to amplification in a PCR reaction for five cycles but without any primer addition. Then, forward primer (Fw_TriExUp) and reverse

primer (Rv_TriExDown) were added and a further thirty cycles were completed. The correct size 1803 bp of overlap PCR product was analyzed as before by agarose gel electrophoresis and the desired overlap band was extracted by gel extraction kit **Figure 5.6.**

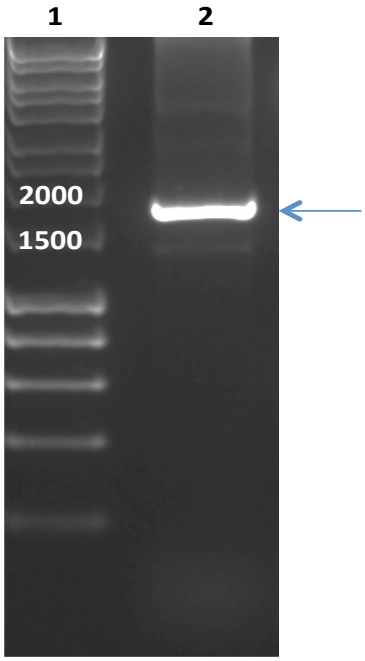

Figure 5.6. Agarose gel electrophoresis of overlap product of Abyssovirus structural protease mutation amplified by PCR. Lane 1: Marker (1kb) DNA ladder, Lane 2: PCR1+PCR2 for Abyssovirus structural protease mutation S243A (1772bp).

The purified product was quantified by Nanodrop spectrophotometer and digested with *Xho*I restriction enzyme for 30 minutes at 37°C and then a clean up kit was used to re-purify the product. PTriEx1.1 plasmid was also digested with *Xho*I restriction enzyme. In order to prevent the self-ligation of the vector, as the vector was digested with a single restriction enzyme, FastAP thermosensitive alkaline

phosphatase treatment was used to remove the phosphate group from the 5' ends of the vector by incubation for 10 min at 37°C. The reaction was then incubated for another 5 min at 75°C to inactivate the enzyme. The linearized vector was mixed with the overlap mutant fragment and a T4 ligation kit was used for their ligation according to the manufacturer's instruction. The ligation mix was transformed into Stellar^{TM} *E. coli* competent cells and a separate transformation with a small amount of non-digested plasmid acted as a control to check the efficiency of transformation.

5.2.5 Colony PCR screening of transformed bacteria

Colony PCR was performed using forward and reverse primers **(Table 5.3)** to check for the presence of the mutant Abyssovirus structural protease and the positive colonies with bands of the expected size were selected **Figure 5.7.**

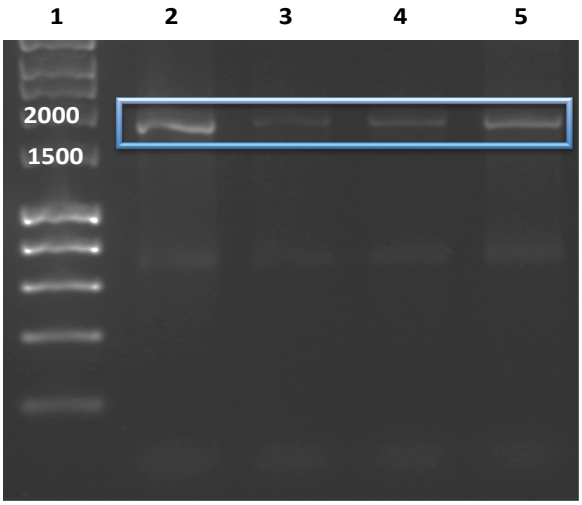

Figure 5.7. Mutant Abyssovirus structural protease screened by PCR. 4 colonies were screened for checking the presence of mutant structural protease insert. Lane 1: Marker (1kb) DNA ladder. Lanes 2, 3, 4 and 5: colonies showed positive results with correct sizes (1772 bp).

Colonies that had shown the correct size of insert by colony PCR were grown at 37°C overnight in 10 ml LB broth with ampicillin and plasmid DNA extracted. The isolated DNA was sent for sequencing to confirm the relevant oriental mutation and, once confirmed, a glycerol stock was prepared and the new strain stored at -80°C.

As a result of the cloning experiments described here, plasmids capable of the expression of wild type Abyssovirus structural protease and four mutants, which change residues predicted to be at the active site of the enzyme, were confirmed. In all cases the fragments were flanked by epitopes for widely available antibodies, for the HSV and HIS tags, so that the complete translation product representing either

end of the fragment (one or other epitope) could be detected. These plasmids represent the resources used for an investigation of possible S^{Pro} enzyme activity.

5.2.6 Expression and proteolytic activity of Abyssovirus structural protease

As noted, the base vector in use, pTriEx1.1, allows the expression of target proteins in multiple expression systems (*E. coli*, insect and vertebrate). Expression in *E. coli* strains, including *E. coli* BL21 (DE3)- pLysS and Rosetta, is enabled by these strains having been engineered to express the T7 polymerase under control of the lac promoter. IPTG induction turns on T7 polymerase expression which in turn transcribes from the plasmid resident T7 promoter (Studier and Maizel, 1969). Expression in insect cells is enabled by the presence of the p10 promoter, one of the very late promoters widely used in the baculovirus expression system (Kelly and Possee, 2016). None of these hosts is expected to contain endogenous proteins that will react with either the HSV or HIS tag antibodies meaning that only the plasmid encoded product, along with any processed intermediates, should be detected.

5.2.6.1 Abyssovirus structural protease expression in different *E. coli* strains

In order to test expression of the Abyssovirus structural protease in *E. coli* strains, all positive plasmids used in this study were transformed into *E. coli* BL21 (DE3)-pLysS. BL21 (DE3)- pLysS cells carry pLysS for toxic protein expression. It encodes T7 lysozyme, which is a natural inhibitor of the T7 RNA polymerase. The presence of this gene can inactivate leaky expression by the RNA polymerase but has no direct effect on the level of expression following induction. Transformed

competent cells were plated on LB agar plates containing ampicillin and, following overnight growth, a colony was selected randomly and grown in 10 ml LB broth with ampicillin (10 µg/ml) at 37°C overnight in a shaking incubator. The following day, 200 µl was transferred to a fresh 10 ml LB broth with ampicillin (10 µg/ml) and grown to OD_{600} 0.4-0.6 whereupon protein expression was induced by the addition of isopropyl-β-D-1 thiogalactopyranoside (IPTG) to 0.5mM and incubation with shaking was allowed to continue for 4 hours for optimal expression. Both before and after IPTG induction aliquots of the bacterial culture were harvested for western blot analysis. SDS-PAGE and western blots were performed and the expressed proteins were detected using both anti-HSV and anti-HIS antibodies.

5.2.6.1.1 Structural protease protein expression in *E. coli* BL21 (DE3)-pLysS

The first screening for S^{Pro} expression was completed using the parental S^{Pro} sequence and 4 of the active site mutants, H126A, D163A, S243A and a mutant construct in all these sites (S243, D163 & H126 to Alanine). The predicted size for the complete translated protein in all cases was ~52kDa, including both epitope tags. A sample of the non-induced culture in each case was used as control for the specificity of the blots. In most cases, when probed with the HSV antibody, induction of S^{Pro} expression led to the appearance of a clear band with a molecular mass of ~52kDa after as little as 0.5 hrs (hours) induction with IPTG, which became more intense with time up to the maximum incubation period of 4 hours

post induction except in the case of the mutant of all active site residues (S243, D163 & H126 to Alanine) which did not show any evidence of expression **Figure 5.8**.

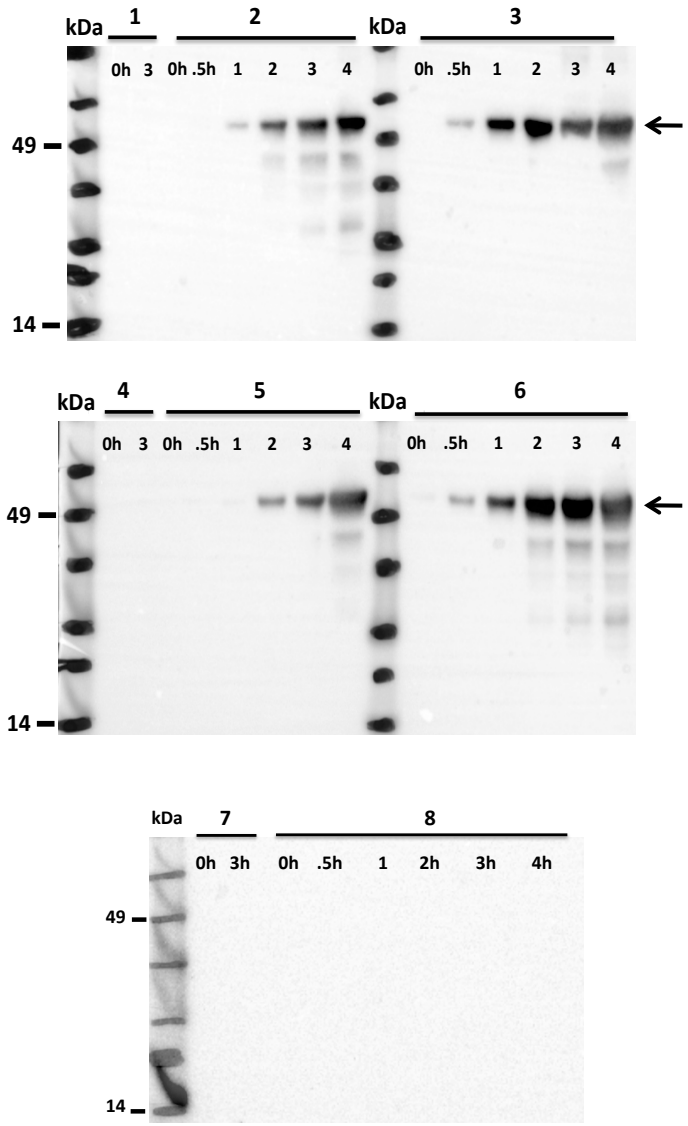

Figure 5.8. Western blot analysis by HSV tag antibody of recombinant S^{Pro} expression of expression in *E. coli* BL21 (DE3)-pLysS. Protein samples were taken prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. Protein samples were subsequently taken after .5h and hourly for 4 hours. Samples 1, 4 and 7 are non-induced cells were used as a control. Sample 2: Wild type (WT) structural protease. Sample 3: active site mutation (H126A). Sample 5: active site mutation (D163A). Sample 6: third active site mutation (S243A). Sample 8: expression of the mutant Abyssovirus structural protein in all previous active sites. The arrows indicate a melucular mass of ~52 kDa. Markers on the left are in kilodaltons, kDa.

Similarly, when the same samples were processed for Western blot analysis using the anti HIS -tag antibody a band of ~52kDa was highlighted in all constructs except the all-active site mutations, consistent with the complete translation product – **Figure 5.9.**

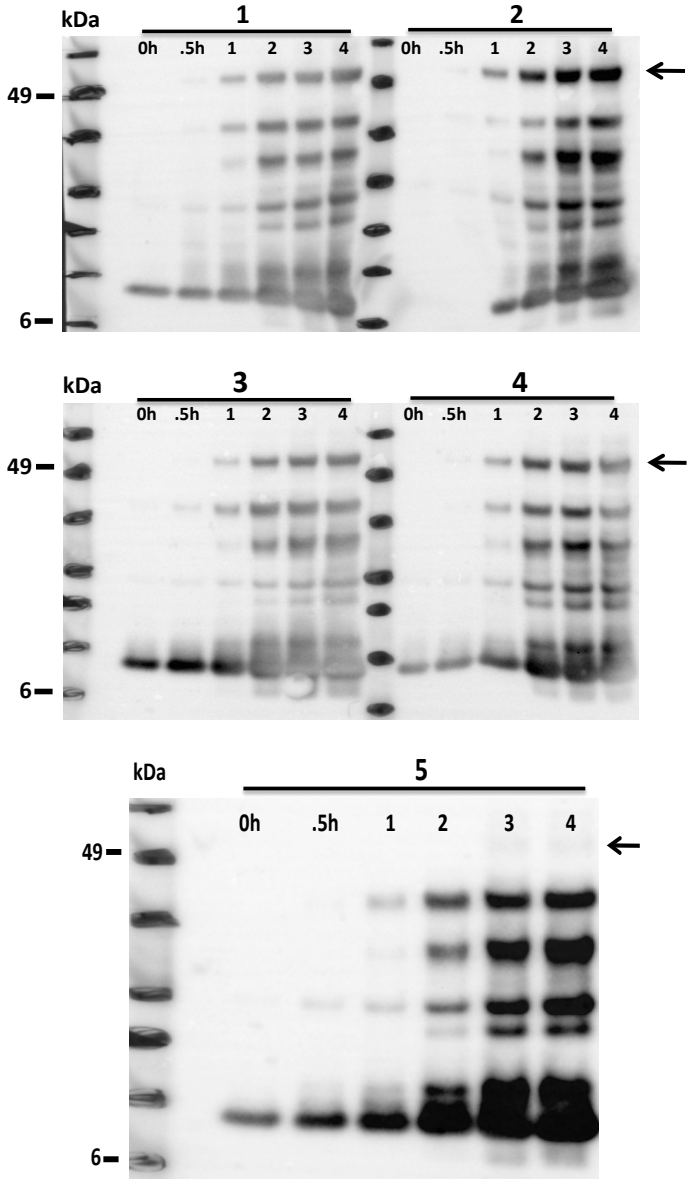

Figure 5.9. Western blot analysis by HIS tag antibody of recombinant S^{Pro} expression of expression in *E. coli* BL21 (DE3)-pLysS. Protein samples were taken prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. Protein samples were subsequently taken after .5h and hourly for 4 hours. Non-induced cells were used as a control. Sample 1: Wild type structural protease. Sample 2: active site mutation (H126A). Sample 3: active site mutation (D163A). Sample 4: third active site mutation (S243A). Sample 5: expression of the mutant structural protease in all active sites. The predicted expressed protein size is ~52 kDa. Markers on the left are in kilodaltons, kDa.

The HSV tag western blots were very clear with the predominant band recognized by the antibody at the molecular mass expected of the primary translation product except the mutant case of all active site residues. By contrast the HIS blot, while also detecting a band of the expected molecular mass, also detected a ladder of smaller bands, probably degradation products. The fact that the HSV antibody did not detect these bands suggests that protein breakdown occurs from the amino terminus to leave a nested set of proteins that have lost the N-terminus but retained a common C-terminus, including the HIS tag. The pattern of protein breakdown was the same irrespective of the identity of the S^{Pro} sequence, wild type or mutants, suggesting this is not related to the expressed protein itself but is the product of the *E. coli* protein degradation machinery.

5.2.6.1.2 Structural protease protein expression in *E. coli* Rosetta strain

E. coli Rosetta is a BL21 derivative designed to enhance the expression of eukaryotic proteins that contain codons rarely used in *E. coli*. These rare tRNAs are tRNAs for Arg, Gly, Ile, Leu and Pro and by supplying these the Rosetta strains provide for "universal" translation, which would otherwise be limited by the codon usage of *E. coli*. Protein expression of the WT Abyssovirus structural protease and mutants did not seem limited in strain BL-21 (**Figures 5.8/5.9** above). Nevertheless, two plasmids, encoding WT and the first active site mutation were transformed into the Rosetta strain and protein expression was assessed again as before. The transformants were grown in 10 ml LB containing ampicillin (10

μg/ml) at 37 °C and induced by the addition of IPTG at an OD_{600} of 0.4-0.6. Samples were taken at regular intervals (30 minutes for 3 h). As shown (**Figure 5.10**), after the addition of IPTG, expression of a band of ~52 kDa was observed, whose intensity increased with time. The band was the same molecular mass when probed with either N or C-terminal antibodies **Figure 5.10**. As before extensive degradation was observed on the HIS antibody blot while the primary translation product was the main or only product when detected by the HSV antibody.

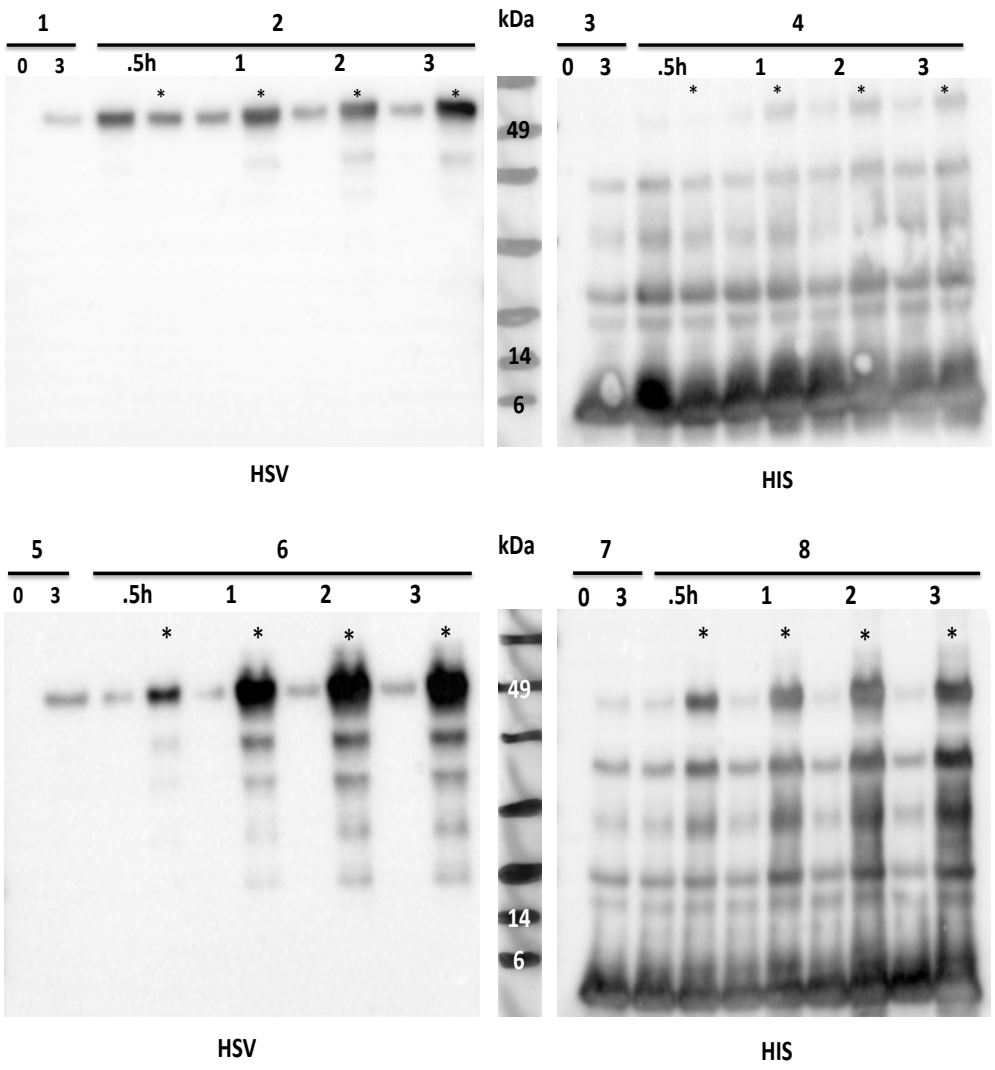

Figure 5.10. Western blot analysis of recombinant expression of wild structural (Samples 2 & 4) and first active site mutation (Samples 6 & 8) S^{Pro} in Rosetta cells using HSV and HIS antibodies. Protein samples were extracted from *E. coli* prior to (non-induced; 0hr) and following induction of protein expression with IPTG at 37°C. Protein samples were subsequently taken after .5h and up to 3 hours. Samples 1, 3, 5 and 7 are non-induced cells were a control. In both membranes, the induced (asterisk) track was more prominent than the non-induced track (no asterisk). The calculated protein molecular mas for the primary translational product is ~52 kDa.

5.2.6.2 Abyssovirus structural protease expression in insect Sf9 cells

Expression of the Abyssovirus structural protease in *E. coli* was successful but no indication of any specific cleavage was noted. One possibility was that the expressed S^{Pro} was expressed but nonfunctional in –*E. coli* as a result of its state, plausibly an insoluble inclusion body product (Fakruddin *et al*., 2013). To assess if an alternate expression system might allow expression and functional assessment, the S^{Pro} encoding plasmids, parental and mutant, were used to generate recombinant baculoviruses as described in Materials and Methods. Briefly, in addition to the multiple promoter described, the pTriEx1.1 vector also contains short sections of the baculovirus *Autographa californica* genome sequence from either side of the polyhedron locus, a site in the genome which can be altered without inactivating virus replication. When transfer vector and baculovirus DNA are transfected into insect *Spodoptera frugiperda* cells recombination occurs between the sequence on the pTriEx based vector and those of the virus genome, resulting in a recombinant in which the sequence between the recombination sequences, in this case the structural protease sequence or its mutated variants, is inserted into the virus genome. The recombinant genome then initiates a full infection and a recombinant virus expressing the target protein as part of its life cycle is isolated.

To isolate the recombinant viruses, the Sf9 insect cell line was cultured in EX-CELL® 420 Serum-Free Medium and transfections were done in 6 well dishes seeded at ~50% confluence with ~1 x 10^6 cells per well. In addition to the structural protease plasmids, a control vector pTriEx1.1-GFP, which carries the Green Florescent Protein (GFP) gene inserted between the *Nco*I and *Bsu*361 sites of pTriEx1.1 was used to ensure successful recombinant formation (O'Flynn, 2011).

Recombinant viruses were amplified by passage on Sf9 cells until cytopathic effect was evident within 2 days of infection, usually by passage 3. To assess protein expression by the recombinant viruses 6 well dishes were seeded at near 100% confluence and infected with P3 virus stocks at an estimated MOI of 3 or greater, under which condition nearly all cells are infected. The plates were incubated at $28^{O}C$ for three days and the monolayers collected by dislodging the cells with a pipette, transferred to Eppendorf tubes and pelleted at 13,000 rpm for 3 minutes (microfuge).

The supernatant was aspirated and the cell pellet lysed directly by resuspension in 1 x SDS-PAGE loading buffer and heated at $100^{O}C$ for 10 minutes. The equivalent of 5 x 10^4 cells was loaded per well of a standard 10cm precast SDS polyacrylamide gel and electrophoresis was carried out for 30 minutes at 170V. Gels were then transferred to PVDF membranes by semi-dry blotter and the membranes processed for western blot as described in Material and Methods using antibodies to the HSV or HIS tag. The tagged proteins expressed by each structural protease construct were detected by imaging using a chemiluminescent substrate for the anti-tag antibody conjugates on a Syngene XL 1.14 imager **Figure 5.11.**

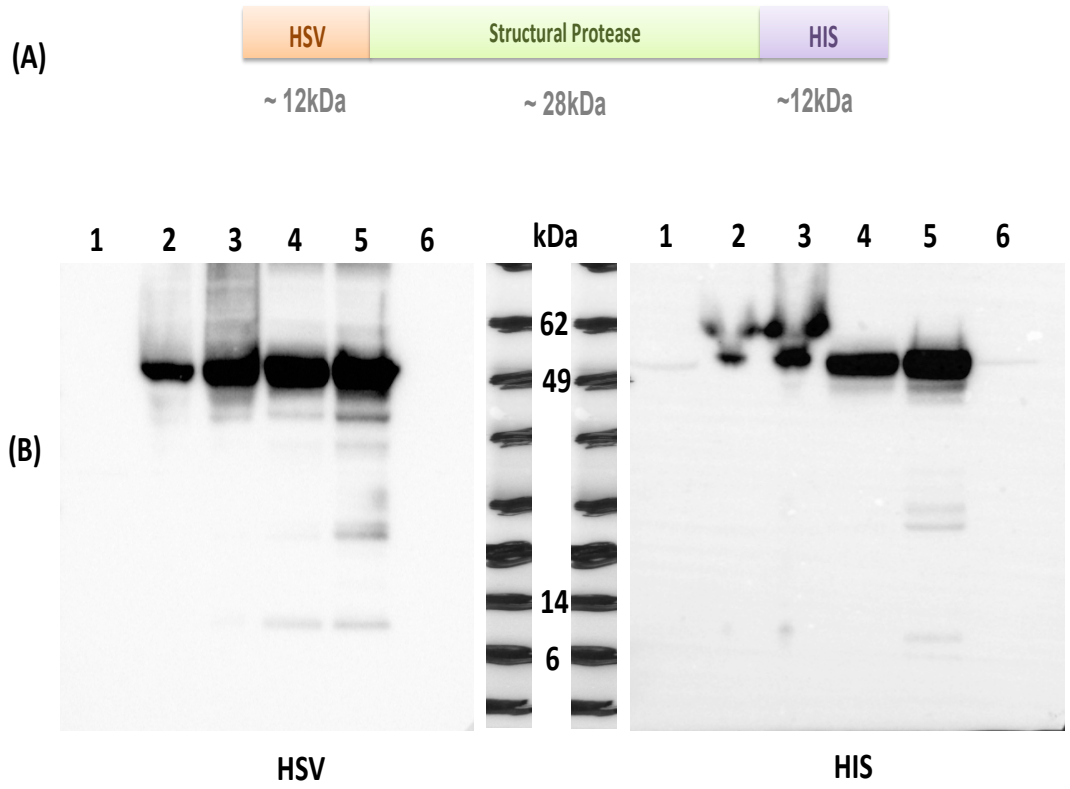

Figure 5.11. Abyssovirus structural protease proteins expressed in Sf9 insect cells. **A:** cartoon of Abyssovirus structural protease construct. **B:** Western blot of wild and mutants structural protease. The anticipated molecular weights are shown above. In both gels, Lane 1: cell control. Lane 2: Abyssovirus structural protease wild type. Lane 3: first active site mutant (H126A). Lane 4: second active site mutant (D163A). Lane 5: third active site mutant (S243A). Lane 6: mutation in all active site residues (S243, D163 & H126 to Alanine). Molecular weight markers are indicated in the center of the blots and are in kilodaltons.

For the wild type Abyssovirus structural protease and all mutants where the protease active site contained one or other mutation, the results of the western blot analysis indicated the production of a 52 kDa band that was visualized by blot with either anti HSV antibody (left gel lanes: 2 to 6 respectively) or the anti HIS

antibody (right gel lanes: 2 and 6 respectively). The size of the 52kDa band and its reactivity with both anti tag antibodies is consistent with the primary translation product as identified following expression in *E. coli*. Notably for the product expressed in insect cells however, no ladder of degraded bands was apparent when the extract was probed with the HIS antibody. The pattern of the expression observed was the same irrespective of the S^{Pro} sequence used and there was no evidence of any smaller band in the WT sample that was missing in any of the mutant samples. Similarly there was no evidence of differential reaction with the antibodies, other than some very low-level background associated particularly with mutant 3 (S243A). As was the case following expression in *E. coli* there was no evidence of expression of the triple active site mutant when the infected cells were probed with the anti HSV antibody. However a very faint band of reactivity was detected, again at ~52kDa, by the anti HIS antibody (**Figure 5.11** B -right panel). This result suggests the triple mutant is expressed but is highly susceptible to degradation. Overall while expression of S^{Pro} in insect cells was efficient there was no pattern of break down to suggest that the Abyssovirus structural protease was active.

5.2.7 Abyssovirus structural protease purification using IMAC from insect cells

The failure to detect any function for the Abyssovirus S^{Pro} despite clear evidence for expression left open the question of its relevance to the Abyssovirus replication cycle. The lack of any auto-cleavage could be for a variety of reasons, not least that the conditions in the expression hosts, *E. coli* or insect cells, were not compatible with activity. However, the failure to auto-cleave the HIS tag offered the

opportunity to make use of it in the purification of S^{Pro}, which would then allow a much greater range of conditions to be tested, possibly to reveal activity. Accordingly, a large scale infection of the WT expressing recombinant baculovirus was done and purification investigated. The addition of the HIS tag to a recombinant protein brings with it an affinity for metal ions, notably nickel, and HIS "tagged" proteins can be therefore purified by binding to an immobilized metal ion matrix in a process termed Immobilized Metal Ion Chromatography (IMAC). Non-HIS tagged proteins do not bind allowing a single step purification of the recombinant from a background of many other cellular proteins. Elution is achieved by competition with free imidazole, the basis of the ion binding activity.

To purify S^{Pro} a 300ml culture of insect Sf9 cells at a density $1.5x10^6$ cells/ml was infected with 30mls of the structural protease recombinant virus stock and incubated at 27 C° for 3 days in a rotary shaker. Then, cells were harvested by centrifugation at 4000rpm for 20 min at 4 C° and the cell pellets placed on ice and lysed by resuspension in lysis buffer (50mM phosphate buffer, 500mM NaCl, 1% NP-40, pH 7.4). A single tablet of a protein inhibitor cocktail (Roche) was then added to the lysate and lysis was completed by sonication on ice for 10 min / 20 seconds on/off pulse and 80% amplitude using a probe sonicator (Sonics). The cell lysate was centrifuged at 15,000 rpm for 20 minutes at 4C° using a Sorval RC5B centrifuge and the supernatant collected. An integrated low pressure purification apparatus (Biologic-BioRad) was used to purify the protein of interest using a prefilled 5ml IMAC column (GE). The column was washed with binding buffer (20 mM PBS, 500mM NaCl, 20 mM Imidazole, 0.1% Tween-20, pH 7.4) at 5ml/min and the cell lysate was loaded into the column at 1ml/min. The column was washed with binding buffer until the OD_{280} was at background levels and HIS-tagged

proteins were then eluted using elution buffer (20 mM PBS, 500mM NaCl, 500 mM Imidazole, 0.1% Tween-20, pH 7.4) as a linear gradient of 0% to 100% elution buffer over a volume of 6 column volumes at 1ml/ min. The eluate fractions containing the protein of interest are shown in **Figure 5.12** and were collected separately by fraction collector. To assess the level of purifcation, selected fractions were analyzed by SDS-PAGE followed by coomassie staining (**Figure 5.13**) and western blot (**Figure 5.14**) to confirm the presence of the protein of interest. Positive fractions were pooled and concentrated using a Viva spin column with a 10K MW cut-off (Sartorius) by spinning at 4000 rpm for 30 minutes at 4C˚.

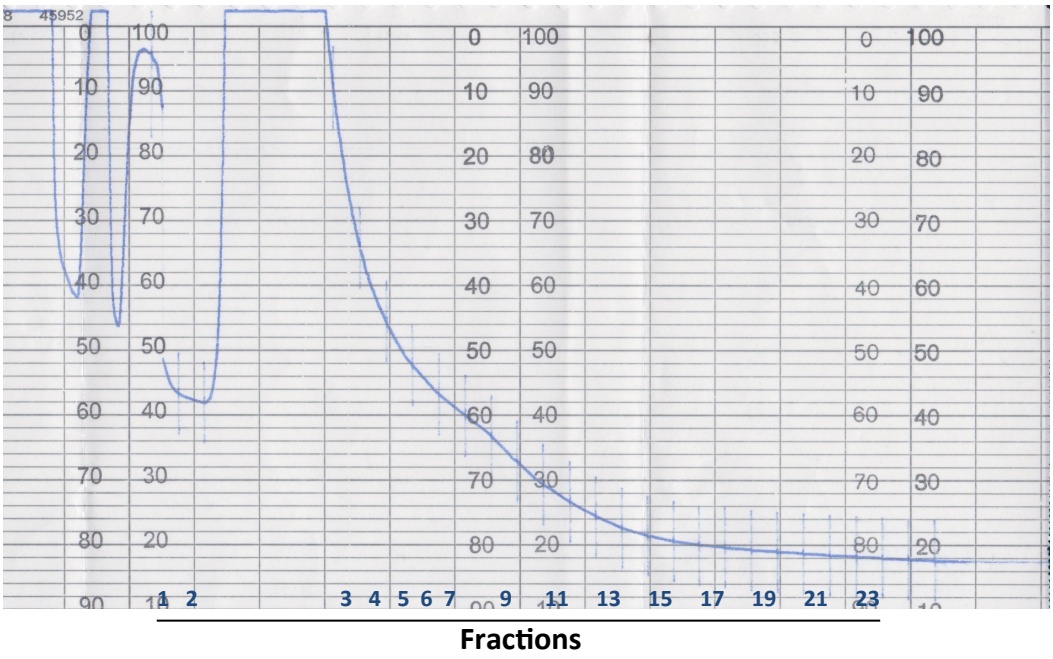

Figure 5.12. Elution profile of the Abyssovirus structural protease after affinity chromatography. The lysate was applied to the column was washed to a background OD_{280} with loading buffer before the start of an elution gradient of 20mM-500mM imidazole in binding buffer. Fractions of 5ml were automatically collected as shown on the elution profile. A large peak early in the elution profile, representing loosely bound cellular proteins, was followed by a shoulder peak starting from fraction 7 and extending to fraction 13. To detect the protein of interest the proteins present in the odd fraction numbers, from 7 to 21, were analyzed by SDS-PAGE.

From **Figure 5.12** it is difficult to see an obvious peak other than a very large early peak which seems too early to be the recombinant product. However, a shoulder between fractions 7 to 21 might be the purified Abyssovirus structural protease and odd fractions between 7 to 21 were analyzed by protein gel (**Figure 5.13**).

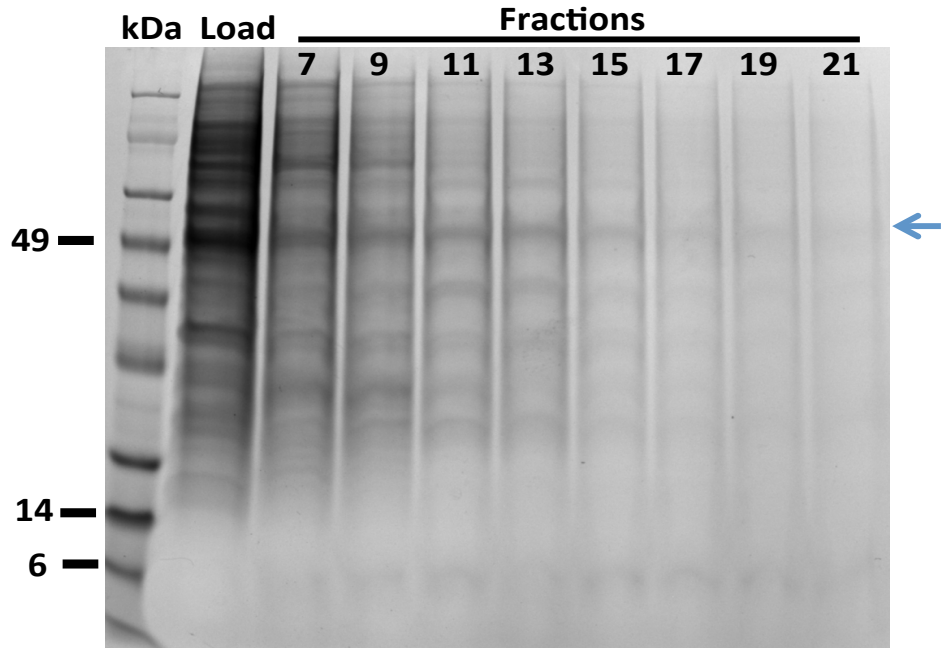

Figure 5.13. Analysis of elution fractions following IMAC of Abyssovirus S^{Pro} extracts from Sf9 insect cells. The SDS PAGE gel was stained with coomassie blue. Insect cells were infected with structural protease recombinant virus stock for 3 days at 27 C˚. Molecular weight marker in kilodaltons are indicated on the left side of the gel followed by load sample and odd numbered eluted fractions from 7 to 21. The arrow indicates a molecular mass of ~52kDa.

As the coomassie blue stained gel did not resolve a clear band in the eluted fractions, rather a smear of bands suggesting poor column selectivity (**Figure 5.13)** a second gel was subjected to western blotting to confirm the presence of the Abyssovirus structural protease (**Figure 5.14**). The S^{Pro} was clearly present in the lysate loaded onto the column and was not present in the early fraction 7 suggesting efficient column binding. In the eluted fractions, the S^{Pro} appears weakly in fraction

9 and become more intense with the level peaking in fraction 13 and declining thereafter, although signal is still present in fraction 21.

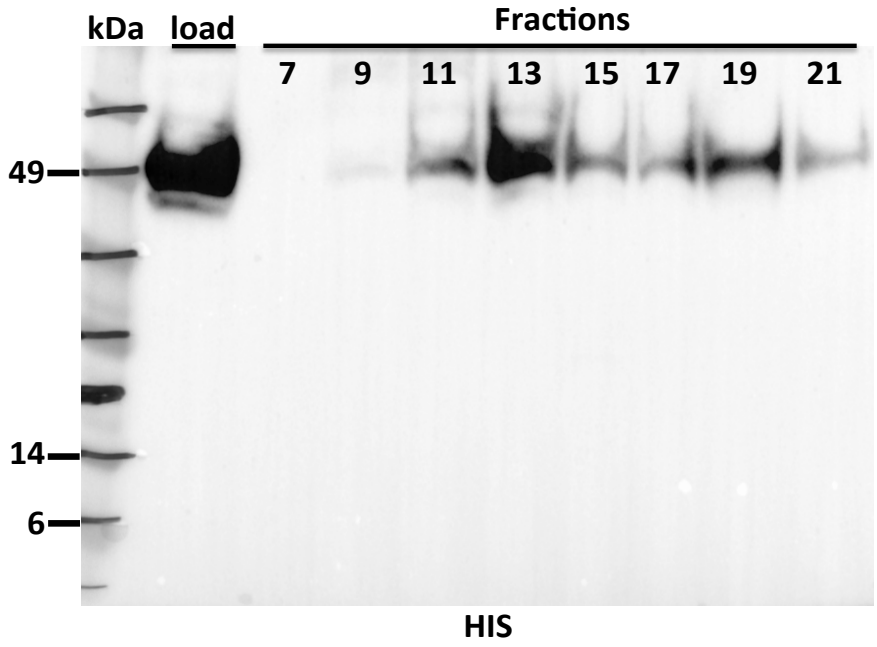

Figure 5.14. Western blot of elution fractions from IMAC purification of structural protease in Sf9 cells. The Molecular weight marker in the left of the blot is in kilodaltons, following by load sample and odd numbered eluted fractions from 7 to 21.

Two different analytical techniques were used to detect the purified Abyssovirus structural Protease. SDS-PAGE followed by coomassie blue failed to showed clear bands of purified protein, rather a smear of proteins indcatative of poor column resolution. In contrast to the results obtained from coomassie blue however, the Western blot analysis indicated that IMAC chromatography had succeeded in binding and eluting the his tagged S^{Pro} with no reactivity in the wash fractions but reactivity in the eluted fractions with a peak in fraction 13. However, the excess of

cellular proteins also present, as shown by coomassie blue staining, meant that further purification would be needed before dedicated protease activity tests in different buffer conditions could be attempted. This was not pursued and remains a goal for future studies.

5.3 Discussion

In this chapter the possible function of the novel Abyssovirus structural protease was investigated. The novelty arises as the chymotrypsin-like serine protease sequence found in the Abyssovirus genome sequence has features analogous to the alphavirus capsid protease (Aggarwal *et al.,* 2015) despite not being described in any nidovirus to date.

The cloning of Abyssovirus structural protease and several mutants used a mixture of In-Fusion cloning and ligation cloning techniques as well as synthetic biology and overlap PCR. Clones capable of expression of HSV and HIS tagged fusion proteins were successfully completed as demonstrated by sequence analysis prior to the expression experiments. All Abyssovirus structural protease sequences were expressed in two different expression systems, including insect cells (Sf9) and *E. coli* strains (strains BL21 (DE3)- pLysS and Rosetta). Proteins showed expression in these systems and were detected using Western blots using anti HSV and anti HIS antibodies. All of the cloned constructs except the construct of the mutant Abyssovirus structural protease in all active site residues showed easily detected protein expression although the *E. coli* background gave rise to a number of breakdown products reasoned to be N terminal degradation as the same bands were not detected by the HSV antibody. The same pattern of degradation was not

observed in insect cells confirming that it was not dependent on S^{Pro} activity. The lack of detectable protein expression for the triple active site residue mutation was notable. This mutant was sequence verified, as were the single residue mutants, yet no evidence for a 52kDa band was found following expression in *E. coli*. While HSV detection also failed to identify this band in insect cells a weakly reactive band was observed for the HIS antibody probed blot. Poor expression in both systems is suggestive of a protein that folds poorly and is rapidly degraded (Arun *et al*., 2009), which could be consistent with the mutations made as active site residues are often in the hydrophobic core of enzymes and their mutation may lead to misfolding (Robertson *et al.,* 2016; Hartl *et al.,* 2011; Nijholt *et al.,* 2011). Thus, this data supports the relevance of these residues even if no protease activity was detected although their location at an active site must await more positive functional data or a three dimensional structure.

The capsid protease of alphavirus (CP), a chymotrypsin-like serine protease, is found at the amino terminus of the structural polyprotein and the first step in processing is the autocatalytic cleavage of the capsid protease to release itself from the rest of the polyprotein (Thomas *et al.,* 2010, Aggarwal *et al.,* 2014). This step is essential for the functions of CP in the virus replication cycle, including the creation of capsomers by intermolecular interactions with other capsid protease monomers, the encapsidation of the genomic RNA in order to generate the nucleocapsid cores, and the interaction with the cytoplasmic domain of glycoproteins that is important for the budding process of the virus (Aggarwal *et al*., 2014, Metsikko and Garoff, 1990, Zhao *et al.,* 1994 and Hong *et al.,* 2006).

By contrast with the alphaviruses capsid protease, the Abyssovirus structural protease did not produce any detectable cleavage products in any expression

system, including insect cells (Sf9) and *E. coli* strains (BL21 (DE3)-pLysS and Rosetta). Western blot analysis showed clear bands for only the complete protein, in both HSV and HIS blots. The failure of the putative structural protease to cleave might be because a result of the construct boundaries, assay conditions or lack of an appropriate substrate.

Chapter 6

6. General discussion

Recent studies have identified a wide variety of virus sequences in intracellular RNA pools but few new members of the *Nidovirales* have been reported, compared to groups such as the *Picornavirales* (Shang *et al.,* 2017). In this project, a novel nidovirus sequence was identifying in the TSA database in cellular RNA pools of *Aplysia californica,* the Californian sea hare (**Figure 6.1-A**), following a bioinformatics trawl and tentatively named the Abyssovirus to reflect its deep sea origins (Bukhari *et al*., 2018). A second independent report of the same sequences, found in the RNA pools of a variety of *Aplysia* adult tissues and several developmental stages of the host organism was reported by Debat (Debat, 2018).

The Abyssovirus genome is represented in its longest and most completely available form by the transcriptome shotgun assembly sequence GBBW01007738. The virus genome (**Figure 6.1-B**) described is the second largest currently reported RNA virus genome behind a new 41.1 kb planarian nidovirus described elsewhere (Saberi *et al*., 2018).

(A)

(B)

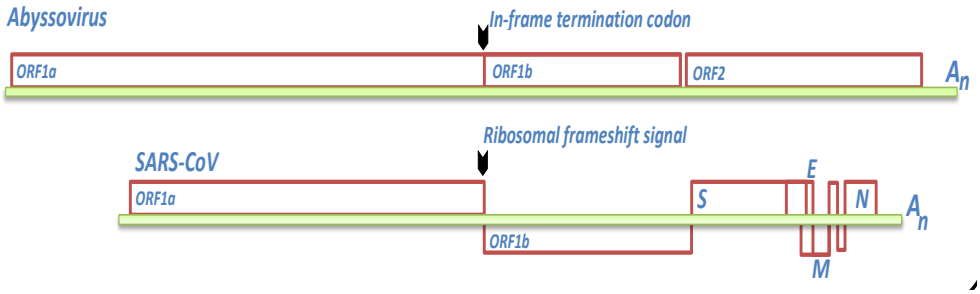

Figure 6.1. Description of Abyssovirus host organism and genome. (**A**) The host organism *Aplysia californica* (adapted from Moroz *et al.,* 2006). (**B**) Genome and coding capacity of Abyssovirus and SARS-CoV are shown to scale.

Remarkably, the organization of the Abyssovirus genome had several features typical of viruses of the Alphavirus genus of the *Togaviridae* family (King *et al*., 2012) that could be contrasted with those conserved in classical nidoviruses. They include: a) two in-frame open reading frames (ORF1a and ORF1b) of the replicase gene that are separated by a stop codon rather than overlapping and including a nidovirus-like ribosomal frameshift signal in the overlap, and b) a single structural polyprotein gene (ORF2) rather than several ORFs encoding structural proteins. The 35913 nt long Abyssovirus genome has a 74 nt 5′- untranslated region, a 964 nt 3′-untranslated region, and a short poly-A tail. Despite these alphavirus-like features, bioinformatics analysis confirmed that the Abyssovirus replicase

polyprotein clustered with the *Nidovirales*.

When identifying viruses through bioinformatics, there is a risk that the sequences are either misassembled, contain errors, or are artifacts of the sequencing and sequence assembly processes. As a result it was necessary to test the function of some of the novel Abyssovirus features and compare the data obtained with nidoviruses and alphaviruses. The targets for this work included the main protease, the pp1a-pp1b junction region and the structural protease. Functional activity, if found, would help to confirm that the Abyssovirus genome was biologically functional and represented an extant replicating virus.

To carry out the functional tests, sequences encoding the Abyssovirus M^{Pro}, the pp1a-pp1b junction region and the S^{Pro} plus surrounding regions were cloned into the pTriEx1.1 vector and expressed as amino-terminal HSV epitope tagged and carboxyl-terminal poly-histidine (HIS) tagged polypeptides (**Figure 6.2)**. Expression was assessed in different cell systems include bacterial cells (several *E. coli* strains), Sf9 insect cells and 17clone-1 mammalian cells and the data analyzed to contribute to the question of whether the Abyssovirus genome represented a replicating virus or was an artifact of sequence assembly.

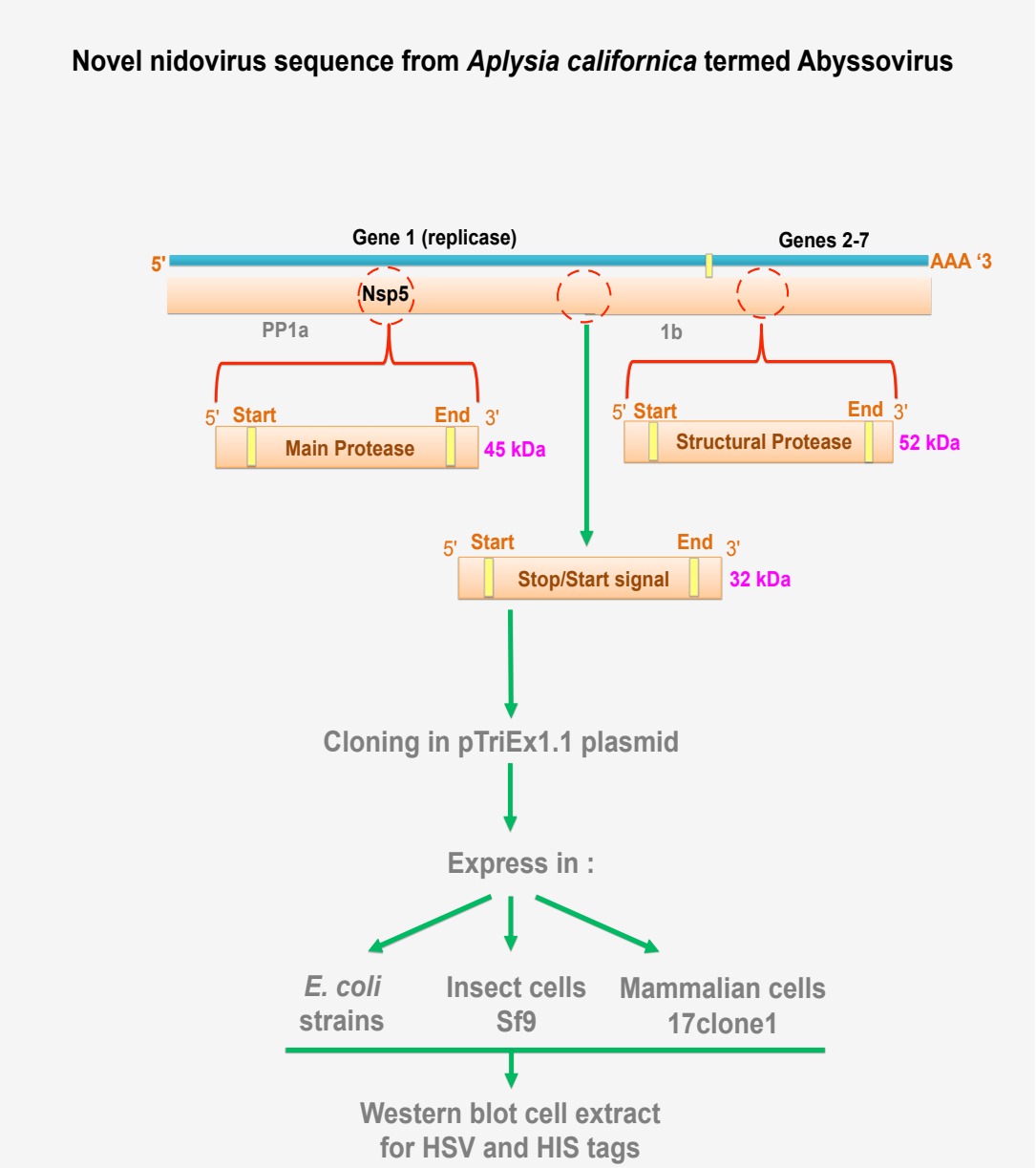

Figure 6.2. Investigation overview of selected Abyssovirus sequences assayed for biological function.

The first target of this report is to detect the activity of Abyssovirus main protease, M^{Pro}, also known as the 3C-like protease. Constructs of this sequence included the wild type sequence and several alanine-scanning mutations which were designed to

alter amino acids that appeared to match the catalytic cysteine and histidine residues of other coronavirus main proteinases. In addition four further mutations located in the N-terminal region of the M^{Pro} sequence were made to probe the presumed cleavage sites, as has been reported in a previous study of nidovirus main protease (Lee *et al*., 1991).

All the Abyssovirus M^{Pro} constructs were tested for expression in several expression systems include *E. coli* BL21, insect cells (Sf9) and mammalian cells (HEK 17clone1).

The expressed product levels of all of the Abyssovirus M^{Pro} constructs varied, for instance, it was not possible to detect the M^{Pro} fragments by western blot after transfection of mammalian cells although part of this was undoubtedly the limited number of cells that were successfully transfected. A positive control, GFP, was detected but the western blot signal was not strong indicating poor overall efficiency. Some protocol changes including the use of different transfection agents did not substantially improve expression and rather than extend the optimisation to every possible variable other expression systems, part of the original experimental design, were investigated. One of these, the *E. coli* expression system, gave clear data following expression in a number of strains (including: BL21 (DE3)-pLysS, Origami, Rosetta and Tuner). A band of the anticipated size of M^{Pro}, detectable by antibodies to both protein tags, was present following protein induction but disappointingly showed no difference among the constructs tested; all gave the same expression outcome, the complete protein size was detected at the predicted molecular weight and with very little degradation detected by the HSV tag antibody. When the blots were probed with the HIS tag antibody substantial non-

specific cleavage was apparent but again no difference among the constructs tested, leading to the conclusion that the Abyssovirus main protease was not active in *E. coli*. Assuming that the protein is a *bona fide* protease, as shown later by insect cell expression, the failure to detect function in *E. coli* is perplexing. Lack of the correct cellular environment, lack of specific cofactors or insolubility of the expressed products, alone or in combination, could explain the data. Further work would be needed to determine which of these predominated.

However, in contrast to the results obtained from *E. coli* the results of the Western blot analysis indicated the production of a 13kDa HSV related band and a 6kDa HIS related band when the wild type M^{Pro} sequence was expressed in insect Sf9cells. Mutation of the putative catalytic cysteine C302 alone did not substantially reduce proteolytic processing, while mutation of both histidine H193 and cysteine C302 blocked proteinase activity completely. Together, these data matched the published data of the analysis of other coronavirus main proteases, adding support for the finding that this is a genuine proteolytic activity of relevance to virus replication (Ziebuhr *et al.,* 2003, Nga *et al.,* 2011, Blanck *et al.,* 2014).

To further characterize the activity and to determine the cleavage specificity of the Abyssovirus M^{Pro}, four initial mutants were made (R, K, N and E to Ala) in the N-terminal region of the main protease. As noted in Chapter 3, two of the mutants, N138 and E139, produced only the higher molecular mass band on western blots suggesting that these two mutants stopped enzyme activity even though they were designed to alter only one of the putative processing sites. Two further mutants, R136 and K137, led to two bands visualized in the western blot, matching the result of the wild type main protease. Thus in contrast to N138 and E139, R136 and K137

appeared to show no effect on protease activity. Overall, the data obtained with the Abyssovirus M^{Pro} N-terminal cleavage junction mutants demonstrate that R, K, N and E play a role in the protease activity, however, this role is unknown as these constructs gave unexpected data in the *cis*-cleavage assay. It has been noted in some previous studies that a M^{Pro} protease peptide cleavage site of SARS-CoV (TSAVLQ/SGFRK) occurs at a Q/S junction (Chen *et al.,* 2005, Hilgenfeld, 2014) and a QS motif appears in the sequence of the Abyssovirus M^{Pro} just downstream of the RKNE

motif. As with mutations N138 and E139 mutation at this residue (Q152A) appeared to inactivate the enzyme, further confusing the cleavage junction used by the Abyssovirus main protease. These data demonstrate that Abyssovirus encodes a functional proteinase, but further work is needed to determine the cleavage specificity by the Abyssovirus M^{Pro}.

In the second target for study, the presence of an in-frame stop codon separating the pp1a and pp1b genes, rather than the expected ribosomal frameshift signal found in most other nidoviruses was investigated. The observation suggested that Abyssovirus might use a translational termination-suppression signal as a way to control expression of the pp1b region. Termination-suppression or readthrough signals are found in several other viruses including alphaviruses and some retroviruses, and typically consist of a UAG or UGA stop codon followed by an RNA secondary structure element and the efficiency of suppression normally depends on the stop codon, the nucleotide following the stop codon, and the free energy of the RNA secondary structure element (Firth *et al.,* 2011, Napthine *et al.,* 2012, Irigoyen *et al.,* 2018, Loughran *et al.,* 2018). To investigate protein

expression across the pp1a-pp1b region, nucleotides 17255 to 17707 of the Abyssovirus sequence were cloned into pTriEx1.1 with amino-terminal HSV and carboxyl-terminal HIS tags. As noted in Chapter 4, this construct was designed to allow detection and quantification of the 25kDa proteins that stopped at the natural UGA stop codon, expected to have an HSV tag only, and a 32kDa readthrough product that would be expected to have both HSV and HIS tags. Expression of this construct in all positive expression systems produced the expected 25kDa termination product and 32kDa readthrough product (shown in chapter 4). Based on densitometry analysis, it was estimated that 9.42-12.58% of translation events resulted in readthrough.

The stop codon and the two elements that follow it, all suggested in previous studies to play a key role in translational termination efficiency were further investigated for the Abyssovirus termination-suppression signal by designing constructs in which the region around the pp1a stop codon was perturbed from the wild-type UGA-C, predicted to produce near optimal termination, to UAA-A, predicted to produce much less optimal termination. In other constructs, the 42 nucleotides predicted to form a predicted RNA stem-loop downstream of the stop readthrough site were deleted. As discussed in Chapter 4, mutation of the Abyssovirus pp1a stop codon itself had little effect on readthrough efficiency. However, deletion of 42 nucleotides appeared to decrease readthrough, and led to a smaller readthrough product as predicted. Together these results indicate that the pp1b region of Abyssovirus can be expressed by readthrough of a UGA stop codon at the 1a/b junction. This readthrough is mediated by a functional termination suppression signal that is dependent partly on a predicted stem-loop structure following the stop codon. The ability to express a constant ratio of a downstream

ORF in relation to that encoded upstream has some potential in biotechnology applications where two proteins are required from the same expression cassata but at different levels. Further work might investigate this possibility for the Abyssovirus stop readthrough signal where two ratios are possible, based on the inclusion or not of the 42bp downstream sequence.

A last target for analysis in this project was the wildtype Abyssovirus structural protease, found as described to be more closely related to a protease normally found in alphaviruses. Whether this protease is active and if so, which active sites residues are required, was the hypothesis to be tested. As before this was approached by S^{Pro} expression and alanine-scanning mutations that appeared to match the conserved residues found in related proteases, i.e. Histidine, Aspartic acid singly and as all active sites mutant (Serine, Histidine and Aspartic acid).

The expression of the wildtype Abyssovirus structural protease (S^{Pro}) and the mutant protease fragments were carried out by using two expression systems, Sf9 insect cells and bacterial cells. In the western blot of the insect cells, both HSV and HIS blots showed that, in most cases, the S^{Pro} was expressed (Chapter 5). Mutation at the presumed catalytic residues, Histidine and Serine+Histidine+Aspartic of S^{Pro} reduced expression slightly but the results did not indicate the presence of any detectable cleavage products. The bacterial expression system gave the same results in the HSV blot although, by contrast, the HIS blot showed small bands, possibly degradation products. Although IMAC protein purification method was attempted in order to isolate a pure protein for test of the protease activity in different conditions, purification was inefficient and this goal was not achieved. Further work on the purification method will be needed before multiple *in vitro* tests can be

completed. Overall, there was no pattern of break down on the western blots to suggest that the Abyssovirus structural protease was active and it has to be concluded that the S^{Pro} construct appeared to have no protease activity. Further work will be needed to determine whether the failure of the putative S^{Pro} to cleave was a result of the construct boundaries, assay conditions, lack of an appropriate substrate, or lack of the correct cellular environment, including other viral proteins, necessary for activity.

References

Aggarwal M, Tapas S, Preeti, Siwach A, Kumar P, Kuhn RJ, Tomar S. 2012. Crystal structure of aura virus capsid protease and its complex with dioxane: new insights into capsid-glycoprotein molecular contacts. *PLoS One,* 7, e51288.

Aggarwal, M., Dhindwal, S., Kumar, P., Kuhn, R. J. & Tomar, S. 2014. trans-Protease activity and structural insights into the active form of the alphavirus capsid protease. *Journal of virology,* 88**,** 12242-53.

Aggarwal, M., Sharma, R., Kumar, P., Parida, M. & Tomar, S. 2015. Kinetic characterization of trans-proteolytic activity of Chikungunya virus capsid protease and development of a FRET-based HTS assay. *Scientific reports,* 5**,** 14753.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment search tool. *Journal of molecular biology,* 215**,** 403-410.

Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J. & Hilgenfeld, R. 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *The EMBO Journal,* 21**,** 3213-3224.

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R. & Hilgenfeld, R. 2003. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science,* 300, 1763-1767.

Angelini, M. M., Akhlaghpour, M., Neuman, B. W. & Buchmeier, M. J. 2013. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *MBio,* 4**,** e00524-13.

Aouadi, W., Blanjoie, A., Vasseur, J. J., Debart, F., Canard, B. & Decroly, E. 2017. Binding of the Methyl Donor S-Adenosyl-l-Methionine to Middle East Respiratory Syndrome Coronavirus 2'-O-Methyltransferase nsp16 Promotes Recruitment of the Allosteric Activator nsp10. *Journal of virology,* 91**,** e02217-16.

Aricescu, A. R., Lu, W. & Jones, E. Y. 2006. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallographica Section D: Biological Crystallography,* 62, 1243-1250.

Attoui, H., Mohd Jaafar, F., Biagini, P., Cantaloube, J. F., DE Micco, P., Murphy, F. A. & De Lamballerie, X. 2002. Genus Coltivirus (family Reoviridae): genomic and morphologic characterization of Old World and New World viruses. *Archives of virology,* 147**,** 533-561.

Baker, S. C., Yokomori, K., Dong, S., Carlisle, R., Gorbalenya, A. E., Koonin, E. V. & Lai, M. M. 1993. Identification of the catalytic sites of a papain-like cysteine proteinase of murine coronavirus. *Journal of virology*, 67, 6056-6063.

Bakkers, M. J., Zeng, Q., Feitsma, L. J., Hulswit, R. J., Li, Z., Westerbeke, A., Van Kuppeveld, F. J., Boons, G. J., Langereis, M. A., Huizinga, E. G. & De Groot, R. J. 2016. Coronavirus receptor switch explained from the stereochemistry of protein-carbohydrate interactions and a single mutation. *Proceedings of the National Academy of Sciences,* 113**,** E3111-E3119.

Baranov, P. V., Gesteland, R. F. & Atkins, J. F. 2002. Recoding: translational bifurcations in gene expression. *Gene,* 286**,** 187-201.

Baranov, P. V., Gesteland, R. F. & Atkins, J. F. 2002. Release factor 2 frameshifting sites in different bacteria. *EMBO reports,* 3**,** 373-377.

Baranov, P. V., Gurvich, O. L., Hammer, A. W., Gesteland, R. F. & Atkins, J. F. 2003. Recode 2003. *Nucleic acids research,* 31**,** 87-89.

Baranov, P. V., Henderson, C. M., Anderson, C. B., Gesteland, R. F., Atkins, J. F. & Howard, M. T. 2005. Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology,* 332**,** 498-510.

Bárcena, 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. *Proceedings of the National Academy of Sciences,* 106**,** 582-587.

Barrett, A. J., Tolle, D. P. & Rawlings, N. D. 2003. Managing peptidases in the genomic era. *Biological chemistry*, 384, 873-882.

Becares, M., Pascual-Iglesias, A., Nogales, A., Sola, I., Enjuanes, L. & Zuniga, S. 2016. Mutagenesis of Coronavirus nsp14 Reveals Its Potential Role in Modulation of the Innate Immune Response. *Journal of virology,* 90**,** 5399-5414.

Belouzard, S., Millet, J. K., Licitra, B. N. & Whittaker, G. R. 2012. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses,* 4**,** 1011-1033.

Bertram, G., Innes, S., Minella, O., Richardson, J. P. & Stansfield, I. 2001. Endless possibilities: translation termination and stop codon recognition. *Microbiology,* 147**,** 255-269.

Bhardwaj, K., Liu, P., Leibowitz, J. L. & Kao, C. C. 2012. The coronavirus endoribonuclease Nsp15 interacts with retinoblastoma tumor suppressor protein. *Journal of virology,* 86**,** 4294-4304.

Birktoft, J. J. & Blow, D. 1972. Structure of crystalline α-chymotrypsin: V. The atomic structure of tosyl-α-chymotrypsin at 2 Å resolution. *Journal of molecular biology,* 68**,** 187-240.

Blanck, S., Stinn, A., Tsiklauri, L., Zirkel, F., Junglen, S. & Ziebuhr, J. 2014. Characterization of an alphamesonivirus 3C-like protease defines a special group of nidovirus main proteases. *Journal of virology,* 88**,** 13747-13758.

Bleckmann, M., Schurig, M., Chen, F. F., Yen, Z. Z., Lindemann, N., Meyer, S., Spehr, J. & Van Den Heuvel, J. 2016. Identification of Essential Genetic Baculoviral Elements for Recombinant Protein Expression by Transactivation in Sf21 Insect Cells. *PLoS One*, 11, e0149424.

Bouvet, M., A. Lugari, C. C. Posthuma, J. C. Zevenhoven, S. Bernard, S. Betzi, I. Imbert, B. Canard, J. C. Guillemot, P. Lecine, S. Pfefferle, C. Drosten, E. J. Snijder, E. Decroly & X. Morelli. 2014. Coronavirus Nsp10, a Critical Co-Factor for Activation of Multiple Replicative Enzymes. *Journal of Biological Chemistry,* 289, 25783-25796.

Brierley, I., Jenner, A. J. & Inglis, S. C. 1992. Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *Journal of molecular biology,* 227**,** 463-479.

Brierley, I., P. Digard & S. C. Inglis. 1989. Characterization of an Efficient Coronavirus Ribosomal Frameshifting Signal - Requirement for an Rna Pseudoknot. *Cell,* 57, 537-547.

Brierley, I., Rolley, N. J., Jenner, A. J. & Inglis, S. C. 1991. Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *Journal of molecular biology,* 220**,** 889-902.

Bukhari, K., Mulley, G., Gulyaeva, A. A., Zhao, L., Shu, G., Jiang, J. & Neuman, B. W. 2018. Description and initial characterization of metatranscriptomic nidovirus-like genomes from the proposed new family Abyssoviridae, and from a sister group to the Coronavirinae, the proposed genus Alphaletovirus. *Virology,* 524**,** 160-171.

Cavanagh, D. 1997. Nidovirales: a new order comprising Coronaviridae and Arteriviridae. *Archives of virology,* 142**,** 629.

Chang, H. P., Chou, C. Y. & Chang, G. G. 2007. Reversible unfolding of the severe acute respiratory syndrome coronavirus main protease in guanidinium chloride. *Biophysical Journal*, 92, 1374-83.

Chen, C. N., Lin, C. P., Huang, K. K., Chen, W. C., Hsieh, H. P., Liang, P. H. & Hsu, J. T. 2005. Inhibition of SARS-CoV 3C-like Protease Activity by Theaflavin-3,3'-digallate (TF3). *Evidence-Based Complementary and Alternative Medicine,* 2**,** 209-215.

Choi, H. K., Lee, S., Zhang, Y. P., Mckinney, B. R., Wengler, G., Rossmann, M. G. & Kuhn, R. J. 1996. Structural analysis of Sindbis virus capsid mutants involving assembly and catalysis. *Journal of molecular biology,* 262**,** 151-167.

Choi, H. K., Lu, G., Lee, S., Wengler, G. & Rossmann, M. G. 1997. Structure of Semliki Forest virus core protein. *Proteins,* 27**,** 345-359.

Choi, H. K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M. G. & Wengler, G. 1991. Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. *Nature,* 354**,** 37-43.

Cottam, E. M., M. C. Whelband & T. Wileman. 2014. Coronavirus Nsp6 Restricts Autophagosome Expansion. *Autophagy,* 10, 1426-1441.

Cowley, J. A., C. M. Dimmock, K. M. Spann and P. J. Walker. 2000. Gill-Associated Virus of Penaeus Monodon Prawns: An Invertebrate Virus with Orf1a and Orf1b Genes Related to Arteri- and Coronaviruses. *Journal of General Virology* 81, 1473-1484.

Cridge, A. G., Crowe-Mcauliffe, C., Mathew, S. F. & Tate, W. P. 2018. Eukaryotic translational termination efficiency is influenced by the 3′ nucleotides within the ribosomal mRNA channel. *Nucleic acids research,* 46**,** 1927-1944.

Csibra, E., Brierley, I. & Irigoyen, N. 2014. Modulation of stop codon read-through efficiency and its effect on the replication of murine leukemia virus. *Journal of virology,* 88**,** 10364-10376.

Debat, H. J. 2018. Expanding the size limit of RNA viruses: Evidence of a novel divergent nidovirus in California sea hare, with a~ 35.9 kb virus genome. *bioRxiv***,** 307678.

den Boon, J. A., E. J. Snijder, E. D. Chirnside, A. A. de Vries, M. C. Horzinek and W. J. Spaan. 1991. Equine Arteritis Virus Is Not a Togavirus but Belongs to the Coronaviruslike Superfamily. *Journal of virology,* 65, 2910-2920.

Deng, X., Hackbart, M., Mettelman, R. C., O'brien, A., Mielech, A. M., Yi, G., Kao, C. C. & Baker, S. C. 2017. Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proceedings of the National Academy of Sciences,* 114**,** E4251-E4260.

Deng, X., Stjohn, S. E., Osswald, H. L., O'brien, A., Banach, B. S., Sleeman, K., Ghosh, A. K., Mesecar, A. D. & Baker, S. C. 2014. Coronaviruses resistant to a 3C-like protease inhibitor are attenuated for replication and pathogenesis, revealing a low genetic barrier but high fitness cost of resistance. *Journal of virology,* 88**,** 11886-11898.

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 biology,* 8**,** 270-279.

Denison, M. R., Zoltick, P. W., Hughes, S. A., Giangreco, B., Olson, A. L., Perlman, S., Leibowitz, J. L. & Weiss, S. R. 1992. Intracellular processing of the N-terminal ORF 1a proteins of the coronavirus MHV-A59 requires multiple proteolytic events. *Virology*, 189, 274-284.

Dervas, E., Hepojoki, J., Laimbacher, A., Romero-Palomo, F., Jelinek, C., Keller, S., Smura, T., Hepojoki, S., Kipar, A. & Hetzel, U. 2017. Nidovirus-associated proliferative pneumonia in the green tree python (Morelia viridis). *Journal of virology,* 91**,** e00718-17.

Dinman, J. D. 2012. Control of gene expression by translational recoding. *Advances in protein chemistry and structural biology,* 86**,** 129-149.

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. *Proceedings of the National Academy of Sciences,* 88**,** 174-178.

Dougherty, W. G. & Semler, B. L. 1993. Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiology and Molecular Biology Reviews,* 57**,** 781-822.

Dreher, T. W. & Miller, W. A. 2006. Translational control in positive strand RNA plant viruses. *Virology,* 344**,** 185-197.

Drosten, C., S. Gunther, W. Preiser, S. van der Werf, H. R. Brodt, S. Becker, H. Rabenau, M. Panning, L. Kolesnikova, R. A. Fouchier, A. Berger, A. M. Burguiere, J. Cinatl, M. Eickmann, N. Escriou, K. Grywna, S. Kramme, J. C. Manuguerra, S. Muller, V. Rickerts, M. Sturmer, S. Vieth, H. D. Klenk, A. D. Osterhaus, H. Schmitz & H. W. Doerr. 2003. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *New England journal of medicine,* 348**,** 1967-1976.

Duckmanton, L., Luan, B., Devenish, J., Tellier, R. & Petric, M. 1997. Characterization of torovirus from human fecal specimens. *Virology,* 239**,** 158-168.

Egloff, M. P., F. Ferron, V. Campanacci, S. Longhi, C. Rancurel, H. Dutartre, E. J. Snijder, A. E. Gorbalenya, C. Cambillau & B. Canard. 2004. The Severe Acute Respiratory Syndrome-Coronavirus Replicative Protein Nsp9 Is a Single-Stranded Rna-Binding Subunit Unique in the RNA virus World. *Proceedings of the National Academy of Sciences,* 101**,** 3792-3796.

Fakruddin, M., Mohammad Mazumdar, R., Bin Mannan, K. S., Chowdhury, A. & Hossain, M. N. 2013. Critical Factors Affecting the Success of Cloning, Expression, and Mass Production of Enzymes by Recombinant *E. coli*. *ISRN Biotechnol,* 2013**,** 590587.

Fan, K., Wei, P., Feng, Q., Chen, S., Huang, C., Ma, L., Lai, B., Pei, J., Liu, Y., Chen, J. & Lai, L. 2004. Biosynthesis, purification, and substrate specificity of severe acute respiratory syndrome coronavirus 3C-like proteinase. *Journal of Biological Chemistry,* 279**,** 1637-1642.

Farabaugh, P. J. 1996. Programmed translational frameshifting. *Microbiol Rev,* 60**,** 103-34.

Fauquet, C., Mayo, M.A., Maniloff, J., Desselberger, U. and Ball, L.A. (Eds.). Virus Taxonomy: Classification and nomenclature of viruses: Family Bromoviridae. Eighth Report of the International Committee on Taxonomy of Viruses. Amsterdam, Elsevier/Academic Press; 2005, pp. 937-954.

Fehr, A. R., & Perlman, S. Coronaviruses: an overview of their replication and pathogenesis: In Coronaviruses. New York, Humana Press; 2015, pp. 1-23.

Fiedler, T. J., Hudder, A., Mckay, S. J., Shivkumar, S., Capo, T. R., Schmale, M. C. & Walsh, P. J. 2010. The transcriptome of the early life history stages of the California Sea Hare Aplysia californica. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 5, 165-170.

Firth, A. E. & J. F. Atkins. 2009. Evidence for a Novel Coding Sequence Overlapping the 5 '-Terminal Similar to 90 Codons of the Gill-Associated and Yellow Head Okavirus Envelope Glycoprotein Gene. *Virology Journal,* 6, 1.

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

Fouchier, R. A. M., N. G. Hartwig, T. M. Bestebroer, B. Niemeyer, J. C. de Jong, J. H. Simon & A. D. M. E. Osterhaus. 2004. A Previously Undescribed Coronavirus Associated with Respiratory Disease in Humans. *Proceedings of the National Academy of Sciences,* 101**,** 6212-6216.

Frank, J., Gao, H., Sengupta, J., Gao, N. & Taylor, D. J. 2007. The process of mRNA-tRNA translocation. *Proceedings of the National Academy of Sciences,* 104**,** 19671-19678.

Freeman, M. C., Graham, R. L., Lu, X., Peek, C. T. & Denison, M. R. 2014. Coronavirus replicase-reporter fusions provide quantitative analysis of replication and replication complex formation. *Journal of virology,* 88**,** 5319-5327.

Fung, T. S. & D. X. Liu. 2014. Coronavirus infection, ER Stress, Apoptosis and Innate Immunity. *Frontiers in microbiology,* 5**,** 296.

Gadlage, M. J. & M. R. Denison. 2010. Exchange of the Coronavirus Replicase Polyprotein Cleavage Sites Alters Protease Specificity and Processing. *Journal of Virology,* 84, 6894-6898.

Giedroc, D. P., Theimer, C. A. & Nixon, P. L. 2000. Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *Journal of molecular biology,* 298**,** 167-185.

Gonzaalez, J. M., P. Gomez-Puertas, D. Cavanagh, A. E. Gorbalenya & L. Enjuanes. 2003. A Comparative Sequence Analysis to Revise the Current Taxonomy of the Family Coronaviridae. *Archives of Virology,* 148, 2207-2235.

Gonzalez-Ochoa, G., Menchaca, G. E., Hernandez, C. E., Rodriguez, C., Tamez, R. S. & Contreras, J. F. 2013. Mutation distribution in the NSP4 protein in rotaviruses isolated from Mexican children with moderate to severe gastroenteritis. *Viruses,* 5**,** 792-805.

Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko & V. M. Blinov. 1989. Coronavirus Genome: Prediction of Putative Functional Domains in the Non-Structural Polyprotein by Comparative Amino Acid Sequence Analysis. *Nucleic acids research,* 17**,** 4847-4861.

Gorbalenya, A. E., L. Enjuanes, J. Ziebuhr & E. J. Snijder. 2006. Nidovirales: Evolving the Largest RNA Virus Genome. *Virus Research,* 117, 17-37.

Gorbalenya, A.E., Brinton, M.A., Cowley, J., de Groot, R., Gulyaeva, A., Lauber, C., Neuman, B.W., Ziebuhr, J., 2017. ICTV Pending Proposal 2017.015S. Reorganization and expansion of the order Nidovirales at the family and sub-order ranks.

Gouilh, M. A., Puechmaille, S. J., Gonzalez, J. P., Teeling, E., Kittayapong, P. & Manuguerra, J. C. 2011. SARS-Coronavirus ancestor's foot-prints in South-East Asian bat colonies and the refuge theory. *Infection, Genetics and Evolution,* 11**,** 1690-1702.

Grossoehme, N. E., L. C. Li, S. C. Keane, P. H. Liu, C. E. Dann, J. L. Leibowitz & D. P. Giedroc. 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.

Hahn, C. S. & Strauss, J. H. 1990. Site-directed mutagenesis of the proposed catalytic amino acids of the Sindbis virus capsid protein autoprotease. *Journal of Virology,* 64**,** 3069-3073.

Hardy, W. R. & Strauss, J. H. 1989. Processing the nonstructural polyproteins of sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in cis and in trans. *Journal of virology,* 63**,** 4653-4664.

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

Hartl, F. U., Bracher, A. & Hayer-Hartl, M. 2011. Molecular chaperones in protein folding and proteostasis. *Nature,* 475**,** 324-332.

Hedstrom, L. 2002. Serine protease mechanism and specificity. *Chemical reviews,* 102**,** 4501-4524.

Heyland, A., Vue, Z., Voolstra, C. R., Medina, M. & Moroz, L. L. 2011. Developmental transcriptome of Aplysia californica'. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution,* 316**,** 113-134.

Hilgenfeld, R. 2014. From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *The FEBS journal,* 281**,** 4085-4096.

Hogue, B. G. & C. E. Machamer. 2008. Coronavirus Structural Proteins and Virus Assembly. *Nidoviruses*, 179-200.

Hong, E. M., Perera, R. & Kuhn, R. J. 2006. Alphavirus capsid protein helix I controls a checkpoint in nucleocapsid core assembly. *Journal of virology,* 80**,** 8848-8855.

Howard, M. T., Aggarwal, G., Anderson, C. B., Khatri, S., Flanigan, K. M. & Atkins, J. F. 2005. Recoding elements located adjacent to a subset of eukaryal selenocysteine-specifying UGA codons. *The EMBO journal,* 24**,** 1596-1607.

Hsieh, P. K., S. C. Chang, C. C. Huang, T. T. Lee, C. W. Hsiao, Y. H. Kou, I. Y. Chen, C. K. Chang, T. H. Huang & M. F. Chang. 2005. Assembly of Severe Acute Respiratory Syndrome Coronavirus RNA Packaging Signal into Virus-Like Particles Is Nucleocapsid Dependent. *Journal of Virology,* 79, 13848-13855.

Hu, H. T., Cho, C. P., Lin, Y. H. & Chang, K. Y. 2016. A general strategy to inhibiting viral -1 frameshifting based on upstream attenuation duplex formation. *Nucleic acids research,* 44**,** 256-266.

Huang, C., Wei, P., Fan, K., Liu, Y. & Lai, L. 2004. 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. *Biochemistry,* 43, 4568-4574.

Huang, Y., Z. Y. Yang, W. P. Kong & G. J. Nabel. 2004. Generation of Synthetic Severe Acute Respiratory Syndrome Coronavirus Pseudoparticles: Implications for Assembly and Vaccine Production. *Journal of Virology,* 78, 12557-12565.

Irigoyen, N., Dinan, A. M., Brierley, I. & Firth, A. E. 2018. Ribosome profiling of the retrovirus murine leukemia virus. *Retrovirology,* 15**,** 10.

Ivanov,I.P., Gesteland,R.F. & Atkins,J.F. 2000. Antizyme expression: a subversion of triplet decoding, which is remarkably conserved by evolution, is a sensor for an autoregulatory circuit. *Nucleic acids research,* 28, 3185–3196.

Jacobus, A. P. & Gross, J. 2015. Optimal cloning of PCR fragments by homologous recombination in Escherichia coli. *PLoS One,* 10**,** e0119221.

Jarvis, D. L. 2009. Baculovirus-insect cell expression systems. *Methods Enzymol,* 463, 191-222.

Jauregui, A. R., Savalia, D., Lowry, V. K., Farrell, C. M. & Wathelet, M. G. 2013. Identification of residues of SARS-CoV nsp1 that differentially affect inhibition of gene expression and antiviral signaling. *PLoS One,* 8**,** e62416.

John, S. E. S., Tomar, S., Stauffer, S. R. & Mesecar, A. D. 2015. Targeting zoonotic viruses: Structure-based inhibition of the 3C-like protease from bat coronavirus HKU4—The likely reservoir host to the human coronavirus that causes Middle East Respiratory Syndrome (MERS). *Bioorganic & medicinal chemistry,* 23**,** 6036-6048.

Kelly, B. J., King, L. A. & Possee, R. D. 2016. Introduction to Baculovirus Molecular Biology. *Springer,* 1350, 25-50.

Ketteler, R. 2012. On programmed ribosomal frameshifting: the alternative proteomes. *Frontiers in genetics,* 3**,** 242.

Kirchdoerfer, R. N., Cottrell, C. A., Wang, N., Pallesen, J., Yassine, H. M., Turner, H. L., Corbett, K. S., Graham, B. S., Mclellan, J. S. & Ward, A. B. 2016. Pre-fusion structure of a human coronavirus spike protein. Nature, 531, 118-121.

Kleine-Weber, H., Elzayat, M. T., Hoffmann, M. & Pohlmann, S. 2018. Functional analysis of potential cleavage sites in the MERS-coronavirus spike protein. *Scientific reports,* 8**,** 16597.

Knoops, K., M. Barcena, R. W. A. L. Limpens, A. J. Koster, A. M. Mommaas & E. J. Snijder. 2012. Ultrastructural Characterization of Arterivirus Replication Structures: Reshaping the Endoplasmic Reticulum to Accommodate Viral Rna Synthesis. *Journal of Virology,* 86, 2474-2487.

Krishna, S Sri, Indraneel Majumdar & Nick V Grishin. 2003. Structural Classification of Zinc Fingers Survey and Summary. *Nucleic acids research* 31, 532-550.

Ksiazek, T. G., D. Erdman, C. S. Goldsmith, S. R. Zaki, T. Peret, S. Emery, S. Tong, C. Urbani, J. A. Comer, W. Lim, P. E. Rollin, S. F. Dowell, A. E. Ling, C. D. Humphrey, W. J. Shieh, J. Guarner, C. D. Paddock, P. Rota, B. Fields, J. DeRisi, J. Y. Yang, N. Cox, J. M. Hughes, J. W. LeDuc, W. J. Bellini, L. J. Anderson & Sars Working Group. 2003. A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *New England journal of medicine,* 348**,** 1953-1966.

Kuhn, J. H., Lauck, M., Bailey, A. L., Shchetinin, A. M., Vishnevskaya, T. V., Bao, Y., Ng, T. F., Lebreton, M., Schneider, B. S., Gillis, A., Tamoufe, U., Diffo Jle, D., Takuo, J. M., Kondov, N. O., Coffey, L. L., Wolfe, N. D., Delwart, E., Clawson, A. N., Postnikova, E., Bollinger, L., Lackemeyer, M. G., Radoshitzky, S. R., Palacios, G., Wada, J., Shevtsova, Z. V., Jahrling, P. B., Lapin, B. A., Deriabin, P. G., Dunowska, M., Alkhovsky, S. V., Rogers, J., Friedrich, T. C., O'connor, D. H. & Goldberg, T. L. 2016. Reorganization and expansion of the nidoviral family Arteriviridae. *Archives of Virology,* 161, 755-68.

La Monica, N., Yokomori, K. & Lai, M. M. 1992. Coronavirus mRNA synthesis: identification of novel transcription initiation signals which are differentially regulated by different leader sequences. *Virology*, 188, 402-407.

Lai MM, Perlman S, Anderson LJ. Coronaviridae. In: Knipe DM, Howley PM, eds. Fields Virology. Philadelphia. Lippincott Williams & Wilkins; 2007.

Lau, S. K. P., Wong, E. Y. M., Tsang, C. C., Ahmed, S. S., Au-Yeung, R. K. H., Yuen, K. Y., Wernery, U. & Woo, P. C. Y. 2018. Discovery and Sequence Analysis of Four Deltacoronaviruses from Birds in the Middle East Reveal Interspecies Jumping with Recombination as a Potential Mechanism for Avian-to-Avian and Avian-to-Mammalian Transmission. *Journal of virology,* 92**,** e00265-18.

Lauber, C., J. J. Goeman, M. D. Parquet, P. T. Nga, E. J. Snijder, K. Morita & A. E. Gorbalenya. 2013. The Footprint of Genome Architecture in the Largest Genome Expansion in Rna Viruses. *Plos Pathogens* 9, e1003500.

Lauber, C., J. Ziebuhr, S. Junglen, C. Drosten, F. Zirkel, P. T. Nga, K. Morita, E. J. Snijder &. A. E. Gorbalenya. 2012. Mesoniviridae: A Proposed New Family in the Order Nidovirales Formed by a Single Species of Mosquito-Borne Viruses. *Archives of Virology,* 157, 1623-1628.

Lauck, M., Alkhovsky, S. V., Bao, Y., Bailey, A. L., Shevtsova, Z. V., Shchetinin, A. M., Vishnevskaya, T. V., Lackemeyer, M. G., Postnikova, E., Mazur, S., Wada, J., Radoshitzky, S. R., Friedrich, T. C., Lapin, B. A., Deriabin, P. G., Jahrling, P. B., Goldberg, T. L., O'connor, D. H. & Kuhn, J. H. 2015. Historical Outbreaks of Simian Hemorrhagic Fever in Captive Macaques Were Caused by Distinct Arteriviruses. *Journal of virology,* 89**,** 8082-8087.

Lee, H. J., Shieh, C. K., Gorbalenya, A. E., Koonin, E. V., LA Monica, N., Tuler, J., Bagdzhadzhyan, A. & Lai, M. M. 1991. The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. *Virology,* 180, 567-582.

Lee, J. M., Cho, J. B., Ahn, H. C., Jung, W. & Jeong, Y. J. 2017. A Novel Chemical Compound for Inhibition of SARS Coronavirus Helicase. *Journal of microbiology and biotechnology,* 27**,** 2070-2073.

Lei, J. & Hilgenfeld, R. 2017. RNA-virus proteases counteracting host innate immunity. *FEBS letters,* 591**,** 3190-3210.

Lei, J., Kusov, Y. & Hilgenfeld, R. 2018. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral research,* 149**,** 58-74.

Li G, Rice CM. 1993. The signal for translational readthrough of a UGA codon in Sindbis virus RNA involves a single cytidine residue immediately downstream of the termination codon. *Journal of virology*, 67, 5062–5067.

Li, F. 2015. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *Journal of virology,* 89**,** 1954-1964.

Li, L., Jose, J., Xiang, Y., Kuhn, R. J. & Rossmann, M. G. 2010. Structural changes of envelope proteins during alphavirus fusion. *Nature,* 468**,** 705-708.

Licznar, P., Mejlhede, N., Prere, M., Wills, N., Gesteland, R.F., Atkins, J.F. & Fayet, O. 2003. Programmed translational -1 frameshifting on hexanucleotide motifs and the wobble properties of tRNAs. *The EMBO journal,* 22**,** 4770-4778.

Liphardt, J., Napthine, S., Kontos, H., Brierley, I., 1999. Evidence for an RNA pseudoknot loop–helix interaction essential for efficient-1 ribosomal frameshifting. *Journal of molecular biology,* 288**,** 321-335.

Liu, D. X. & Brown, T. D. 1995. Characterisation and mutational analysis of an ORF 1a-encoding proteinase domain responsible for proteolytic processing of the infectious bronchitis virus 1a/1b polyprotein. *Virology,* 209, 420-427.

Loh PG & Song H. 2010. Structural and mechanistic insights into translation termination. *Current opinion in structural biology,* 20**,** 98-103.

Loughran, G., Jungreis, I., Tzani, I., Power, M., Dmitriev, R. I., Ivanov, I. P., Kellis, M. & Atkins, J. F. 2018. Stop codon readthrough generates a C-terminally extended variant of the human vitamin D receptor with reduced calcitriol response. *Journal of Biological Chemistry,* 293**,** 4434-4444.

Lu, Y., Lu, X. & Denison, M. R. 1995. Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. *Journal of virology,* 69, 3554-3559.

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**,** 1399-1404.

Masters, P. S. 2006. The molecular biology of coronaviruses. *Advances in virus research,* 66**,** 193-292.

Mcbride, R., Van Zyl, M. & Fielding, B. C. 2014. The coronavirus nucleocapsid is a multifunctional protein. *Viruses,* 6**,** 2991-3018.

McCaughan KK, Brown CM, Dalphin ME, Berry MJ, Tate WP. 1995. Translational termination efficiency in mammals is influenced by the base following the stop codon. *Proceedings of the National Academy of Sciences,* 92**,** 5431-5435.

Melancon, P. & Garoff, H. 1987. Processing of the Semliki Forest virus structural polyprotein: role of the capsid protease. *Journal of virology,* 61**,** 1301-1309.

Menachery, V. D., Gralinski, L. E., Mitchell, H. D., Dinnon, K. H., 3RD, Leist, S. R., Yount, B. L., JR., Mcanarney, E. T., Graham, R. L., Waters, K. M. & Baric, R. S. 2018. Combination Attenuation Offers Strategy for Live Attenuated Coronavirus Vaccines. *Journal of virology,* 92**,** e00710-18.

Metsikko, K. & Garoff, H. 1990. Oligomers of the cytoplasmic domain of the p62/E2 membrane protein of Semliki Forest virus bind to the nucleocapsid in vitro. *Journal of virology,* 64**,** 4678-83.

Miranda, J. A., Culley, A. I., Schvarcz, C. R. & Steward, G. F. 2016. RNA viruses as major contributors to Antarctic virioplankton. *Environmental microbiology,* 18**,** 3714-3727.

Moroz, L. L., Edwards, J. R., Puthanveettil, S. V., Kohn, A. B., Ha, T., HEYLAND, A., Knudsen, B., Sahni, A., Yu, F., Liu, L., Jezzini, S., Lovell, P., Iannucculli, W., Chen, M., Nguyen, T., Sheng, H., Shaw, R., Kalachikov, S., Panchin, Y. V., Farmerie, W., Russo, J. J., Ju, J. & Kandel, E. R. 2006. Neuronal transcriptome of Aplysia: neuronal compartments and circuitry. *Cell,* 127**,** 1453-1467.

Nakagawa, K., Lokugamage, K. G. & Makino, S. 2016. Viral and Cellular mRNA Translation in Coronavirus-Infected Cells. *Advances in virus research. Elsevier,* 96**,** 165-192.

Namy O, Rousset J-P. 2010. Specification of standard amino acids by stop codons. In Recoding: Expansion of decoding rules enriches gene expression (ed. JF Atkins, RF Gesteland), pp 79–100. Springer Science and Business Media, New York.

Namy O, Rousset JP, Napthine S, Brierley I. 2004. Reprogrammed genetic decoding in cellular gene expression. *Molecular cell,* 13**,** 157-168.

Napthine, S., Liphardt, J., Bloys, A., Routledge, S., Brierley, I., 1999. The role of RNA pseudoknot stem 1 length in the promotion of efficient-1 ribosomal frameshifting. *Journal of molecular biology,* 288**,** 305-320.

Napthine, S., Yek, C., Powell, M. L., Brown, T. D. & Brierley, I. 2012. Characterization of the stop codon readthrough signal of Colorado tick fever virus segment 9 RNA. *RNA,* 18**,** 241-252.

Narayanan, K., Ramirez, S. I., Lokugamage, K. G. & Makino, S. 2015. Coronavirus nonstructural protein 1: Common and distinct functions in the regulation of host and viral gene expression. *Virus Research,* 202**,** 89-100.

Nauwynck, H., Snijder, E.J., Faaberg, Balasuriya, U., Brinton, M., Gorbalenya, A.E., Leung, F.-C., Stadejek, T., Yang, H., Yoo, D., 2012. Family Arteriviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) Virus Taxonomy:IXth report of the International Committee of Taxonomy.

Needle, D., Lountos, G. T., & Waugh, D. S. 2015. Structures of the Middle East respiratory syndrome coronavirus 3C-like protease reveal insights into substrate specificity. *Acta Crystallographica Section D: Biological Crystallography,* 71**,** 1102-1111.

Neuman, B. W. 2008. Supramolecular Architecture of the Coronavirus Particle. *Nidoviruses*, 201-210.

Neurath, H. 1984. Evolution of proteolytic enzymes. *Science*, 224, 350-357.

Nga, P. T., Parquet Mdel, C., Lauber, C., Parida, M., Nabeshima, T., Yu, F., Thuy, N. T., Inoue, S., Ito, T., Okamoto, K., Ichinose, A., Snijder, E. J., Morita, K. & Gorbalenya, A. E. 2011. Discovery of the first insect nidovirus, a missing evolutionary link in the emergence of the largest RNA virus genomes. *PLoS Pathogens,* 7**,** e1002215.

Niemeyer, D., Mosbauer, K., Klein, E. M., Sieberg, A., Mettelman, R. C., Mielech, A. M., Dijkman, R., Baker, S. C., Drosten, C. & Muller, M. A. 2018. The papain-like protease determines a virulence trait that varies among members of the SARS-coronavirus species. *PLoS Pathogenus,* 14**,** e1007296.

Nijholt, D. A., DE Kimpe, L., Elfrink, H. L., Hoozemans, J. J. & Scheper, W. 2011. Removing protein aggregates: the role of proteolysis in neurodegeneration. *Current medicinal chemistry,* 18**,** 2459-2476.

O'dea, M. A., Jackson, B., Jackson, C., Xavier, P. & Warren, K. 2016. Discovery and Partial Genomic Characterisation of a Novel Nidovirus Associated with Respiratory Disease in Wild Shingleback Lizards (Tiliqua rugosa). *PLoS One,* 11**,** e0165209.

O'flynn, N. M., Patel, A., Kadlec, J. & Jones, I. M. 2012. Improving promiscuous mammalian cell entry by the baculovirus Autographa californica multiple nuclear polyhedrosis virus. *Bioscience reports,* 33**,** 23-36.

Owen KE, Kuhn RJ. 1997. Alphavirus budding is dependent on the interaction between the nucleocapsid and hydrophobic amino acids on the cytoplasmic domain of the E2 envelope glycoprotein. *Virology*, 230**,** 187-196.

Park, J. E., Li, K., Barlan, A., Fehr, A. R., Perlman, S., Mccray, P. B., Jr. & Gallagher, T. 2016. Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism. *Proceedings of the National Academy of Sciences,* 113**,** 12262-12267.

Pasternak, A. O., W. J. M. Spaan & E. J. Snijder. 2006. Nidovirus Transcription: How to Make Sense ... ?. *Journal of General Virology*, 87, 1403-1421.

Peiris, J. S. M., S. T. Lai, L. L. M. Poon, Y. Guan, L. Y. C. Yam, W. Lim, J. Nicholls, W. K. S. Yee, W. W. Yan, M. T. Cheung, V. C. C. Cheng, K. H. Chan, D. N. C. Tsang, R. W. H. Yung, T. K. Ng, K. Y. Yuen & Sars Study Grp. 2003. Coronavirus as a Possible Cause of Severe Acute Respiratory Syndrome. *The Lancet,* 361**,** 1319-1325.

Petry, S., Weixlbaumer, A. & Ramakrishnan, V. 2008. The termination of translation. *Current Opinion Structural Biology,* 18, 70-77.

Posthuma, C. C., D. D. Nedialkova, J. C. Zevenhoven-Dobbe, J. H. Blokhuis, A. E. Gorbalenya & E. J. Snijder. 2005. Site-Directed Mutagenesis of the Nidovirus Replicative Endoribonuclease Nendou Exerts Pleiotropic Effects on the Arterivirus Life Cycle. *Journal of Virology*, 80, 1653-1661.

Powers, A., Huang, H., Roehrig, J., Strauss, E., Weaver, S. Family Togaviridae. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. New York, Elsevier Academic Press; 2012.

Rascon-Castelo, E., Burgara-Estrella, A., Mateu, E. & Hernandez, J. 2015. Immunological features of the non-structural proteins of porcine reproductive and respiratory syndrome virus. *Viruses*, 7**,** 873-886.

Robertson, A. L., Headey, S. J., Ng, N. M., Wijeyewickrema, L. C., Scanlon, M. J., Pike, R. N. & Bottomley, S. P. 2016. Protein unfolding is essential for cleavage within the alpha-helix of a model protein substrate by the serine protease, thrombin. *Biochimie,* 122**,** 227-234.

Rota, P. A., M. S. Oberste, S. S. Monroe, W. A. Nix, R. Campagnoli, J. P. Icenogle, S. Penaranda, B. Bankamp, K. Maher, M. H. Chen, S. X. Tong, A. Tamin, L. Lowe, M. Frace, J. L. DeRisi, Q. Chen, D. Wang, D. D. Erdman, T. C. T. Peret, C. Burns, T. G. Ksiazek, P. E. Rollin, A. Sanchez, S. Liffick, B. Holloway, J. Limor, K. McCaustland, M. Olsen-Rasmussen, R. Fouchier, S. Gunther, A. D. M. E. Osterhaus, C. Drosten, M. A. Pallansch, L. J. Anderson & W. J. Bellini. 2003. Characterization of a Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *Science,* 300, 1394-1399.

Ruch, T. R. & C. E. Machamer. 2012. The Coronavirus E Protein: Assembly and Beyond. *Viruses-Basel,* 4, 363-382.

Rupp, J. C., Sokoloski, K. J., Gebhart, N. N. & Hardy, R. W. 2015. Alphavirus RNA synthesis and non-structural protein functions. *The Journal of general virology,* 96**,** 2483.

Saberi, A., Gulyaeva, A. A., Brubacher, J. L., Newmark, P. A. & Gorbalenya, A. E. 2018. A planarian nidovirus expands the limits of RNA genome size. *PLoS pathogens,* 14**,** e1007314.

Saikatendu, K. S., Joseph, J. S., Subramanian, V., Clayton, T., Griffith, M., Moy, K., Velasquez, J., Neuman, B. W., Buchmeier, M. J. & Stevens, R. C. 2005. Structural basis of severe acute respiratory syndrome coronavirus ADP-ribose-1 ″-phosphate dephosphorylation by a conserved domain of nsP3. *Structure,* 13**,** 1665-1675.

Sakai, Y., Kawachi, K., Terada, Y., Omori, H., Matsuura, Y. & Kamitani, W. 2017. Two-amino acids change in the nsp4 of SARS coronavirus abolishes viral replication. *Virology,* 510**,** 165-174.

Sawicki, S. G. & D. L. Sawicki. 1995. Coronaviruses Use Discontinuous Extension for Synthesis of Subgenome-Length Negative Strands. *Springer,* 380, 499-506.

Sawicki, S. G., D. L. Sawicki & S. G. Siddell. 2007. A Contemporary View of Coronavirus Transcription. *Journal of Virology,* 81, 20-29.

Sawicki, S. G., D. L. Sawicki, D. Younker, Y. Meyer, V. Thiel, H. Stokes & S. G. Siddell. 2005. Functional and Genetic Analysis of Coronavirus Replicase-Transcriptase Proteins. *PLoS pathogens,* 1, e39.

Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature methods,* 9**,** 671.

Schutze, H., R. Ulferts, B. Schelle, S. Bayer, H. Granzow, B. Hoffmann, T. C. Mettenleiter & J. Ziebuhr. 2006. Characterization of White Bream Virus Reveals a Novel Genetic Cluster of Nidoviruses. *Journal of Virology,* 80, 11598-11609.

Seibert, C. H. & Pinto, A. R. 2012. Challenges in shrimp aquaculture due to viral diseases: distribution and biology of the five major penaeid viruses and interventions to avoid viral incidence and dispersion. *Brazilian Journal of Microbiology,* 43**,** 857-864.

Sexton, N. R., Smith, E. C., Blanc, H., Vignuzzi, M., Peersen, O. B. & Denison, M. R. 2016. Homology-Based Identification of a Mutation in the Coronavirus RNA-Dependent RNA Polymerase That Confers Resistance to Multiple Mutagens. *Journal of Virology,* 90**,** 7415-7428.

Seybert, A., C. C. Posthuma, L. C. van Dinten, E. J. Snijder, A. E. Gorbalenya & J. Ziebuhr. 2005. A Complex Zinc Finger Controls the Enzymatic Activities of Nidovirus Helicases. *Journal of Virology,* 79, 696-704.

Seybert, A., Ziebuhr, J. & Siddell, S. G. 1997. Expression and characterization of a recombinant murine coronavirus 3C-like proteinase. *Journal of General Virology,* 78**,** 71-75.

Shan, Y. F., Li, S. F. & Xu, G. J. 2004. A novel auto-cleavage assay for studying mutational effects on the active site of severe acute respiratory syndrome coronavirus 3C-like protease. *Biochemical and biophysical research communications,* 324**,** 579-583.

Shang, J., Zheng, Y., Yang, Y., Liu, C., Geng, Q., Luo, C., Zhang, W. & Li, F. 2018. Cryo-EM structure of infectious bronchitis coronavirus spike protein reveals structural and functional evolution of coronavirus spike proteins. *PLoS pathogens,* 14, e1007009.

Shang, P., Misra, S., Hause, B. & Fang, Y. 2017. A Naturally Occurring Recombinant Enterovirus Expresses a Torovirus Deubiquitinase. *Journal of virology,* 91**,** e00450-17.

Shi, J., Wei, Z. & Song, J. 2004. Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme: defining the extra domain as a new target for design of highly specific protease inhibitors. *Journal of Biological Chemistry,* 279**,** 24765-24773.

Shi, M., Lin, X. D., Chen, X., Tian, J. H., Chen, L. J., Li, K., Wang, W., Eden, J. S., Shen, J. J., Liu, L., Holmes, E. C. & Zhang, Y. Z. 2018. The evolutionary history of vertebrate RNA viruses. *Nature,* 556**,** 197-202.

Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X., Qin, X. C., Li, J., Cao, J. P., Eden, J. S., Buchmann, J., Wang, W., Xu, J., Holmes, E. C. & Zhang, Y. Z. 2016. Redefining the invertebrate RNA virosphere. *Nature*, 540**,** 539.

Shi, M., Zhang, Y. Z. & Holmes, E. C. 2018. Meta-transcriptomics and the evolutionary biology of RNA viruses. *Virus research,* 243**,** 83-90.

Sivashanmugam, A., Murray, V., Cui, C., Zhang, Y., Wang, J. & LI, Q. 2009. Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. *Protein Science,* 18**,** 936-948.

Skoging, U. & Liljestrom, P. 1998. Role of the C-terminal tryptophan residue for the structure-function of the alphavirus capsid protein. *Journal of molecular biology,* 279**,** 865-872.

Skuzeski JM, Nichols LM, Gesteland RF, Atkins JF. 1991. The signal for a leaky UAG stop codon in several plant viruses includes the two downstream codons. *Journal of molecular biology,* 218, 365-373.

Snijder, E. J., M. Kikkert & Y. Fang. 2013. Arterivirus Molecular Biology and Pathogenesis. *Journal of General Virology,* 94, 2141-2163.

Snijder, Eric J, Stuart G Siddell & Alexander E Gorbalenya. 2005. The Order Nidovirales. *Topley and Wilson's microbiology and microbial infections*.

Sonawane, N. D., Szoka, F. C., JR. & Verkman, A. S. 2003. Chloride accumulation and swelling in endosomes enhances DNA transfer by polyamine-DNA polyplexes. *Journal of Biological Chemistry,* 278**,** 44826-44831.

Stahl, G., McCarty, G.P., Farabaugh, P.J., 2002. Ribosome structure: revisiting the connection between translational accuracy and unconventional decoding. *Trends Biochem Trends in biochemical sciences,* 27**,** 178-183.

Steel, J. J. & Geiss, B. J. 2015. A novel system for visualizing alphavirus assembly. *Journal of virological methods,* 222**,** 158-163.

Stobart, C. C., Sexton, N. R., Munjal, H., Lu, X., Molland, K. L., Tomar, S., Mesecar, A. D. & Denison, M. R. 2013. Chimeric exchange of coronavirus nsp5 proteases (3CLpro) identifies common and divergent regulatory determinants of protease activity. *Journal of virology,* 87**,** 12611-12618.

Strauss, E. G., Rice, C. M. & Strauss, J. H. 1984. Complete nucleotide sequence of the genomic RNA of Sindbis virus. *Virology,* 133**,** 92-110.

Strauss, J. H. & Strauss, E. G. 1994. The alphaviruses: gene expression, replication, and evolution. *Microbiology and Molecular Reviews,* 58**,** 491-562.

Studier, F. W. & Maizel, J. V., JR. 1969. T7-directed protein synthesis. *Virology,* 39, 575-586.

Sun, D., Chen, S., Cheng, A., & Wang, M. 2016. Roles of the picornaviral 3C proteinase in the viral life cycle and host cells. *Viruses,* 8**,** 82.

Tan, J., George, S., Kusov, Y., Perbandt, M., Anemüller, S., Mesters, J. R., Norder, H., Coutard, B., Lacroix, C. & Leyssen, P. 2013. 3C protease of enterovirus 68: structure-based design of Michael acceptor inhibitors and their broad-spectrum antiviral effects against picornaviruses. *Journal of virology,* 87**,** 4339-4351.

Tan, J., Verschueren, K. H., Anand, K., Shen, J., Yang, M., Xu, Y., Rao, Z., Bigalke, J., Heisen, B. & Mesters, J. R. 2005. pH-dependent conformational flexibility of the SARS-CoV main proteinase (Mpro) dimer: molecular dynamics simulations and multiple X-ray structure analyses. *Journal of molecular biology*, 354, 25-40.

Tan, Y. W., S. G. Fang, H. Fan, J. Lescar & D. X. Liu. 2006. Amino Acid Residues Critical for Rna-Binding in the N-Terminal Domain of the Nucleocapsid Protein Are Essential Determinants for the Infectivity of Coronavirus in Cultured Cells. *Nucleic Acids Research,* 34, 4816-4825.

Tanaka, Y., Sato, Y. & Sasaki, T. 2013. Suppression of coronavirus replication by cyclophilin inhibitors. *Viruses,* 5**,** 1250-1260.

Ten Dam, E. B., Pleij, C. W. & Bosch, L. 1990. RNA pseudoknots: translational frameshifting and readthrough on viral RNAs. *Virus genes,* 4**,** 121-136.

Thiel, V., Ivanov, K. A., Putics, A., Hertzig, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E. J., Rabenau, H., Doerr, H. W., Gorbalenya, A. E. & Ziebuhr, J. 2003. Mechanisms and enzymes involved in SARS coronavirus genome expression. *Journal of General Virology,* 84**,** 2305-2315.

Thomas, S., Rai, J., John, L., Gunther, S., Drosten, C., Putzer, B. M. & Schaefer, S. 2010. Functional dissection of the alphavirus capsid protease: sequence requirements for activity. *Virology journal,* 7**,** 327.

Tibbles, K. W., Brierley, I., Cavanagh, D. & Brown, T. D. 1996. Characterization in vitro of an autocatalytic processing activity associated with the predicted 3C-like proteinase domain of the coronavirus avian infectious bronchitis virus. *Journal of Virology,* 70**,** 1923-1930.

Tokarz, R., Sameroff, S., Hesse, R. A., Hause, B. M., Desai, A., Jain, K. & Lipkin, W. I. 2015. Discovery of a novel nidovirus in cattle with respiratory disease. *Journal of General Virology,* 96**,** 2188-93.

Tseng, Y. T., S. M. Wang, K. J. Huang, A. I. R. Lee, C. C. Chiang & C. T. Wang. 2010. Self-Assembly of Severe Acute Respiratory Syndrome Coronavirus Membrane Protein. *Journal of Biological Chemistry,* 285, 12862-12872.

Ulasli, M., M. H. Verheije, C. A. M. de Haan & F. Reggiori. 2010. Qualitative and Quantitative Ultrastructural Analysis of the Membrane Rearrangements Induced by Coronavirus. *Cellular Microbiology,* 12, 844-861.

Ulferts, R., Mettenleiter, T. C. & Ziebuhr, J. 2011. Characterization of Bafinivirus main protease autoprocessing activities. *Journal of Virology,* 85**,** 1348-59.

Van der Hoek, L., K. Pyrc, M. F. Jebbink, W. Vermeulen-Oost, R. J. M. Berkhout, K. C. Wolthers, P. M. E. Wertheim-van Dillen, J. Kaandorp, J. Spaargaren & B. Berkhout. 2004. Identification of a New Human Coronavirus. *Nature Medicine,* 10, 368-373.

Van Dinten, L. C., H. van Tol, A. E. Gorbalenya & E. J. Snijder. 2000. The Predicted Metal-Binding Region of the Arterivirus Helicase Protein Is Involved in Subgenomic Mrna Synthesis, Genome Replication, and Virion Biogenesis. *Journal of Virology,* 74, 5213-5223.

Van Hemert, M. J., Van Den Worm, S. H., Knoops, K., Mommaas, A. M., Gorbalenya, A. E. & Snijder, E. J. 2008. SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity in vitro. *PLoS pathogens,* 4**,** e1000054.

Vasilakis, N., Guzman, H., Firth, C., Forrester, N. L., Widen, S. G., Wood, T. G., Rossi, S. L., Ghedin, E., Popov, V., Blasdell, K. R., Walker, P. J. & Tesh, R. B. 2014. Mesoniviruses are mosquito-specific viruses with extensive geographic distribution and host range. *Virology journal,* 11**,** 97.

Velthuis, A. J. W. T., J. J. Arnold, C. E. Cameron, S. H. E. van den Worm & E. J. Snijder. 2010. The RNA polymerase Activity of Sars-Coronavirus Nsp12 Is Primer Dependent. *Nucleic Acids Research,* 38, 203-214.

Velthuis, A. J. W. T., S. H. E. van den Worm & E. J. Snijder. 2012. The Sars-Coronavirus Nsp7+Nsp8 Complex Is a Unique Multimeric RNA polymerase Capable of Both De Novo Initiation and Primer Extension. *Nucleic Acids Research,* 40, 1737-1747.

Von der Haar T, Tuite MF. 2007. Regulated translational bypass of stop codons in yeast. *Trends in microbiology,* 15**,** 78-86.

Wahl-Jensen, V., Johnson, J. C., Lauck, M., Weinfurter, J. T., Moncla, L. H., Weiler, A. M., Charlier, O., Rojas, O., Byrum, R., Ragland, D. R., Huzella, L., Zommer, E., Cohen, M., Bernbaum, J. G., Cai, Y., Sanford, H. B., Mazur, S., Johnson, R. F., Qin, J., Palacios, G. F., Bailey, A. L., Jahrling, P. B., Goldberg, T. L., O'connor, D. H., Friedrich, T. C. & Kuhn, J. H. 2016. Divergent Simian Arteriviruses Cause Simian Hemorrhagic Fever of Differing Severities in Macaques. *MBio,* 7**,** e02009-15.

Wang, Jibin, Shouguo Fang, Han Xiao, Bo Chen, James P Tam & Ding Xiang Liu. 2009. Interaction of the Coronavirus Infectious Bronchitis Virus Membrane Protein with B-Actin and Its Implication in Virion Assembly and Budding. *PLoS One,* 4, e4908.

Wang, R., Xiong, J., Wang, W., Miao, W. & Liang, A. 2016. High frequency of +1 programmed ribosomal frameshifting in Euplotes octocarinatus. *Scientific reports,* 6**,** 21139.

Woo, P. C. Y., S. K. P. Lau, C. M. Chu, K. H. Chan, H. W. Tsoi, Y. Huang, B. H. L. Wong, H. L. Wong, R. W. S. Poon, J. J. Cai, W. K. Luk, L. L. M. Poon, S. S. Y. Wong, Y. Guan, J. S. M. Peiris & K. Y. Yuen. 2005. Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus Hku1, from Patients with Pneumonia. *Journal of Virology,* 79, 884-895.

WU, C. H., CHEN, P. J. & YEH, S. H. 2014. Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcription. *Cell Host Microbe,* 16**,** 462-72.

Yang, H., Yang, M., Ding, Y., Liu, Y., Lou, Z., Zhou, Z., Sun, L., Mo, L., Ye, S., Pang, H., Gao, G. F., Anand, K., Bartlam, M., Hilgenfeld, R. & RAO, Z. 2003. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proceedings of the National Academy of Sciences,* 100**,** 13190-13195.

Zaki, J., Weber, J. & Ochsner, K. 2012. Task-dependent neural bases of perceiving emotionally expressive targets. *Frontiers in human neuroscience,* 6**,** 228.

Zeng, Z., Deng, F., Shi, K., Ye, G., Wang, G., Fang, L., Xiao, S., Fu, Z. & Peng, G. 2018. Dimerization of Coronavirus nsp9 with Diverse Modes Enhances Its Nucleic Acid Binding Affinity. *Journal of virology,* 92**,** e00692-18.

Zhang, L., Li, L., Yan, L., Ming, Z., Jia, Z., Lou, Z. & Rao, Z. 2018. Structural and Biochemical Characterization of Endoribonuclease Nsp15 Encoded by Middle East Respiratory Syndrome Coronavirus. *Journal of virology,* 92**,** e00893-18.

Zhang, Q., Shi, K. & Yoo, D. 2016. Suppression of type I interferon production by porcine epidemic diarrhea virus and degradation of CREB-binding protein by nsp1. *Virology,* 489**,** 252-68.

Zhao H, Lindqvist B, Garoff H, von Bonsdorff C, Liljeström P. 1994. A tyrosine-based motif in the cytoplasmic domain of the alphavirus envelope protein is essential for budding. *The EMBO Journal,* 13**,** 4204-4211.

Zhao, Q., Li, S., Xue, F., Zou, Y., Chen, C., Bartlam, M. & Rao, Z. 2008. Structure of the main protease from a global infectious human coronavirus, HCoV-HKU1. *Journal of virology,* 82**,** 8647-8655.

Zhao, Y., Chapman, D. A. & Jones, I. M. 2003. Improving baculovirus recombination. *Nucleic acids research,* 31**,** e6-e6.

Zhengchun, Lu., Jianqiang, Zhang., Chengjin M. Huang., Young, Yun, Go., Kay S. Faaberg., Raymond R.R. Rowland., Peter J. Timoney., Udeni B.R. Balasuriya. 2012. Chimeric viruses containing the N-terminal ectodomains of GP5 and M proteins of porcine reproductive and respiratory syndrome virus do not change the cellular tropism of equine arteritis virus. *Virology,* 432, 99-109.

Zhu, X., Wang, D., Zhou, J., Pan, T., Chen, J., Yang, Y., Lv, M., Ye, X., Peng, G., Fang, L. & Xiao, S. 2017. Porcine Deltacoronavirus nsp5 Antagonizes Type I Interferon Signaling by Cleaving STAT2. *Journal of virology,* 91**,** e00003-17.

Ziebuhr, J. 2005. The Coronavirus Replicase. *Current Topics in Microbiology and Immunology,* 287, 57-94.

Ziebuhr, J., Bayer, S., Cowley, J. A. & Gorbalenya, A. E. 2003. The 3C-like proteinase of an invertebrate nidovirus links coronavirus and potyvirus homologs. *Journal of virology,* 77**,** 1415-26.

Ziebuhr, J., Herold, J. & Siddell, S. G. 1995. Characterization of a human coronavirus (strain 229E) 3C-like proteinase activity. *Journal of virology,* 69, 4331-4338.

Ziebuhr, J., Heusipp, G. & Siddell, S. G. 1997. Biosynthesis, purification, and characterization of the human coronavirus 229E 3C-like proteinase. *Journal of virology,* 71, 3992-3997.

Ziebuhr, J., Snijder, E. J. & Gorbalenya, A. E. 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *Journal of General Virology,* 81**,** 853-879.

Zirkel, F., A. Kurth, P. L. Quan, T. Briese, H. Ellerbrok, G. Pauli, F. H. Leendertz, W. I. Lipkin, J. Ziebuhr, C. Drosten & S. Junglen. 2011. An Insect Nidovirus Emerging from a Primary Tropical Rainforest. *MBio,* 2**,** e00077-11.

Zirkel, F., H. Roth, A. Kurth, C. Drosten, J. Ziebuhr & S. Junglen. 2013. Identification and Characterization of Genetically Divergent Members of the Newly Established Family Mesoniviridae. *Journal of Virology,* 87, 6346-6358.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research,* 31**,** 3406-3415.

Appendix

Appendix 1: Publications

171

Appendix 2: Poster Presentations

- RNA structure and function in Middle Eastern Respiratory Syndrome Coronavirus. The Scientific Society for Saudi Students in the UK 2016. Presented poster.

- RNA structure and function signals in a novel Nidovirus sequence from the Sea Hare *Aplaysia californica*. Society of General Microbiology annual conference 2017. The Edinburgh International Conference Centre (EICC), Edinburgh. Presented poster.

- RNA structure and function signals in a novel Nidovirus sequence from the Sea Hare *Aplaysia californica*. EMBO conference in Protein synthesis and translational control 2017. Heidelberg – Germany. Presented poster.

Appendix 3: Synthetic DNA Fragments

DNA fragment name	Sequence	Length
Wildtype main protease	TTGCGAATGATTTTGTCTACCATTATACGTTTTTCCATTGTGAG AGTTCCTCTGGCCTATTTGTTGAGGTATACGTGGTATTTGTTTT GTATTCCTCCATGGAAGTACACTCATCTTGTTTCAGTGGAAGA GGCTTATTTGCACTTGCCACTTATTAGTTTGGATGATAATGCTA TGTTGAGTTATGGACATGTTTGTTCTACTTGTATTAACAAGGAT GAGAGTACTTTCTTTTGGGCAACTACAACGGGGGGCCGTAAG AATGAAGCAATGCTTGAATCTCAGCGTGGGAAAAGGGTGTTG CAGTCTGATGCAAATTTGTTTTCTGATGTGTTTGGTTCTCTGGT AACTGTGTCGTATGGCTCTAGTACAAGTGTTAACATGAATGGT GTGATTGTTTCTATGTTCACTATTGTTGTTCCACGTCATCTGTTT TTAACGTCTGGCAACACGCAGGCATATGCCGCTATGGAGCGT GACATTGCTGCCTTGGCTAGTAAGGGAGAGCTCTCTATTACTC TCAACAAGAAGACTGTTAAGGTTAGCAATGTAGTGTTTGTTGG TTGTTTGGCGTATATTACGTTTGACTCTCACGTTGGTCGCCCTC TTGGTTATGAAAAAATCTCTTTGGACAGTGCTTTTAAAAACCG CCCTGATGTACAAGTGGGTGCAGCAATTAATTACAGGAGTGG TGTCACACCTGTGGTTATTGGAAATAACGGGACTGTCTATGGA GCTTTAGAGGGTGGCGATTGTGGTACGCCGTTGTTTGTCTATG ATGGTCATGGCTGGAGATGGTTGGGTGTACACAACACTGGGA TTAATGCCAATGTTGTGGATGGGCAAACGTTTCCGGTCAATGT GTTCTTGGATATAACTACTGGTGTGAGTCAGCATGGAAAGATT GACCAATCGGATGTGAAAATTGGGCCTTTCAGGGTGCCTATTA TGGGTGCTATGGTGAAGGCGTTGATGAAAAAGGTTCTACCTG GTGAGTTCTTAAATGATGCCACTAAGATCTTCTTGTGTGATGA CACGTCGCTTCTCCTTCAGTACGATGGTCATGTTTTGGCCGAG GAATATAAGAGGGTTAGTGATATGACGTCCATAGGATTGAAC TTTTCTACGGCAGTTTTTGAGTCTATTGGTGGAAATGAAGGTA TTGTTCTTCACACTTCCGATGGTCCAGTTGTTCTTGATGAAAAG AAGGGTACGACTGCGTTGTTCAGATTTTTGTTGTGTTGTCTGT GT	1251 bp

DNA fragment name	Sequence	Length
Mutant main protease Q → A	GCGGCCGCACAGCTGTATACACGTGCAAGCCAGCCAGAACTCGCCCCGG AAGACCCCGAGGATCTCGAGTTGCGAATGATTTTGTCTACCATTATACGTT TTTCCATTGTGAGAGTTCCTCTGGCCTATTTGTTGAGGTATACGTGGTATTT GTTTTGTATTCCTCCATGGAAGTACACTCATCTTGTTTCAGTGGAAGAGGC TTATTTGCACTTGCCACTTATTAGTTTGGATGATAATGCTATGTTGAGTTAT GGACATGTTTGTTCTACTTGTATTAACAAGGATGAGAGTACTTTCTTTTGG GCAACTACAACGGGGGGCCGTAAGAATGAAGCAATGCTTGAATCTCAGC GTGGGAAAAGGGTGTTGCAGTCTGATGCAAATTTGTTTTCTGATGTGTTT GGTTCTCTGGTAACTGTGTCGTATGGCTCTAGTACAAGTGTTAACATGAAT GGTGTGATTGTTTCTATGTTCACTATTGTTGTTCCACGTCATCTGTTTTTAA CGTCTGGCAACACGCAGGCATATGCCGCTATGGAGCGTGACATTGCTGCC TTGGCTAGTAAGGGAGAGCTCTCTATTACTCTCAACAAGAAGACTGTTAA GGTTAGCAATGTAGTGTTTGTTGGTTGTTTGGCGTATATTACGTTTGACTC TCACGTTGGTCGCCCTCTTGGTTATGAAAAAATCTCTTTGGACAGTGCTTT TAAAAACCGCCCTGATGTACAAGTGGGTGCAGCAATTAATTACAGGAGTG GTGTCACACCTGTG	774 bp
Wildtype in frame stop codon TGA-C	GAGTCTTGTCGTGTGAAGTGTATACATAATAATGTTTTGGGTGACTGTTGT CTGAATTTGTTATGGCAGGGTCACACTGGCCACTTACATGAACATTTCGCG CGAGAGGGGAGTGATGAGTTGTTTTTGGCTTGTGTTGCAGGACTGGATAT TAGACATTTGGTGTTTGATAGGAAATGTTATTGTGATACTGTTTTGAAGGC ATGTAAGAATTGTGTTATTCGTACCTCTGTAGATAAGCATGTTCGTGCGGT GTATGCACGTCTTTTCCAAGCTTATGAGTCTGGTGCTATTACGAAAGTGGG TAAGTGTGATGAATGTTCTAATCCTACTGTGTTTGATAGATGTTTTGTGTG CGAACACGATAGGTATGCTAGTGAGGGAAAAATAGGGCCTAAGAGACCA CGGGCCACAGCGAATCCCTTTGATATAAAGTGACGGTGCGAAAACGAGTT CAGTTGCTCCTCGGTGACTGAAGTGCCTCAGATCCCAACAGCTAAAAACA GCAAGTGACGAAAGGTTTCGGAAAGTTTGAAACAACTCTCACTGAGGATG AAGTAGCTGGGCTCGTGGCGTGTCAGCTTTTTGATCAGGGAACTTTGTTT GACAGACGGATTAATCCT	624 bp

DNA fragment name	Sequence	Length
Mutant in frame stop codon TAA-A	GAGTCTTGTCGTGTGAAGTGTATACATAATAATGTTTTGGGTGACTGT TGTCTGAATTTGTTATGGCAGGGTCACACTGGCCACTTACATGAACAT TTCGCGCGAGAGGGGAGTGATGAGTTGTTTTTGGCTTGTGTTGCAGG ACTGGATATTAGACATTTGGTGTTTGATAGGAAATGTTATGTGATACT GTTTTGAAGGCATGTAAGAATTGTGTTATTCGTACCTCTGTAGATAAG CATGTTCGTGCGGTGTATGCACGTCTTTTCCAAGCTTTGAGTCTGGTG CTATTACGAAAGTGGGTAAGTGTGATGAATGTTCTAATCCTACTGTGT TTGATAGATGTTTTGTGTGCGAACACGATAGGTATGCTAGTGAGGGA AAAATAGGGCCTAAGAGACCACGGGCCACAGCGAATCCCTTTGATAT AAAGTAAAGGTGCGAAAACGAGTTCAGTTGCTCCTCGGTGACTGAAG TGCCTCAGATCCCAACAGCTAAAAACAAGCAAGTGACGAAAGGTTTC GGAAAGTTTGAAACAACTCTCACTGAGGATGAAGTAGCTGGGCTCGT GGCGTGTCAGCTTTTTGATCAGGGAACTTTGTTTGACAGACGGATTAA TCCT	624 bp
Wildtype structural protease	TCTGGTGTAAGGTCTTTTGCCCTTCGGTCTGACGATAATGAAGTTGTC TCTGGTGGTAGGTATGAACACGTGTCTGAAGACACACCTGATGTTCGT TTTCTTGTTGAGGGTGCTACGTATGAAAGCAATGTAAAATTTGCTGAT GTGGTTGGTTCCATTATTTCTTTTACTGATGGTTTGTGTTCTGGTGTTC TTGTGGGTGAAAATCTTGTCTTGACTGCTGCGCATTGTTTGGTTCAGC CGGCTCAGGTGCTACGTTTCTGTTTGGGAGATGACTGTTATCAAATCC TTTCTTATGAGGTTGTTGACAAAGAGTTTACTAAAATAGGGGGGAAT GATCTTGGTCTTGCTGTGATTGATGCTAAGTTGGACAATTTTGTTAGA GTTTTGCCTACGAATGCTGCTATGTCTTCTTGTTACTTTTTTGGTTATG GTGGTCCAACTGATTCTGGGCTGAGTCGTACGCTTGGTTCTTTTGCTG TTACACCTTTTATGTGTTCGTATGAAGGGGAGGTTCTTGATAATCTTGT TTGTAGTTCGTATGAAACTGATGGCGCTCGTTTGTGTGGAGGTGACTC TGGTGGCCCATTGATTTGTGATAGTCCTTGGGGTCTTGGTGTTTATGC GGTTGCGAGTGCTGGCGGAGGTTGTTCTGGTGGTTCTACCATTTCTAT GTATGTAGATGTTCGTCGTGATTATCCATTGTTGCGTGCGCAAGTGTG TGATAGTTTGACTGTGCCTAGGTCTTTACATTTGGCTTCAACGGAGAT GGATGTTTTAAACATAACTGAGGCGGATTCCCTTTTTACGAATGCTAC	1266 bp

	TCCTGTTAAGTCTAACTTTTCTTGTGCATCTGTGTCTCCTGAGCTTTTGG GTCCACTTGTTTCTGGTCATGCTTCTGTGTACACTGTAGATTCACGTTA TGTTGAAGGTTGTGTTAAACCTTCACCTTTGTCTGCGTATTCGGAAAAT GTTGAAATGGCTGATAATTATTTTTTAGTTTCTAGTCTTGGTTCTAATC CTAAGGCGAAGGGTGGTGATGATTTTGTTTTGTTGATTGGGCCAAAT ACTGAGTTTGTTTGTCCTTGTTATTACTTGTATCAATCTTCGGTTTACCG TTATGATGTGGGTAAGTACGTTTCTTCAGTTAAAGTGGGATCACGTAC AAAACAGTTAAACACTACTTCTGGTTATTGGCAGGAGTATATGTTGAG GGGGAAGTCAGTTAAGTATCAAGCTACACTGGGGCATGTTCTGTATG CGAACAAACCG	
Structural protease first catalytic residue mutation H → A	GTATACACGTGCAAGCCAGCCAGAACTCGCCCCGGAAGACCCCGAGG ATCTCGAGTCTGGTGTAAGGTCTTTTGCCCTTCGGTCTGACGATAATG AAGTTGTCTCTGGTGGTAGGTATGAACACGTGTCTGAAGACACACCT GATGTTCGTTTTCTTGTTGAGGGTGCTACGTATGAAAGCAATGTAAAA TTTGCTGATGTGGTTGGTTCCATTATTTCTTTTACTGATGGTTTGTGTTC TGGTGTTCTTGTGGGTGAAAATCTTGTCTTGACTGCTGCGGCTTGTTT GGTTCAGCCGGCTCAGGTGCTACGTTTCTGTTTGGGAGATGACTGTTA TCAAATCCTTTCTTATGAGGTTGTTGACAAAGAGTTTACTAAAATAGG GGGGAATGATCTTGGTCTTGCTGTGATTGATGCTAAGTTGGACAATTT TGTTAGAGTTTTGCCTACGAATGCTGCTATGTCTTCTTGTTACTTTTTTG GTTATGGTGGTCCAACTGATTCTGGGCTGAGTCGTACG	520 bp

Structural protease second catalytic residue mutation D → A	GTATACACGTGCAAGCCAGCCAGAACTCGCCCCGGAAGACCCCGAGG ATCTCGAGTCTGGTGTAAGGTCTTTTGCCCTTCGGTCTGACGATAATG AAGTTGTCTCTGGTGGTAGGTATGAACACGTGTCTGAAGACACACCT GATGTTCGTTTTCTTGTTGAGGGTGCTACGTATGAAAGCAATGTAAAA TTTGCTGATGTGGTTGGTTCCATTATTTCTTTTACTGATGGTTTGTGTTC TGGTGTTCTTGTGGGTGAAAATCTTGTCTTGACTGCTGCGCATTGTTT GGTTCAGCCGGCTCAGGTGCTACGTTTCTGTTTGGGAGATGACTGTTA TCAAATCCTTTCTTATGAGGTTGTTGACAAAGAGTTTACTAAAATAGG GGGGAATGCTCTTGGTCTTGCTGTGATTGATGCTAAGTTGGACAATTT TGTTAGAGTTTTGCCTACGAATGCTGCTATGTCTTCTTGTTACTTTTTTG GTTATGGTGGTCCAACTGATTCTGGGCTGAGTCGTACG	520 bp
Structural protease all catalytic residues mutation H, D and S → A	GTATACACGTGCAAGCCAGCCAGAACTCGCCCCGGAAGACCCCGAGG ATCTCGAGTCTGGTGTAAGGTCTTTTGCCCTTCGGTCTGACGATAATG AAGTTGTCTCTGGTGGTAGGTATGAACACGTGTCTGAAGACACACCT GATGTTCGTTTTCTTGTTGAGGGTGCTACGTATGAAAGCAATGTAAAA TTTGCTGATGTGGTTGGTTCCATTATTTCTTTTACTGATGGTTTGTGTTC TGGTGTTCTTGTGGGTGAAAATCTTGTCTTGACTGCTGCGGCTTGTTT GGTTCAGCCGGCTCAGGTGCTACGTTTCTGTTTGGGAGATGACTGTTA TCAAATCCTTTCTTATGAGGTTGTTGACAAAGAGTTTACTAAAATAGG GGGGAATGCTCTTGGTCTTGCTGTGATTGATGCTAAGTTGGACAATTT TGTTAGAGTTTTGCCTACGAATGCTGCTATGTCTTCTTGTTACTTTTTTG GTTATGGTGGTCCAACTGATTCTGGGCTGAGTCGTACGCTTGGTTCTT TTGCTGTTACACCTTTTATGTGTTCGTATGAAGGGGAGGTTCTTGATA ATCTTGTTTGTAGTTCGTATGAAACTGATGGCGCTCGTTTGTGTGGAG GTGACGCTGGTGGCCCATTGATTTGTGATAGTCCTTGGGGTCTTGGTG TTTATGCGGTTGCGAGTGCTGGCGGAGGTTGTTCTGGTGGTTCTACCA TTTCTATGTATGTAGATGTTCGTCGTGATTATCCATTGTTGCGTGCGCA AGTGTGTGATAGTTTGACTGTGCCTAGG	799 bp