The analysis of a conserved RNA structure in the 3D polymerase encoding region of *Human Parechovirus* 1

Cheryl K. Eno-Ibanga

A thesis submitted for the degree of Doctorate of Philosophy
in Microbiology

School of Biological Sciences
University of Essex
October 2016

Dedication

I would like to dedicate this thesis to all scientists who shall find this study interesting and also those who may wish to carry out further research on it.

Table of Contents

Acknowledgments.....viii

Abstract.....x

Abbreviations.....xii

List of Figures.....xvii

List of Tables.....xxi

Chapter 1.....1

General introduction.....1

1.1 Introduction to viruses.....2

1.1.1 What is a virus?.....2

1.1.2 Origin, history and uses of viruses.....2

1.2 Classification and nomenclature of viruses.....6

1.2.1 General classification.....6

1.2.2 Baltimore's classification.....7

1.2.3 Sequence based classification.....9

1.2.4 ICTV classification.....9

1.3 Picornaviruses.....12

1.3.1 Introduction.....12

1.3.2 Medical significance of picornaviruses.....16

1.3.2.1 Parechovirus.....21

1.3.3 Picornavirus replication.....25

1.3.4 Picornavirus genome organization and expression.....28

1.3.4.1 Leader protein (L^{pro}).....30

1.3.4.2 Capsid proteins.....30

1.3.4.3 P2 and P3 proteins and their functions.....33

1.3.5 Picornavirus receptors.....35

1.4 Determination and Prediction of RNA structures.....38

1.4.1 Determination of RNA structures.....38

1.4.1.1 Chemical and enzymatic reactivity.....38

1.4.1.2 Physical method.....39

1.4.2 Prediction of RNA structures.....41

1.4.2.1 Thermodynamic method 41
1.4.2.2 Phylogenetic analysis and mutational analysis 42
1.5 General Overview of RNA Structures in Viral Genomes 45
1.5.1 Phages 45
1.5.2 Negative stranded RNA Viruses 46
1.5.3 Positive Strand RNA Viruses 46
1.5.4 *Retroviridae* 47
1.6 RNA structures in *Picornaviridae* 48
1.6.1 Non-coding regions 48
1.6.1.1 3' UTR 48
1.6.1.2 5' UTR 49
1.6.2 Cis-acting replication element (cre) 52
1.6.3 Other conserved elements 53
1.7 Aims of present study 56
Chapter 2 57
Materials and Methods 57
2.1 Materials 58
2.1.1 Agarose gel electrophoresis 58
2.1.2 Bacterial cell culture media 60
2.1.3 Bacterial strain 61
2.1.4 Cloning vectors 61
2.1.5 Ligation reagent 61
2.1.6 *In vitro* transcription 61
2.1.7 Mammalian cell culture 62
2.1.8 Mammalian cell over-lay 64
2.1.9 Mammalian cell preparation for bioimaging 64
2.1.10 Oligonucleotides 65
2.1.11 Plasmid DNA preparation 65
2.1.12 PCR reagents 65
2.1.13 Reagents for determining cell's stress response 71
2.1.14 Restriction Enzymes 71
2.1.15 Sequencing 71

2.1.16 Transfection reagents 71
2.1.17 Viruses 72
2.1.18 Virus cDNA 72
2.1.19 Luciferase assay: 74
2.1.20 Cell viability assay: 74
2.1.21 General purpose materials 74
2.2 Methods 76
2.2.1 Cell Culture 76
2.2.1.1 Cell propagation 76
2.2.1.2 Cell storage by freezing 76
2.2.2 Cell biology techniques 77
2.2.2.1 Transfection 77
2.2.2.2 Fixation 78
2.2.2.3 Immunofluorescence staining 79
2.2.2.4 Flow cytometry analysis with the Fluorescence Activated Cell Sorter (FACS) 79
2.2.3 Annealing complementary oligonucleotides 79
2.2.4 PCR 82
2.2.4.1 PCR mutagenesis 82
2.2.4.2 Colony PCR 82
2.2.4.3 Overlap PCR 83
2.2.5 Cloning DNA fragments into plasmid vectors 84
2.2.5.1 Digestion with restriction enzymes 84
2.2.5.2 Linearization of pEGFP-C1 and pmCherry-C1 84
2.2.5.3 Ligation of annealed oligonucleotides/DNA fragments to cut vector 84
2.2.5.4 Ligation to TOPO Vector 85
2.2.5.5 Recombinant plasmid construction from primary vector (pGEM-T Easy) 85
2.2.6 Construction of mutant plasmids 86
2.2.7 Transformation process 86
2.2.8 Nucleic acid extraction and purification 87
2.2.8.1 Small-scale DNA extraction 87
2.2.8.2 Large-scale DNA extraction 87
2.2.8.3 DNA isolation from agarose gel 87

2.2.9 Agarose gel electrophoresis 88
2.2.10 Plaque assay 88
2.2.11 Crystal violet growth assay/Wst 1 assay 89
2.2.12 *In vitro* transcription 90
2.2.13 Luciferase assay 90
2.2.14 Computer analysis of DNA sequences 91
2.2.15 Highlighting of sequences 92
Chapter 3 93
Examining the role of an RNA structure in the $3D^{pol}$-encoding region of HPeV1 on RNA stability 93
3.1 Introduction 94
3.2 Investigation of the 3D conserved structure 96
3.2.1 Confirmation of the structure 96
3.2.2 A promiscuous RNA structure in the 3'UTR of several picornaviruses 112
3.3 Constructs to investigate the role of the RNA structure in the $3D^{pol}$-encoding region 114
3.3.1 Fluorescence microscopy 117
3.3.1.1 Co-transfection of pEGFP3DS and pEGFPeC with pmCherry-C1 120
3.3.1.2 Co-transfection of pMCh3DS and pMCheC with pEGFP-C1 122
3.3.1.3 Co-transfection of pEGFP3DS with pMCheC and pEGFPeC with pMCh3DS 124
3.3.2 IRES driven translation experiment 126
3.3.3 Analysis of different HPeV1 regions compared to the 3DS conserved structure 131
3.3.4 The highly conserved structure in the $3D^{pol}$ might affect viral RNA stability 133
3.4 Investigation of fluorescence using FACS analysis 136
3.4.1 Investigation via comparison of DNA construct sizes 142
3.4.2 Investigation of the 'CG' dinucleotide in pEGFP3DS 147
3.4.3 Mutant plasmid construction 150
3.4.3.1 The highly conserved structure in the $3D^{pol}$ appears to affect viral RNA stability 150
3.5 Discussion 155
3.5.1 Investigation of the role of the 3DS structure via microscopy 164
3.5.2 Investigation of the 3DS structure using IRES driven translation 165
3.5.3 Further investigation of the role of the 3DS structure via microscopy and FACS 166
3.5.4 The relationship between insert size and EGFP fluorescence produced 167
3.5.5 Mutation of the 3DS structure 168

3.5.6 Other RNA structures 169
Chapter 4 172
Investigation of the effect of RNA secondary structures on RNA stability in response to cell stress 172
4.1 Introduction 173
4.2 Analysis of optimum Actinomycin-D concentration to prevent transcription 175
4.2.1 Regulation of the expression of transfected constructs by Actinomycin-D 179
4.2.2 Normalization to individual (pEGFP-C1 + pmCherry-C1) for each experiment 179
4.2.3 Normalization to only untreated (pEGFP-C1 + pmCherry-C1) 180
4.3 Investigation of stress factors 183
4.3.1 Analysis of NaCl concentrations for osmotic stress 184
4.3.1.1 Normalization to individual (pEGFP-C1 + pmCherry-C1) for each experiment 186
4.3.1.2 Normalization to only unstressed (pEGFP-C1 + pmCherry-C1) 186
4.3.2 Analysis of H_2O_2 concentrations for oxidative stress 192
4.3.2.1 Normalization to (pEGFP-C1 + pmCherry-C1) for each experiment 194
4.3.2.2 Normalization to only unstressed (pEGFP-C1 + pmCherry-C1) 194
4.3.3 Analysis of Thapsigargin concentrations for ER stress 200
4.3.3.1 Normalization to (pEGFP-C1 + pmCherry-C1) for each experiment 202
4.3.3.2 Normalization to only untreated (pEGFP-C1 + pmCherry-C1) 202
4.4 The effect of secondary structures on RNA stability in cells responding to stress due to infection 208
4.4.1 Quantification of HPeV1 used for Infection of cells transfected with DNA constructs 208
4.4.2 Quantification of CAV9 used for Infection of cells transfected with DNAs 208
4.4.2.1 Normalization to (pEGFP-C1 + pmCherry-C1) for each experiment 209
4.4.2.2 Normalization to only non-infected (pEGFP-C1 + pmCherry-C1) 209
4.5 Discussion 215
4.5.1 The effect of secondary structures on RNA stability in cells responding to stress 216
4.5.2 Infection with HPeV1 218
4.5.3 Infection with CAV9 219
Chapter 5 220
Assembly of a the *Renilla* Luciferase reporter construct to analyse RNA structures in HPeV1 220
5.1 Introduction 221
5.2 The generation of an HPeV1 replicon using a luciferase reporter 222

5.2.1 Backbone Preparation (pHPeV1Luc) 222
5.2.2 PCR 223
5.2.3 Final ligation 227
5.3 Use of the construct to study the putative HPeV cre 229
5.3.1 Production of pHPeV1LucLoopcreMut and pHPeV1LucRScreMut 229
5.3.2 Analysis of the cre 232
5.3.3 Transcription 232
5.3.4 Analysis of the potential cre 233
5.4 The 3DS RNA secondary structure in RNA replication 237
5.4.1 Production of pHPeV1(21mut)Luc 237
5.4.2 pHPeV1(21mut) 237
5.4.3 pHPeV1YGAA 242
5.5 Discussion 245
Chapter 6 249
General discussion and future work 249
6.1 General discussion and future work 250
6.1.1 General discussion 250
6.1.2 Future work 253
Chapter 7 254
References 254
Appendices 272

Acknowledgement

I thank God Almighty for his grace, guidance and preservation in the cause of this research and also for His help to complete this thesis. It is indeed in God that I live, move and have my being.

Many thanks to my supervisor, Prof. Glyn Stanway for his unending patience, professional guidance and unrelenting support in the cause of this research. It has indeed been a privilege to work closely with him all these years.

I also thank Dr. Julie Lloyd for the thorough scrutiny of my work for my board meetings and all that I have learnt from her. I am grateful to Dr. Andrea Mohr and Dr. Ralf Zwacka for their assistance with the flow cytometry experiments and fruitful research advice. I thank Dr. Greg Brooke for his unreserved help with the two different assays used to measure relative cell death in this research and also for his assistance with experiments involving the illuminometer. Many thanks to Mrs. Emma Revill for creating a relaxed and friendly study environment for me, and for her professional help and guidance where necessary.

I am grateful to “the 5.16 family”, Dr. Shaia Almalki, Dr. Arsalan Salimi, Dr. Naseebh Baeshen, Dr. Ashjan Shami, Miss Marina Ioannou, Osamah Hasan and my very dear colleague Mr. Ali Khrid, for their unreserved assistance in my Laboratory experiments. Also thank you to Mrs. Natalie Gray and Mrs Hannah Dundon for making available all the needed reagents used to carry out this research.

Thanks to members of Laboratories, 5.08, 5.36 and 5.32, Biological Sciences, University of Essex for letting me use the equipment in their Laboratory.

Thank you to my very wonderful friend Marina Ioannou for being so dear to me, looking after me and also who’s kind efforts have helped me to organise this thesis. I thank Christos Kaniklis for accommodating me and dropping me off from the University especially in the late cold nights. I am indebted to all my special friends, Chukuka Ochonogor, Irabonosi Obomighie, Onyeka Asuzu, Ndifreke Ette, Daniel Silas-Olu, Olayinka Uadiale, Joe & Semira Daramola, Kelechi Anumnu, LCpl Daniel Ojo and

Alexandra Uma, none of whom may ever read this thesis, however, their friendship, support and care kept me sane during the cause of this Ph.D.

I wish to thank the Tertiary Education Trust (TET) Fund, Nigeria, which provided full funding for this work.

Lastly an extra special thank you to my beloved and inspiring parents and brothers for their all-round support, encouragement and prayers. They have been my rock and motivation thus far. I am indeed very grateful to God that they are all alive to witness the completion of this thesis.

Abstract

Picornaviruses are important causes of human illness and it is necessary to understand more about how these viruses function. Human parechoviruses (HPeV) are common pathogens and studies have shown that 95% of people become infected with HPeV at a very early age, usually with symptoms such as mild diarrhoea and fever. However, one virus type HPeV3, is implicated in much more serious cases of neonatal disease and so it is important to understand HPeVs to increase the opportunity to develop drugs or vaccines against the infection. The HPeV1 genome encodes a single polyprotein that is cleaved into structural and non-structural proteins. Analysis of one region of the genome (encoding the polymerase, $3D^{pol}$) shows that some codons are perfectly conserved, suggesting functions in addition to protein coding. This region seems to fold into an RNA secondary structure made up of three stem-loops and a tertiary structure “kissing” interaction. The structure was validated by comparing all the available HPeV sequences and found to be highly conserved. To investigate if the structure has a role in RNA stability, an EGFP fluorescent assay was used. Sequences containing the structure was added to the 3’ UTR of the EGFP gene. A mutant with 21 mutations which completely destroys the RNA structure was also used. A FACS-based method was used to measure expression levels of EGFP. The results showed that there was a significant reduction in fluorescence from the mutant construct. The effect of the structure was also investigated in infected cells and in cells exposed to different stresses which could mimic virus infection. The results suggest that the structure can positively affect RNA stability/translation. Further investigation on other possible roles such as RNA replication and translation should be performed to improve the

understanding of the biology of the structure in HPeVs and a Renilla Luciferase reporter gene system was assembled to facilitate the studies in the future.

Abbreviations

A	Adenine (nb)
AEV	*Avian encephalomyelitis virus*
AiV	*Aichivirus*
ASV	*Avian Sapelovirus*
AsV	*Avisivirus*
ATP	Adenosine triphosphate
AqV	*Aquamavirus*
B.C.	Before Christ
BCV	*Boone Cardiovirus*
BGPV	*Bluegill Picornavirus*
bp	Base pair
BPV	*Bat Picornavirus*
BRAV	*Bovine rhinitis A virus*
BRBV	*Bovine rhinitis B virus*
C	Cytosine (nb)
$CaCl_2$	Calcium chloride (Chemical formula)
CanPV	*Canine picornavirus*
CaPdV	*Cadicivirus*
CAR	Coxsackie-adenovirus receptor
CD	Cluster of differenciation
cDNA	Complementary DNA
CHO	Chinese Hamster Ovary
CMC	Carboxymethyl cellulose

CNS	Central Nervous System
Conc.	Concentration
CoV	*Cosavirus*
Cre	Cis-acting replication element
CV	Crystal violet
CVA	Coxsackie A viruses
CVB	Coxsackie B viruses
DAF	Decay Accelerating Factor
dH_2O	Distilled H_2O
DHAV	*Duck hepatitis A virus*
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic Acid
dsDNA	Double strand DNA
dsRNA	Double strand RNA
E. coli	*Escherichia coli*
EBOV	*Ebolavirus*
EDTA	Ethylene diamine tetra-acetate
EMCV	*Encephalomyocarditis virus*
ER	Endoplasmic reticulum
ERAV	*Equine rhinitis A virus*
ERBV	*Equine rhinitis B virus*
EV	*Enterovirus*
FBS	Fetal Bovine Serum
FePV	*Feline picornavirus*
FMDV	*Foot and mouth disease virus*

g	gram
G	Guanine (nb)
GMK	Green Monkey Kidney
GpV	*Greplavirus*
GV	*Gallivirus*
H_2O	Water (Chemical formula)
HAV	*Hepatitis A virus*
HAVcr	Hepatitis A virus cellular receptor
HIV	*Human Immunodeficiency Virus*
HPeV	*Human parechovirus*
hr	Hour
HuV	*Hungarovirus*
ICAM	Intercellular adhesion molecule
ICTV	International Committee on Taxonomy of Viruses
IgSF	Immunoglobulin superfamily
IiPV	*Ia io picornavirus*
IRES	Internal ribosome entry site
kb	Kilo base
l/L	Litre
LB	Luria broth
LDL	Low-density lipoprotein
LV	*Ljungan virus*
L^{Pro}	Leader protein
m	milli/metre
M	Molar

MCF-7	Michigan Cancer Foundation 7
MDA-MB-231	M.D. Anderson - metastatic breast 231
MeV	*Melegrivirus*
min	Minute
MiV	*Miniopterus schreibersii picornavirus*
MoV	*Murine mosavirus*
mRNA	Messenger RNA
n	nano
nt	Nucleotide
ORF	Open reading frame
OsV	*Oscivirus*
P	Polyprotein
PaV	*Pasivirus*
PasV	*Passerivirus*
PBS	Phosphate-Buffered Saline
PC3	Prostate cancer 3
PCR	Polymerase Chain Reaction
PiPV	*Pigeon picornavirus*
Poly(C)	Polyribocytidylic acid
PSV	*Porcine sapelovirus*
PTV	*Porcine teschovirus*
PV	*Poliovirus*
QPV	*Quail picornavirus*
RGD	Arginine-glycine-aspartic acid
RNA	Ribonucleic acid

+RNA	positive sense RNA
-RNA	negative sense RNA
RNase	Ribonuclease
RNP	Ribonucleoprotein
RoV	*Murine rosavirus*
rpm	Rounds per minute
rRNA	ribosomal RNA
RsetRNAV01	*Rhinozosolenia setigera RNA virus 01*
RT	Room Temperature
RV	*Rhinovirus*
SARS	Severe Acute Respiratory Syndrome
SaV	*Salivivirus*
SEBV	*Sebokele virus*
SVV	*Seneca valley virus*
ssDNA	Single strand DNA
ssRNA	Single strand RNA
SSV	*Simian sapelovirus*
T	Thymine (nb)
TMEV	*Theiler's murine encephalomyelitis virus*
U	Uracil (nb)
UTR	Untranslated region
VP	Viral protein
VPg	Viral protein genome
WT	Wild type

List of Figures

Figure 1.1: Very early recording of viral infection **5**

Figure 1.2: The Baltimore system of classification of viruses into 7 different types **8**

Figure 1.3: Rooted phylogenetic tree of picornavirus species..... **14**

Figure 1.4: The life cycle of a picornavirus..... **27**

Figure 1.5: Picornavirus genome showing the 12 proteins derived from a typical polyprotein and their functions..... **29**

Figure 1.6: Schematic representation of the icosahedral capsid..... **32**

Figure 1.7: Schematic representation of picornaviruses receptors..... **37**

Figure 1.8: Alignment of the 3D structure in 10 completely sequenced strains of HPeVs..... **44**

Figure 1.9: Schematic diagram of the 5'UTR structures of picornaviruses with different IRES types, I-V. .. **51**

Figure 1.10: Predicted RNA structure in the 3D encoding region of HPeV1 Harris strain **55**

Figure 2.1: 1 Kb DNA ladder **59**

Figure 3.1: Schematic representation of the predicted RNA secondary structure in the 3D encoding region of HPeV1..... **98**

Figure 3.2: Alignment of the nucleotide sequences of 254 HPeV isolates in the region of the predicted structure..... **110**

Figure 3.3: Alignment of the sequences of HPeV1 Harris strain and Sebokele virus in the region of the predicted HPeV RNA structure..... **111**

Figure 3.4: A conserved RNA structure found in several diverse picornaviruses **113**

Figure 3.5: Schematic representation of constructs pEGFP3DS and pMCh3DS **115**

Figure 3.6: MCF-7 cells transfected with pEGFP3DS, pEGFPeC and pEGFP-C1 ..... **118**

Figure 3.7: GMK cells transfected with pEGFP3DS, pEGFPeC and pEGFP-C1 ..... **119**

Figure 3.8: Co-transfection of pmCherry-C1 with pEGFP3DS, pEGFPeC and pEGFP-C1 into GMK cells…. . **121**

Figure 3.9: Co-transfection of pEGFP-C1 with pMCh3DS and pMCheC into GMK cells **123**

Figure 3.10: Co- transfection of pEGFP3DS with pMCheC, pEGFPeC with pMCh3DS and pEGFP-C1 with pmCherry-C1 into GMK cells **125**

Figure 3.11: Schematic diagram to show the structure of RNA transcribed *in vivo* from constructs ligated to a TOPO GFP vector **127**

Figure 3.12: Theory behind the constructs made to analyse whether the 3DS structure enhances IRES-driven translation **128**

Figure 3.13: GMK cells were transfected with T5dsRed, T5dsRed3DS and T5dsRed3D using Lipofectin.. **130**

Figure 3.14: Co-transfection of pmCherry-C1 with pEGFP-C1, pEGFP3UTR, pEGFPeC, pEGFPS, pEGFPVP1, pEGFPN3 and pEGFP3DS respectively into GMK cells **135**

Figure 3.15: Flow cytometry analysis of constructs after co-transfection of pmCherry-C1 with pEGFP-C1, pEGFP3UTR, pEGFPeC, pEGFPS, pEGFPVP1, pEGFPN3 and pEGFP3DS respectively into GMK cells **140**

Figure 3.16: Percentage fluorescence given by constructs gated at 10^4 using Fluorescence Activated Cell Sorter (FACS) **141**

Figure 3.17: Mean fluorescence of EGFP constructs gated on mCherry fluorescence **145**

Figure 3.18: Scatter plot of the fluorescence expressed by each construct against size of insert **146**

Figure 3.19: Mean fluorescence of the structure in the 3D region (pEGFP3DS) in comparison to the CG dinucleotide mutations **149**

Figure 3.20A: Schematic representation of the predicted RNA secondary structure in the 3D encoding region of HPeV1 ..... **152**

Figure 3.20B: 21 mutations introduced into the structure thereby disrupting the three stem-loops **152**

Figure 3.21: Mean fluorescence of the structure in the 3D region **154**

Figure 3.22: Analysis of the structure predicted to be present in the HPeV 3D-encoding region (Witwer *et al*., 2001) which overlaps with the Williams *et al*. (2009) structure **157**

Figure 3.23: Alignment of the amino acid sequences of the 3D protein C-terminus from PV1 Mahoney and HPeV1 Harris **160**

Figure 3.24: Schematic representation of pEGFP-C1 with all other constructs analyzed indicating the length of the inserted sequence **163**

Figure 4.1: Concentrations of Actinomycin-D analysed using crystal violet assay **177**

Figure 4.2: Actinomycin-D result normalized to its individual pEGFP-C1 + pmCherry-C1 control **181**

Figure 4.3: Actinomycin-D result normalized to only unstressed pEGFP-C1 + pmCherry-C1 control **182**

Figure 4.4: GMK cells treated with different concentrations of NaCl for 6 hr.....**185**

Figure 4.5: 140 mM NaCl result normalized to its individual pEGFP-C1 + pmCherry-C1 control **188**

Figure 4.6: 140 mM NaCl result normalized to only unstresssed pEGFP-C1 + pmCherry-C1 control **189**

Figure 4.7: 400 mM NaCl result normalized to its individual pEGFP-C1 + pmCherry-C1 control **190**

Figure 4.8: 400 mM NaCl result normalized to only unstressed pEGFP-C1 + pmCherry-C1 control **191**

Figure 4.9: GMK cells treated with different concentrations of H_2O_2 for 6 hr **193**

Figure 4.10: 50 µM H_2O_2 result normalized to its individual pEGFP-C1 + pmCherry-C1 control.....**196**

Figure 4.11: 50 µM H_2O_2 result normalized to only unstressed pEGFP-C1 + pmCherry-C1 control.....**197**

Figure 4.12: 500 µM H_2O_2 result normalized to its individual pEGFP-C1 + pmCherry-C1 control.....
.....**198**

Figure 4.13: 500 µM H_2O_2 result normalized to only unstressed pEGFP-C1 + pmCherry-C1 control **199**

Figure4.14: GMK cells treated with different concentrations of Thapsigargin for 6 hr **201**

Figure4.15: 0.3 µM Thapsigargin result normalized to its individual pEGFP-C1 + pmCherry-C1 control…**204**

Figure 4.16: 0.3 µM Thapsigargin result normalized to only unstressed pEGFP-C1 + pmCherry-C1 control
.....**205**

Figure 4.17: 1 µM Thapsigargin result normalized to its individual pEGFP-C1 + pmCherry-C1 control.....**206**

Figure4.18: 1 µM Thapsigargin result normalized to only unstressed pEGFP-C1 + pmCherry-C1 .....**207**

Figure4.19: HPeV1 infection result normalized to its individual pEGFP-C1 + pmCherry-C1 control.....**211**

Figure4.20: HPeV1 infection result normalized to only non-infected pEGFP-C1 + pmCherry-C1 control..**212**

Figure4.21: CAV9 infection result normalized to its individual pEGFP-C1 + pmCherry-C1 control.....**213**

Figure4.22: CAV9 infection result normalized to only non-infected pEGFP-C1 + pmCherry-C1 control....**214**

Figure 5.1: Schematic representation of how the construct pHPeV1Luc was produced from existing
pHPeV1.....**224**

Figure 5.2: Uncut pHPeV1 and single digestion of pHPeV1 with EcoRI, XhoI, MluI, AflII and SacI.....**225**

Figure 5.3: Gel picture showing A) two PCR products; SacI.5UTR.Luc and 5UTR.Luc.NheI respectively, purified and B) two joined PCR product samples, SacI.5UTR + Luc.NheI **226**

Figure 5.4A: Schematic representation of the putative HPeV1 cre structure showing the target areas for mutation **230**

Figure 5.4B: Folding of the mutated loop with mutated bases highlighted **230**

Figure 5.4C: Folding of the mutated right stem with mutated bases highlighted **230**

Figure 5.5: Schematic methodology diagram to illustrate the production and analysis of the mutated HPeV1 cre using *Renilla* Luciferase as a reporter **234**

Figure 5.6: *In vitro* RNA transcription of the DNA samples **235**

Figure 5.7: Mean luminescence of constructs analysed **236**

Figure 5.8: Schematic diagram showing primer target sites in the pHPeV1 cDNA and the pEGFP(21mut)3DS construct which was used to produce pHPeV1(21mut **239**

Figure 5.9A: Schematic representation of the initial product pHPeV1(21mut **241**

Figure 5.9B: Schematic representation of the final product; pHPeV1(21mut)Luc **241**

Figure 5.10A: Schematic representation of the initial product pHPeV1(YGAA) **244**

Figure 5.10B: Schematic representation of the final product pHPeV1(YGAA)Luc **244**

List of Tables

Table 1.1: ICTV classification of viruses (2015 release) **11**

Table 1.2: Names of species and genus of the abbreviations used to construct the phylogenetic tree.. **....15**

Table 1.3: Human Parechovirus types **24**

Table 2.1. Names, origins and suppliers of the cell lines used for the experiments in this report including the summary of methods of maintenance of the cells **63**

Table 2.2: Oligonucleotides used to generate DNA fragments that were ligated to Topo vector **66**

Table 2.3: Oligonucleotides used to generate and analyse DNA fragments that were ligated to the C-terminus of cut pEGFP-C1 vector **67**

Table 2.4: Oligonucleotides used to generate and analyse DNA fragments that were used in the production of the Luciferase replicon **69**

Table 2.5: Oligonucleotides used to generate DNA fragments containing mutations in the Luciferase replicon **70**

Table 2.6: Restriction enzymes used **73**

Table 2.7: Characteristics of each oligonucleotide used for annealing **81**

Table 2.8: Summary of how overlap PCR was carried out **83**

Table 3.1: Description of each construct used for analysis, size and origin of PCR fragment ligated to cut pEGFP-C1 and pmCherry vectors **116**

Table 3.2: Description of each construct used in the expanded set to study if the 3DS structure affects RNA stability **132**

Table 3.3: List of constructs made to study the effect of size of insert on EGFP fluorescence in the assay **144**

Table 3.4: Summary of how overlap PCR was carried out to produce pEGFP(CG1)3DS, pEGFP(CG2)3DS and pEGFP(CG1+2)3DS **148**

Table 3.5: Summary of the 21 silent mutations introduced into the 3DS structure **153**

Table 4.1: Concentrations of Actinomycin-D analysed after 24 hr of treatment to GMK cells by observation by eye and light microscope **176**

Table 4.2: Name, insert size ligated to cut pEGFP-C1 and origin of each construct used for analysis.....**178**

Table 5.1: Plasmids used to give the fragments needed to make pHPeV1Luc and the restriction enzymes used to cut each **228**

Table 5.2: Summary of how overlap PCR was carried out to produce pHPeV1Lucloop*cre*Mut and pHPeV1LucRS*cre*Mut.....**231**

Table 5.3: Summary of how overlap PCR was carried out to produce the final product, pHPeV1(21mut) **240**

Table 5.4: Summary of how overlap PCR was carried out to produce the final product, pHPeV1(YGAA) .**243**

Chapter 1

General introduction

1.1 Introduction to viruses

1.1.1 What is a virus?

A virus is commonly defined as an extremely small, subcellular parasite of cells, composed of either a DNA or RNA strand enclosed in a protein coat (van Etten, 2011). It sometimes possesses a lipid envelope which it acquires during its exit from the cell membrane of its host. This is however specific to different viruses. Viruses cause diseases in animals, plants and even microorganisms such as bacteria. Viruses that are pathogenic to bacteria are generally known as bacteriophages or simply called phages (Carter & Saunders, 2007).

All viruses enter into the host cell and replicate themselves within the host using the cell mechanisms and resources, as they do not have their own source of energy. They leave the cell after it is severely damaged (Van Regenmortel & Mahy, 2004). Viruses are ubiquitous in nature and more are being identified through approaches such as deep sequencing. Knowledge of many of these vast numbers of viruses is limited (Breitbart & Rohwer, 2005; Carter & Saunders, 2007).

1.1.2 Origin, history and uses of viruses

Viruses do not have fossils, so it is difficult to determine their exact origin. However, three theories have been postulated to explain their origin. **The regressive theory** proposes that viruses originated from free living bacteria that integrated into cells of living organisms. Afterward, they lost most of their genetic information except those required to code for replication and preservation of new virions. **The cellular origin theory** proposes that

viruses came from the DNA or RNA material of a host cell. The material auto-replicated, in time it became parasitic to the host. The third conjecture, **the independent-entity theory** asserts that viruses co-evolved with cellular constituents from the beginning of life (Flint *et al.*, 2008).

Viral diseases have a rich history in antiquity. For example, smallpox was believed to have spread from India to China and was believed to be present between 1122 and 249 B.C., as evidence of the disease was noted in China during that period. It was also thought to be present in Egypt at 1200 B.C. Poliomyelitis was also present in Egypt more than 3000 years ago based on a funeral stele from 1403-1365 (18^{th} Dynasty) which shows a priest exhibiting the typical signs of poliomyelitis (Figure 1.1). Yellow fever, another significant viral disease in virus history was predominant in North and South America and portions of Western Europe between 1791 and 1815 (Smith, 1943). The yellow fever virus is alleged to have first originated from Africa and spread to Europe due to the European slave trade as it is basically a tropical disease (Oldstone, 2010).

A Russian botanist Dmitri Iosifovich Ivanovsky, announced the first virus discovery in 1892. He proved something dissimilar to the germ theory of disease (a bacterial cause) when he noticed that a filterable agent (i.e. smaller than a bacterium) could cause disease in a healthy tobacco plant. Ivanovsky was unaware that the agent infecting the plant was a virus. Friedrich Loeffler and Paul Frosch made another striking discovery in 1898 when they detected a minute agent [later known as *Foot and mouth disease virus* (FMDV)] able to pass through a bacteria-proof filter could cause disease in cattle (Smith, 1943).

Viruses have changed the world over the years. It was recorded that some tulips in the 17th century changed colour. It is now known to be due to viral disease. This led to tulipomania where bulbs of tulips of unusual colours were sold for vast amounts of money. Potatoes in 18th century England were infected, resulting in significant crop loss (Smith, 1943). Smallpox is believed to have caused economic, religious and political change in Japan and Europe in the 16th and 17th century, when the death of ruling monarchs ushered in a change in dynasty. In the 19th century, more than 80 % of Napoleon's army, consisting of 25,000 men, died in the West Indies from a yellow fever epidemic. Most of these serious human diseases are now controlled by vaccination, indeed smallpox is completely eradicated from the world at large. Poliomyelitis has also been curbed by vaccination, and a long campaign has tried to eradicate the disease though it is still present in a small number of countries (Oldstone, 2010; Global Polio Eradication Initiative, 2016).

It can be imagined that viruses are harmful and therefore, cannot be used for any good purpose and new viruses such as HIV, SARS, Lassa fever virus, Ebola virus and Hanta virus emerged in the second part of the 20th century to strengthen this opinion. However, viruses are now used as anti-cancer agents (such as modified herpes simplex virus), as the origin of enzymes (reverse transcriptase and RNA polymerase), as pesticides (baculoviruses) and for production of large amounts of protein using virus promoters e.g. baculoviruses and adenoviruses. It is nevertheless justifiable to think that due to the harm that viruses cause, their uses are limited and still undergoing further studies (Carter & Saunders, 2007).

Figure 1.1: Very early recording of viral infection, showing an Egyptian temple priest also known as Ruma, possessing a withered right leg due to paralytic poliomyelitis. Image was taken from (Semler & Wimmer, 2002).

1.2 Classification and nomenclature of viruses

1.2.1 General classification

Virus classification and nomenclature are related because it is impossible to classify viruses without naming them. Before 1940, virologists classified viruses based on the kind of diseases they caused, the kind of host they infected, their morphology and what vector served as a route for transporting the viruses from one host to the other (Carter & Saunders, 2007). Classification systems based on the signs and symptoms of disease caused was inefficient because the same virus can show different symptoms depending on the host. Moreover, one virus can cause more than one disease. For example, the varicella zoster virus causes shingles after virus in the ganglion from an earlier chickenpox infection is reactivated (Gershon, 2006).

Classifications based on the infected host create conflicts as very different viruses can affect the same host. Groupings based on morphology assess the shape of the virus and whether the virus has an envelope or not and this became possible with the invention of the electron microscopy and negative staining techniques. Nevertheless, a drawback is that there is no detail given of the molecular biology or pathology of the virus (Dimmock *et al.*, 2007). Further types of classification now include: a) virus genome, if it is a DNA or RNA and also if it is segmented; b) nature of the nucleic acid strand, if it is single stranded or double; c) type and size of capsid and virus particle, if it is enveloped or not. Viral classification is not aimed at emphasizing hierarchy but rather, a functional classification that groups viruses according to certain similarities which may facilitate studies on the viruses (Dimmock *et al.*, 2007; Fauquet, 2010).

Fundamentally, viruses could be named based on the disease they cause e.g. yellow fever virus, human respiratory syncytial virus etc., while some others are named based on the place where the virus was first discovered e.g., *Ljungan virus*, from the Ljungan river in Medelpad county, Sweden (Stanway & Hyypiä, 1999; Tolf *et al.*, 2009) and Ebola virus, identified near the Ebola river in Zaire (Howard & Fletcher, 2012), amongst others. This also applies to naming system in both invertebrates such as insect virus and plant viruses due to viral effect(s) on the host e.g. *Tipula* iridescent virus (causes the insects to appear iridescent) and Tobacco Mosaic Virus (causes a mottling of leaves) (Carter & Saunders, 2007).

1.2.2 **Baltimore's classification**

David Baltimore came up with a system of classification grouping viruses into seven different classes depending on their mode of synthesizing virus messenger RNA (mRNA) which is then recognized by the host cell's protein synthesis machinery to allow virus replication (Figure 1.2) (Flint *et al.*, 2008). These classes are listed accordingly from group I – VII and they include: dsDNA genome, ssDNA genome, dsRNA genome, positive sense ssRNA genome, negative sense ssRNA genome, ssRNA genome (containing an enzyme that generates dsDNA as a replication intermediate) and dsDNA genome (generates ssRNA as a replication intermediate) (Baltimore, 1971). The problem with this system is clumping as some groups are similar in terms of genetic material e.g., class IV and VI, while some viruses may appear in the same group but have totally different morphology and biology (Carter & Saunders, 2007; Dimmock *et al.*, 2007).

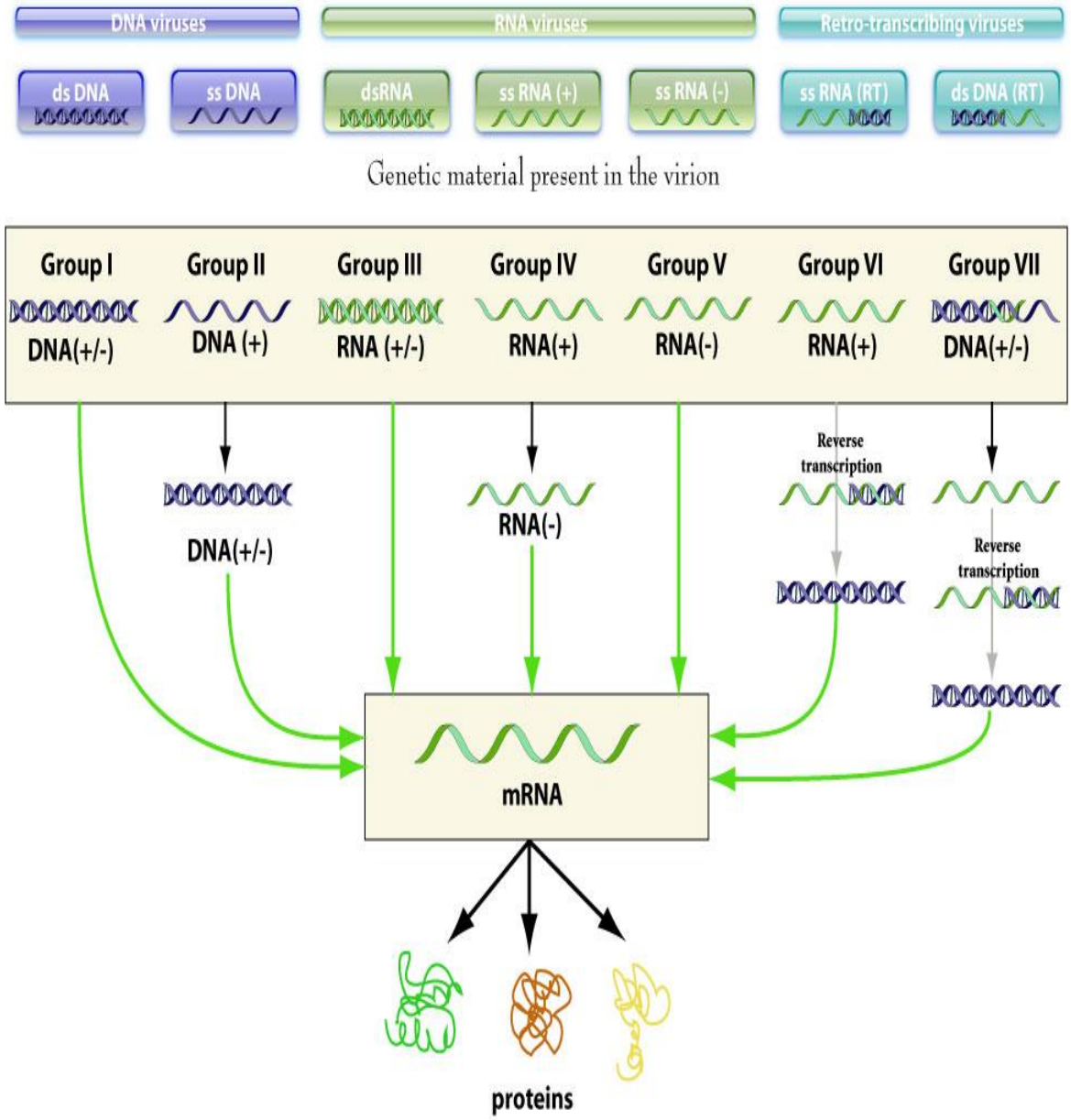


Figure 1.2: The Baltimore system of classification of viruses into 7 different types. The schematic shows how mRNA is formed from nucleic acids to generate proteins. Image was taken from http://viralzone.expasy.org/all_by_species/254.html.

Key: ds – Double Stranded, ss – Single Stranded, (+) – positive sense, (-) – negative sense, mRNA – messenger RNA and RT – Reverse Transcription.

1.2.3 **Sequence based classification**

Classification can also be based on sequence of genomes and it measures the degree to which they are conserved. Trees may be rooted, which means the ancestor of the evolved virus is known or they may be unrooted, which means the ancestry of the virus is unclear. Classification that is rooted also could be scaled or unscaled. When scaled, the various lengths on the phylogenic tree represent the evolutionary relation between the virus strains. It is useful in placing newly discovered viruses which have little host or biological information. It amounts to working from the known to the unknown. Where information is made clearer with time, the phylogenic tree of viruses can be reviewed and edited (Dimmock *et al.*, 2007; ICTV Virus Taxonomy, 2012).

1.2.4 **ICTV classification**

The International Committee on Taxonomy of Viruses (ICTV) was formed in 1966 to make the classification of viruses into group more systematic. The information recovered from the similarities in virion physical properties, genome sequence, organization and replication mode of viral genome is what serves as a base for more sophisticated classification of viruses (Bedard *et al.*, 2004; ICTV Virus Taxonomy, 2012). Virologists now use taxonomic groups to group viruses into; order, family, subfamily (in some cases), genus and species. Due to the debate if the concept of ‘species’ can be used in viral classification given viruses replicate only by cloning, the ICTV defines a virus species as: “a polythetic class of viruses that constitute a replicating lineage and occupy a particular ecological niche”, which now accommodates viruses as being grouped into species (Van Regenmortel & Mahy, 2004). A specie consists of variable strains which differ slightly from

one another in their genome sequence (genotype) and/or their serotype, which are defined by the antigens detected during infection (Carter & Saunders, 2007).Viruses can now be classified according to taxonomic groups where each group has its own suffix as seen below:

- Order : *-virales* e.g., *Mononegavirales*
- Family : *-viridae* e.g., *Paramyxoviridae*
- Subfamily : *-virinae* e.g., *Paramyxovirinae*
- Genus : *-virus* e.g., *Morbillivirus*
- Specie : -virus e.g., *Measles virus*

(Carter & Saunders, 2007; ICTV Virus Taxonomy, 2012).

The current report of virus taxonomy published by ICTV shows that seven orders have been defined namely: *Caudovirales, Herpesvirales, Ligamenvirales, Mononegavirales, Nidovirales, Picornavirales* and *Tymovirales*. There are 113 families, 27 sub-families, 616 genera and 3724 species (ICTV, 2015). Table 1.1 shows a summary of the number of families, subfamilies, genera and species classified according to ICTV. There are still many viruses that have not yet been assigned to an order. Several new viruses are being discovered, a lot of these still need to be assigned to families, genera and species. The ICTV report also contains some other sub-viral agents such as viroids and prions which have not been classified officially but have been listed because of their prior existence (ICTV, 2015). The ICTV taxonomy shows that there is an enormous number and variety of known viruses.

Table 1.1: 2015 release of virus taxonomy according to ICTV classification of viruses

Order	Family	Sub-family	Genus	Specie	Genus not assigned to a subfamily	Specie of genus not assigned to a subfamily	(Unnamed Family)		(Unnamed genus)
							genus	specie	specie
Caudiovirales	3	9	34	165	99	512	-	-	-
Herpesvirales	3	3	19	102	-	-	-	-	(1) 1
Ligamenvirales	2	-	5	12	-	-	-	-	-
Mononegavirales	8	-	30	138	-	-	(1) 5	5	-
Nidovirales	4	2	9	56	-	-	-	-	(2) 3
Picornavirales	5	1	39	142	5	16	(1) 2	4	(1) 3
Tymovirales	4	2	22	181	-	-	0	-	(3) 6
Unassigned	82	10	340	2378	-	-	-	-	-
Total	**111**	**27**	**498**	**3174**	**104**	**528**	**(2) 7**	**9**	**(7) 13**

Source: Produced from: http://ictvonline.org/virusTaxonomy.asp.

* () show the number of the ‘unnamed family’ or ‘unnamed genus’ within the order of classification.

1.3 Picornaviruses

1.3.1 Introduction

Picornaviruses belongs to the family *Picornaviridae*. These viruses have a single stranded positive sense RNA genome enclosed within a non-enveloped icosahedral capsid. The genome is 7.2 kb – 9.0 kb long and the particles have diameters of about 30 nm (Sasaki & Taniguchi, 2003a; De Palma *et al.*, 2008). 'Pico' is a prefix in denoting one million-millionth (10^{-12}), a very small quantity. As the virus consists of an RNA genome, the virus is therefore named pico-RNA-virus (King *et al.*, 2012).

Picornaviruses comprise various genera of small RNA viruses abundant in nature. They cause diseases in man, such as poliomyelitis and in livestock, such as FMDV (Tuthill *et al.*, 2010). Four genera were initially recognized based on the molecular organization of the viruses: *Rhinovirus, Enterovirus, Aphthovirus* and *Cardiovirus* (Stanway, 1990). Over the years, the study of features such as serology and the molecular properties of these viruses grouped them into nine genera (Tuthill *et al.*, 2010). However, current studies now classify picornaviruses into 31 genera and 54 species consisting of *Aphthovirus* (4 species), *Aquamavirus* (1 species), *Avihepatovirus* (1 species), *Avisivirus* (1 species), *Cardiovirus* (3 species), *Cosavirus* (1 species), *Dicipivirus* (1 species), *Enterovirus* (12 species), *Erbovirus* (1 species), *Gallivirus* (1 species), *Hepatovirus* (1 species), *Hunnivirus* (1 species), *Kobuvirus* (3 species), *Kunsagivirus* (1 species) *Limnipivirus* (3 species), *Megrivirus* (1 species), *Mischievous* (1 species), *Mosavirus* (1 species), *Oscivirus* (1 species), *Parechovirus* (2 species), *Pasivirus* (1 species), *Passerivirus* (1 species), *Potamipivirus* (1 species), *Rosavirus* (1 species), *Sakobuvirus* (1 species) *Salivirus* (1

species), *Sapelovirus* (3 species), *Senecavirus* (1 species), *Sicinivirus* (1 species), *Teschovirus* (1 species) and *Tremovirus* (1 species) (ICTV, 2015). Rhinoviruses are now included in the genus, *Enterovirus* and there is now no *Rhinovirus* genus (ICTV, 2015). The phylogenetic relationship between the species making up these genera is shown in Figure 1.3 and the abbreviations of the species are listed in Table 1.2.

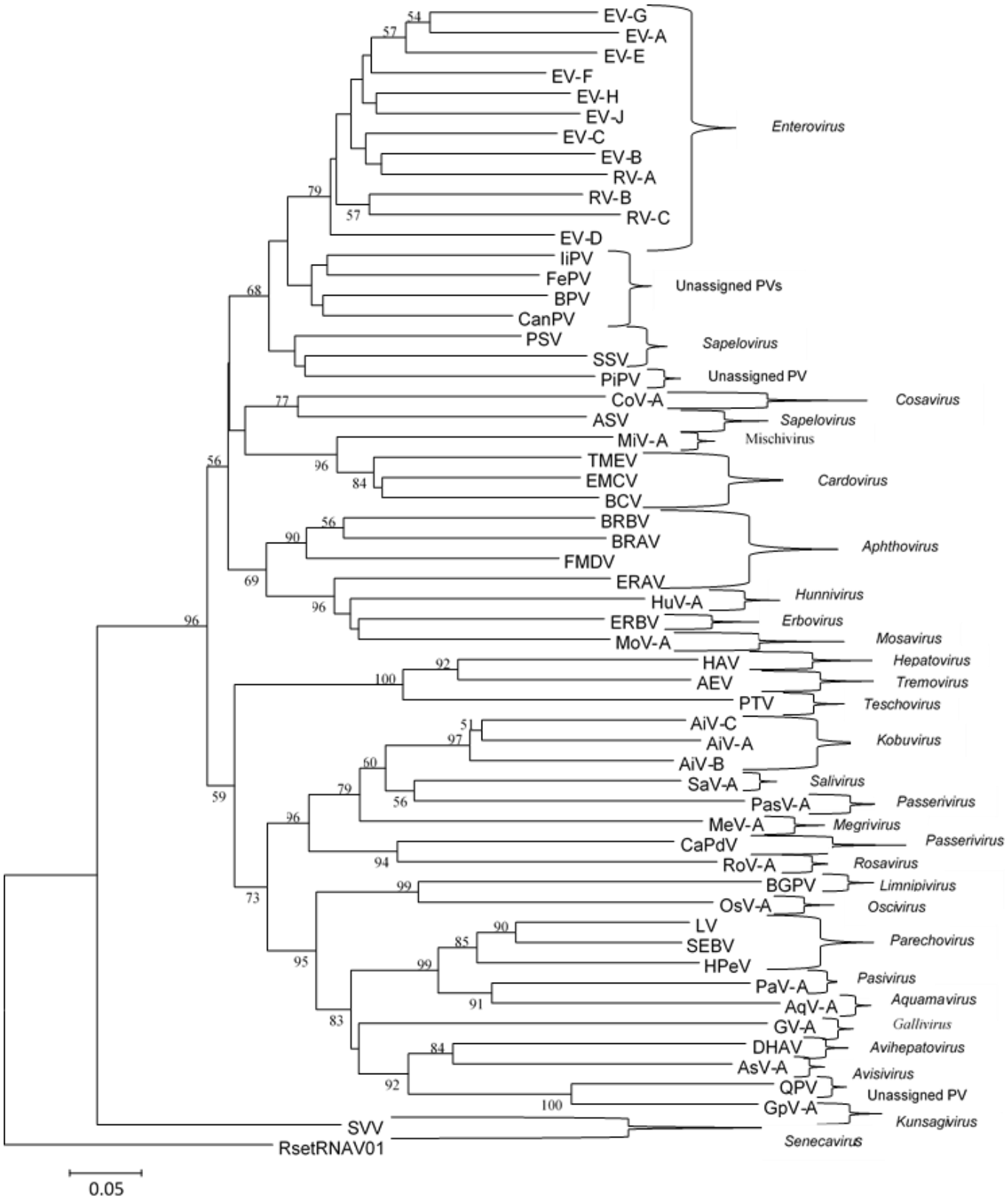


Figure 1.3: Rooted phylogenetic tree of picornavirus species, using *Rhinozosolenia setegera RNA virus 01* (RsetRNAV01) as outgroup, based on comparison of their capsid protein sequences. Alignment was carried out using CLUSTAL W2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) and the phylogenetic tree was constructed with MEGA5.2 using the neighbour-joining tree option. The bootstrap values for branches were computed using one thousand replications. Full names for the abbreviations of viruses used are showed in Table 1.2.

Table 1.2: Names of species and genus of the abbreviations used to construct the phylogenetic tree in Figure 1.3.

S/N	SPECIES	ABBREVIATIONS	GENUS
1	*Aichivirus A*	AiV-A	*Kobuvirus*
2	*Aichivirus B*	AiV-B	
3	*Aichivirus C*	AiV-C	
4	*Avian encephalomyelitis virus*	AEV	*Tremovirus*
5	*Avian Sapelovirus*	ASV	*Sapelovirus*
6	*Sapelovirus A*	PSV	
7	*Sapelovirus B*	SSV	
8	*Avisivirus A*	AsV-A	*Avisivirus*
9	*Aquamavirus A*	AqV-A	*Aquamavirus*
10	*Bluegill picornavirus 1*	BGPV	*Limnipivirus*
11	*Encephalomyocarditis virus*	EMCV	*Cardiovirus*
12	*Theiler's murine encephalomyelitis virus*	TMEV	
13	*Boone cardiovirus*	BCV	
14	*Bovine rhinitis B virus*	BRBV	*Aphthovirus*
15	*Bovine rhinitis A virus*	BRAV	
16	*Equine rhinitis A virus*	ERAV	
17	*Foot-and-mouth disease virus*	FMDV	
18	*Cadicivirus A*	CaPdV-1	*Dicipivirus*
19	*Cosavirus A*	CoV-A	*Cosavirus*
20	*Duck Hepatitis A Virus*	DHAV	*Avihepatovirus*
21	*Enterovirus A*	EV-A	*Enterovirus*
22	*Enterovirus B*	EV-B	
23	*Enterovirus C*	EV-C	
24	*Enterovirus D*	EV-D	
25	*Enterovirus E*	EV-E	
26	*Enterovirus F*	EV-F	
27	*Enterovirus G*	EV-G	
28	*Enterovirus H*	EV-H	
29	*Enterovirus J*	EV-J	
30	*Rhinovirus A*	RV-A	
31	*Rhinovirus B*	RV-B	
32	*Rhinovirus C*	RV-C	
33	*Equine rhinitis B virus*	ERBV	*Erbovirus*
34	*Gallivirus*	GV-A	*Galliivirus*
35	*Greplavirus A*	GpV-A	*Kunsagivirus*
36	*Hepatovirus A*	HAV	*Hepatovirus*
37	*Parechovirus A*	HPeV	*Parechovirus*
38	*Parechovirus B*	LV	
39	*Sebokele virus*	SEBV	
40	*Hunnivirus A*	HuV-A	*Hunnivirus*
41	*Melegrivirus A*	MeV-A	*Megrivirus*
42	*Mischivirus A*	MiV-A	*Mischivirus*
43	*Murine mosavirus*	MoV-A	*Mosavirus*
44	*Murine rosavirus*	RoV-A	*Rosavirus*
45	*Oscivirus A*	OsV-A	*Oscivirus*
46	*Pasivirus A*	PaV-A	*Pasivirus*
47	*Passerivirus A*	PasV-A	*Passerivirus*
48	*Seneca valley virus*	SVV	*Senecavirus*
49	*Porcine teschovirus*	PTV	*Teschovirus*
50	*Salivirus A*	SaV-A	*Salivirus*
51	*Rhinozosolenia setigera RNA virus 01*	RSetRNAV01	*Bacillarnavirus*
52	*Bat picornavirus 1*	BPV	Unassigned picornaviruses
53	*Canine picornavirus 1*	CanPV	
54	*Feline picornavirus 1*	FePV	
55	*Ia io picornavirus 1*	IiPV	
56	*Pigeon picornavirus A*	PiPV A	
57	*Quail picornavirus*	QPV	

1.3.2 Medical significance of picornaviruses

Some of the genera in the *Picornaviridae* family that are significant pathogens of human and animals are discussed below:

Aphthovirus

Aphthovirus contains the species: *Foot-and-mouth disease virus* (FMDV), *Bovine rhinitis A virus*, *Bovine rhinitis B virus* and *Equine rhinitis A virus*. FMDV is the most important and was in fact the very first identified animal virus, as it was able to pass through bacteria-retaining filters (Smith, 1943; Tuthill *et al.*, 2010). It is highly contagious and if affects cloven-hoofed animals such as cattle, sheep, goats and pigs. The primary site of infection is via the upper respiratory tract which is then widely spread to all parts of the body, while the secondary routes of replication are mostly in the epithelial tissues (Jackson *et al.*, 2000). Once infected, the animal continues to shed the virus in their secretions even after recovery, hence the only viable means of dealing with further spread of the virus to healthy animals is by elimination of the affected animals (Chen *et al.*, 2003). An outbreak in the UK in 2001 accounted for the loss of about six million animals and over £8 billion (Carter & Saunders, 2007). There are seven identified serotypes of FMDV which are: A, Asia-1, C, O, South African territories (SAT) types 1, 2 and 3 (Jackson *et al.*, 2000; Chen *et al.*, 2003).

Avihepatovirus

Avihepatovirus has only one specie; *Avihepatovirus A* which was formerly named *Duck hepatitis A virus* (ICTV, 215). It causes a highly contagious liver infection which is also acute and lethal in ducklings which are under the age of 6 weeks (Boros *et al.*, 2013). This

phenomenon is an issue of concern in places where ducks are produced and/or consumed (Wen *et al.*, 2014).

Cardiovirus

Cardiovirus has 3 species; *Cardiovirus A*, *Cardiovirus B* and *Cardiovirus C*. The *Encephalomyocarditis virus* (EMCV) is now placed in Cardiovirus A, *Theiler's murine encephalomyelitis virus* (TMEV), along with *Vilyuisk human encephalomyelitis virus* (VHEV), *Thera virus* (TRV) which is a Theiler-like rat virus and the new human *Saffold virus* (SAFV) discovered in 2007 are placed in *Cardiovirus B*. Cardiovirus C is made up of *Boone cardiovirus* -1 and 2 (ICTV, 2015) (Jones *et al.*, 2007).TMEV causes a loss of nerve-fibre covering in the Central Nervous System (CNS) of mice and myocarditis in humans (De Palma *et al.*, 2008). SAFV has been found to be present in different body fluids and it infects mostly young children giving rise to both asymptomatic and symptomatic infections just like other picornaviruses (Tapia *et al.*, 2015), however, it is frequently detected in patients with respiratory illness and acute gastroenteritis (Lin *et al.*, 2015).

Cosavirus

Cosavirus (common stool associated virus) includes the specie; *Cosavirus A* which was first identified in stools of South Asian children possessing non-polio acute flaccid paralysis (AFP) (Kapoor *et al.*, 2008). Its closest relatives are cardioviruses and *Senecavirus*. It has been found in stool samples of both healthy and ill children from Afghanistan, Australia, China, Pakistan and also the United Kingdom. It has also been detected in adult patients in Brazil, Northern Italy and Thailand (Kapusinszky *et al.*, 2012; Campanini *et al.*, 2013). The presence of cosaviruses was also detected in waste water treatment plants in the United States (Kitajima *et al.*, 2015).

Enterovirus

Enterovirus is a large genus consisting of 12 species; *Enterovirus* (EV) A, EV-B, EV-C, EV-D, EV-E, EV-F, EV-G, EV-H, EV-J, *Rhinovirus* (RV) A, RV-B and RV-C of which EV-A to EV-D and RV-A to RV-C are human pathogens. EV-A to EV-D include coxsackieviruses, echoviruses and polioviruses as well as several numbered enteroviruses, EV-D68 to EV-A121. RV-A to RV-C include human rhinoviruses (Chen *et al.*, 2015) (ICTV, 2015). EVs are now classified into types based on the identity of the VP1 (capsid) sequence (Oberste *et al.*, 2000). The name of the numbered EV types include the species which they belong to e.g. EV-D68 (Species EV-D). Enteroviruses replicate in the enteric tract (hence the name 'entero'- viruses), causing fever and rashes and may produce more complicated infection when the virus spreads to other organs of the body, like the heart or nervous system. Examples include hand foot and mouth disease (HFMD), poliomyelitis, encephalitis, myocarditis, neonatal sepsis and even common cold (Strauss & Strauss, 2007; Chen *et al.*, 2015). It is well known that different EVs are capable of producing the same clinical effects in an infected patient and also, a particular type of EV can produce different clinical signs and symptoms. For example, EV-A71 is capable of causing both HFMD and neurological infections. Also, a good number of EVs can cause HFMD as well as neurological infections (Van Leer-Buter *et al.*, 2016). Poliovirus has caused human disease for a long time and was a common infection until vaccines were introduced. There are two types of vaccines, the Salk and Sabin vaccines, both having been used for over 60 yrs to prevent infection. The Salk is an inactivated polio vaccine (IPV) and Sabin is a live attenuated vaccine (OPV-oral poliovirus vaccine). A major campaign will hopefully lead to the eradication of the polioviruses (Sanders *et al.*, 2016).

Rhinoviruses cause more than half of the common cold cases in the world, as well as being more medically important by making the symptoms of diseases such as asthma worse (De Palma *et al.*, 2008). The three species of RVs have over 160 types. RV-C was the last to be discovered, in 2006, because members of this species are not easily detectable by standard culture techniques. They also cause more severe illness, especially in children, that could even lead to hospitalization when compared with RV-A and RV-B (Bochkov *et al.*, 2015).

Erbovirus

Erbovirus has only one specie, *Erbovirus A* formerly named *Equine rhinitis B virus,* and it has 3 types; ERBV1, ERBV2 and ERBV3. It is also known as the acid-stable equine picornavirus (King *et al.*, 2012). They cause respiratory illness in horses with signs and symptoms such as nasal discharge, cough and fever (Woo *et al.*, 2016). Infection with *Erbovirus A* is of economic significance because it affects the performance of the infected horses and could also be a source of disease to the horse owners. The virus was mainly detected in nasal discharge until recently, when they were also detected in faecal samples of infected horses (Woo *et al.*, 2016).

Hepatovirus

Hepatovirus consists of the *Hepatitis A Virus* (HAV) which causes hepatitis A, an acute liver infection (De Palma *et al.*, 2008). It is transmitted via the faecal-oral route due to the ingestion of contaminated food and water, but principally through person-to-person by occult contamination with faeces. It is therefore endemic in regions where there are inadequate sanitary conditions and poor personal hygiene (Purcell & Emerson, 2001). Studies show that almost everyone in continents such as Africa, Asia and South America become seropositive to HAV by the age of 5 (Purcell & Emerson, 2001)

HAV is a non-enveloped virus but it circulates in the blood of an infected individual with an envelope. This it acquires from the cell membrane of the host during exit from the liver cells (Drexler *et al.*, 2015).

Kobuvirus

Kobuvirus contains 3 species; *Aichivirus A, Aichivirus B and Aichivirus C* (Olarte-Castillo *et al.*, 2015)*. Aichivirus A* was first discovered in 1989 at a place called 'Aichi' in Japan, in humans infected with oyster-associated, non-bacterial gastroenteritis (Yamashita *et al.*, 1995). Aichiviruses have also been known to infect animals such as oxen, sheep, pigs, rodents, canines and horses (Carmona-Vicente *et al.*, 2013; Van Dung *et al.*, 2016). Aichiviruses have not yet been detected in non-mammal species (Pankovics *et al.*, 2015). Kobuvirus is endemic in many countries around the world where there is consumption of oysters as well as other shell-fish and it is also prevalent in the cat and dog populations in the UK (Carmona-Vicente *et al.*, 2013).

Limnipivirus

Limnipivirus is a new genus within the *Picornaviridae* family consisting of 3 species; *Limnipivirus A*, *Limnipivirus B* and *Limnipivirus C*. It is named from the Greek word 'limne' which means lake and *Picornavirus* and it was first isolated from the heart, brain and liver of a carp (*Cyprinus carpio*) but the link to fish disease is not clear (Lange *et al.*, 2014).

Sapelovirus

Sapelovirus is made up of 3 species; *avian sapelovirus*, *Sapelovirus A* (formally named *porcine sapelovirus*) and *Sapelovirus C* (formerly named *Simian sapelovirus*). The name sapelovirus is derived from **s**imian, **a**vian, and **p**orcine **e**ntero-**l**ike virus (Lan *et al.*, 2011). Sapeloviruses cause asymptomatic infection as well as mild and severe diarrhea,

pneumonia, respiratory distress, polioencephalomyelitis and reproductive disorders (Donin *et al.*, 2014; Son *et al.*, 2014)

Senecavirus

Senecavirus contains the *Senecavirus A* (formerly named *Seneca valley virus*) as the only specie in the genus. The virus is still called the Seneca valley virus despite the change of the specie name. It was first isolated from contaminated cell culture media in 2002 (Willcocks *et al.*, 2011; Bracht *et al.*, 2016). The primary host of the Seneca virus are pigs but studies show that it is a potential anti-cancer agent as it selectively replicates in and also destroys neuroendocrine tumour cells in human (Reddy *et al.*, 2007).

Teschovirus

Teschovirus contains one specie; *Teschovirus A* formerly named *Porcine teschovirus*, is associated with teschen (severe) to talfan (mild) diseases in pigs, causing distress to their digestive, respiratory and female reproductive tract as well as the central nervous system (Cano-Gomez & Jimenez-Clavero, 2016). The transmission of the virus is mostly via the faecal-oral route and pigs in endemic regions are mostly asymptomatic (Donin *et al.*, 2015; Tsai *et al.*, 2016).

1.3.2.1 Parechovirus

The genus *Parechovirus* is made up of 2 species; *Parechovirus A* formerly called *Human parechovirus* (HPeV) and *Parechovirus B* formerly called *Ljungan virus* (LV) (Pham *et al.*, 2010). Two new proposed species; “Sebokele virus 1” and “Ferrent parechovirus” have recently been described and may be included in the *Parechovirus* genus (Smits *et al.*, 2013).

Parechovirus B (Ljungan virus) (LV)

LV was initially discovered from Swedish bank voles (*Clethrionomys glareolus*) and over the years it has been reported in small bank voles (*Myodes glareolus*) found in northern Italy and recently also reported in Eurasian red squirrels (*Sciurus vulgaris*) (Romeo *et al.*, 2014). There are four identified genotypes of the LV, two isolated in Sweden and the other two from the United States (Tolf *et al.*, 2009; Zhu *et al.*, 2015). Although records on the LV infections in humans are rare, it was predicted to be the cause of human diseases such as type 1 diabetes mellitus and myorcarditis (Niklasson *et al.*, 2003). It was also reported to be found in cases of sudden infant death syndrome and malfunctions of the central nervous system (CNS) (Tolf *et al.*, 2009; Jääskeläinen *et al.*, 2015).

A recent study in Finland showed that 36 % of patients had LV-reactive antibodies and these were patients who were mostly below the age of 14 suggesting they were infected at an early age. There was a higher proportion of patients with LV-reactive antibodies in the urban areas than in the rural areas therefore suggesting human-to-human transmission of the LV or Ljungan-like virus amongst the human population rather than via direct zoonosis by rodents (Jääskeläinen *et al.*, 2015). This human-to-human transmission is plausible because just like many other picornaviruses, LV is thought to be transmitted via the faecal-to-oral route even though rodents remain the primary host of the virus (Pounder *et al.*, 2015).

Parechovirus A (Human parechovirus) (HPeV)

Two Parechoviruses were first identified in 1956, thought to be enteroviruses and called echovirus 22 and echovirus 23. It later became clear that they are genetically distinct from enteroviruses (Hyypiä *et al.*, 1992; Stanway *et al.*, 1994). They were renamed Human

parechovirus 1 (HPeV1) and HPeV2 and placed in a specie *Human parechovirus* in the *parechovirus* genus. Parechoviruses became more prominent in 2004 when the type HPeV3 was discovered to cause neonatal sepsis in humans. Up till then, HPeVs were thought to cause only mild illness in children (Wildenbeest *et al.*, 2010). HPeV3 could be the cause of most neonatal deaths in Europe and in the United States due to neurological infections but tests for HPeV3 is not routinely performed in hospital laboratories (Cabrerizo *et al.*, 2015; Sharp *et al.*, 2015). There are currently 16 HPeV types (HPeV1 through HPeV16) (Knowles, 2013), listed in Table 1.3 (www.picornaviridae.com).

Over 95 % of humans become seropositive to HPeVs early in life between the ages of two to five years (Stanway *et al.*, 2000; Harvala *et al.*, 2010) and they spread easily most likely because the most probable route for the transmission of the virus is the faecal-oral route. Also, recent studies have shown that HPeV infections are most common in autumn (Lodder *et al.*, 2013; Yip *et al.*, 2014).

HPeV1 remains the most common genotype worldwide, the second is HPeV3 while the remaining genotypes are less common (Cabrerizo *et al.*, 2015). HPeV1 and HPeV2 have been reported to have caused mild gastroenteritis and respiratory illness although, HPeV1 may cause sepsis-like illness or meningitis (Benschop *et al.*, 2008; Harvala *et al.*, 2008). In a recent study, a case of a child diagnosed with dilated cardiomyopathy and myocarditis caused by HPeV1 was successfully treated with intravenous immunoglobulins (IVIGs) (Wildenbeest *et al.*, 2013).

The characteristics of the infections of HPeV types 4-16 are mostly clinically unexplored, although most have been reported to cause diarrhoea in young children (Yip *et al.*, 2014; Sharp *et al.*, 2015). One report has linked HPeV4 with neonatal sepsis (Kolehmainen *et al.*, 2014a).

Table 1.3: Human Parechovirus types.

Type	Prototype strain	Reference
HPeV1	Harris	Hyypiä *et al*., 1992
HPeV2	Williamson	Ghazi *et al*., 1998
HPeV3	A308/99	Ito *et al*., 2004
HPeV4	K251176-02	Benschop *et al*., 2006
HPeV5	CT86-6760	Oberste *et al*., 1998
HPeV6	NII561-2000	Watanabe *et al*., 2007
HPeV7	PAK5045	Li *et al*., 2009
HPeV8	BR/217/2006	Drexler *et al*., 2009
HPeV9	BAN2004-10902	Oberste *et al*., unpub.
HPeV10	BAN2004-10903	Oberste *et al*., unpub.
HPeV11	BAN2004-10905	Oberste *et al*., unpub.
HPeV12	BAN2004-10904	Oberste *et al*., unpub.
HPeV13	BAN2005-10901	Oberste *et al*., unpub.
HPeV14	451564 (NL, 2004)	Benschop *et al*., 2008
HPeV15	BAN-11614	Oberste *et al*., unpub.
HPeV16	BAN-11615	Oberste *et al*., unpub.

1.3.3 **Picornavirus replication**

An overview of picornavirus replication is shown in Figure 1.4. When the virion binds to receptors and co-receptors, it gains entry into the cell, usually in endocytic vesicles, and uncoats. Uncoating is due to capsid instability, caused by interaction with the receptor and/or pH changes in the vesicles. VP4 is externalized and interacts with the host cell or vesicle membrane which is one of the steps in uncoating (De Palma *et al.*, 2008). The viral genome which functions as mRNA, is released from the capsid. The open reading frame (ORF) is translated by cell ribosomes to one polyprotein which is cleaved by virus proteases to 10, 11 or 12 proteins (Figure 1.5) (Lin *et al.*, 2009).

Translation of the viral genome takes place in the cytoplasm of infected cells using host ribosomes which are recruited by the virus IRES. Virus proteases are necessary for cleavage of the picornavirus polyprotein which is generated. The virus polymerase together with the aid of VPg and an RNA structure called the cis-acting replication element (*cre*) help to bring about synthesis of negative sense RNA. This serves as a template for synthesis and elongation of new viral genomic RNA (plus sense) (Steil & Barton, 2009). After replication, only synthesized plus strand RNA undergo encapsidation suggesting it is a highly specific procedure (Nugent *et al.*, 1999). The progeny virus exits the cell as it is ruptured. The release of newly formed viral particles ends the replication process of the virus (Sasaki & Taniguchi, 2003a).

The replication process is aided by the actions of the encoded proteins, as well as host cell proteins (Figure 1.4). During the course of replication in several picornaviruses, host

cell transcription and translation are shut off completely to aid efficient utilization of host cell resources by the viral genome (Bushell & Sarnow, 2002).

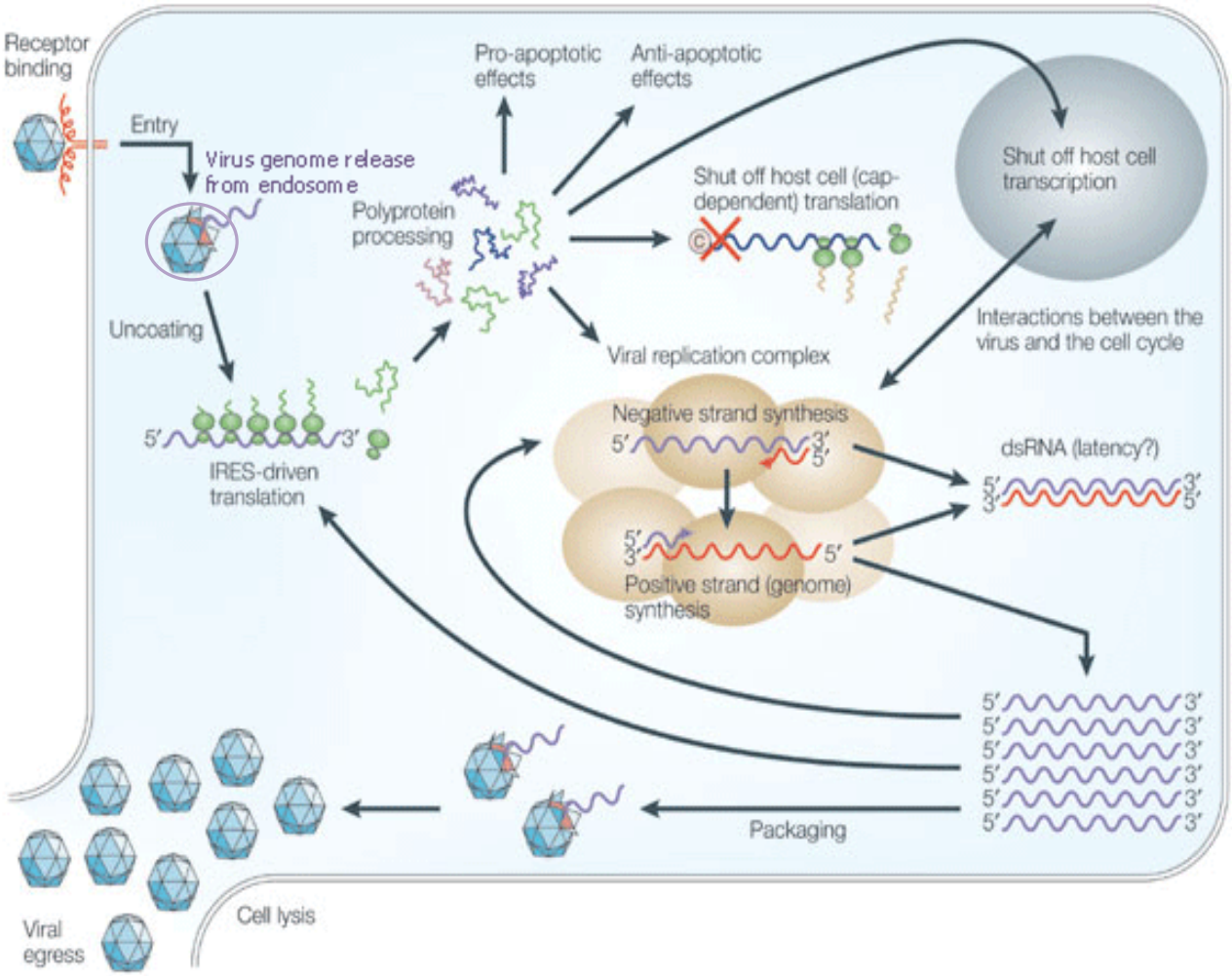


Figure 1.4: The life cycle of a picornavirus, showing the major stages of replication. Positive strand RNA is represented with purple while negative strand RNA is represented with red (Adapted from Whitton *et al*., 2005).

1.3.4 **Picornavirus genome organization and expression**

The picornavirus RNA genome (Figure 1.5) functions as a messenger RNA (mRNA) which directly encodes a single precursor polyprotein that is cleaved to both structural and non-structural proteins (Dimmock *et al.*, 2007; Liu *et al.*, 2009a). At the 5' end of the genome is the 5' Untranaslated region (UTR) and there is a 3' UTR followed by a Poly(A). The length of the Poly(A) tail varies between 65 and 100 nucleotides bases (Lin *et al.*, 2009). The 5' UTR is made up of an Internal Ribosome Entry Site (IRES) that assembles ribosome precursors to begin translation, a mechanism which is cap-independent. The 5' UTR and 3' UTR contain RNA structures needed for replication of the virus genome (Liu *et al.*, 2009a). Structural proteins are encoded in the P1 region at the 5' end in the viral open reading frame (ORF) and non-structural proteins by P2 and P3. The 5' UTR is about 600 - 1200 nt in length, while the 3' UTR is around 30 -350 nt. (Bergamini *et al.*, 2000; Mattick, 2003; Lin *et al.*, 2009). *Aphthovirus* and *Cardiovirus* contain a Polyribocytidylic acid (Poly(C)) tract shown to be near the 5' end of the viral RNA, which is important for viral replication (Strauss & Strauss, 2007). Five classes of IRES (I - V) have been identified in picornaviruses. Type I e.g. in *Enterovirus*, type II e.g. in *Cardiovirus, Aphthovirus* and *Parechovirus*, type III only in *Hepatovirus,* type IV e.g. in *Teschovirus, Senecavirus* and *Avihepatovirus* (Liu *et al*., 2009) and type V e.g. in *Kobuvirus* (Sweeney *et al*., 2012).

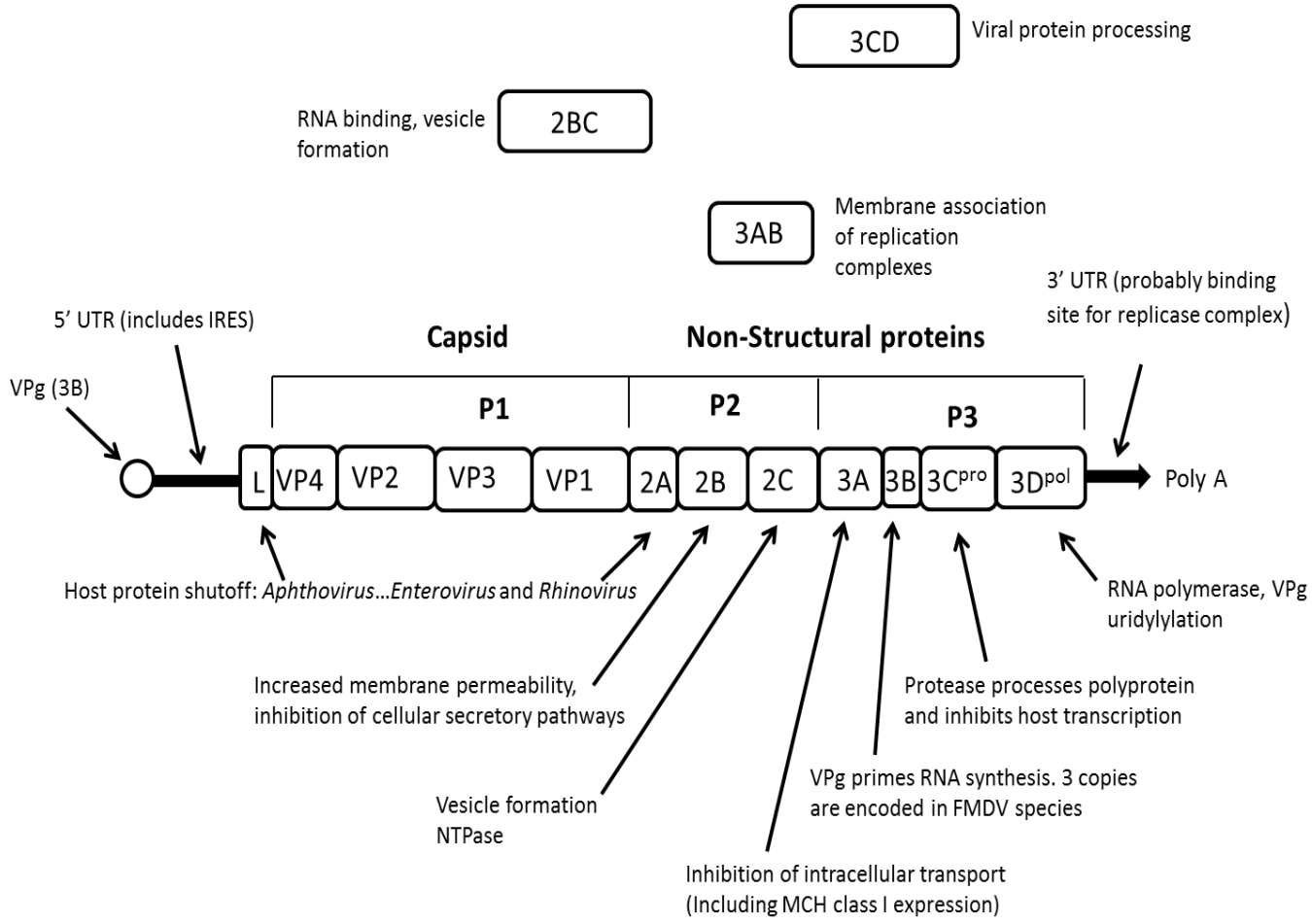


Figure 1.5: Picornavirus genome showing the 12 proteins derived from a typical polyprotein and their functions. 2BC, 3AB, 3CD are precursors of some of the final proteins and have separate functions. Only some picornaviruses have an L protein and this has a different structure/function in different genera. The 2A protein is also very diverse between different genera and some picornaviruses have more than one protein at this position (Adapted from Whitton *et al*., 2005).

1.3.4.1 Leader protein (L^{pro})

Picornaviruses such as *Aphthovirus*, *Hepatovirus*, *Cardiovirus* and *Kobuvirus* encode a Leader protein (L^{pro}) (Delhaye *et al.*, 2004). This is located just before the structural protein encoding region. It has different structures and functions in different picornaviruses. In aphthoviruses, it inhibits alpha/beta interferon produced early after viral infection, catalyzes its release from VP4 and aids the shut off of host cellular protein synthesis as it is a protease (Strauss & Strauss, 2007; De Palma *et al.*, 2008).

1.3.4.2 Capsid proteins

The capsid of picornaviruses consists of 60 copies of each viral protein VP1 to VP4, encoded by the P1 region of the viral genome (Figure 1.6). VP1, VP2 and VP3 are mainly external and VP4 is internal. VP1, VP2 and VP3 have the same core structure, an 8-strand β-barrel, while VP4 has little secondary structure and is much shorter (Carter & Saunders, 2007). Parts of the C-termini of VP1 to VP3 and several loops joining β-strands are expressed on the surface of the virus, while all the parts of VP4 and the remaining parts of the other 3 structural proteins are located in the interior part of the virion. Some of these regions of the proteins on the surface of the capsid are variable between different viruses and this is a major factor in determining the antigenic properties of the virus and defining serotypes. The VP0 of viruses like *Parechovirus* and *Kobuvirus* is not cleaved to VP2 and VP4, but remain as VP0 (Figure 1.6). Thus there are only three structural proteins after processing of P1 (Stanway *et al.*, 1994; Yang *et al.*, 2009).

Receptor attachment sites are located on the surface of the capsid. In some genera, such as *Enterovirus*, these receptor attachment sites are located in canyon structures surrounding the 5 fold axis. Residues of amino acids in the canyons of polioviruses and

rhinoviruses bind to the amino-terminal domain of receptors, which prompts the instability and uncoating of the capsid (De Palma *et al.*, 2008; Lin *et al.*, 2009).

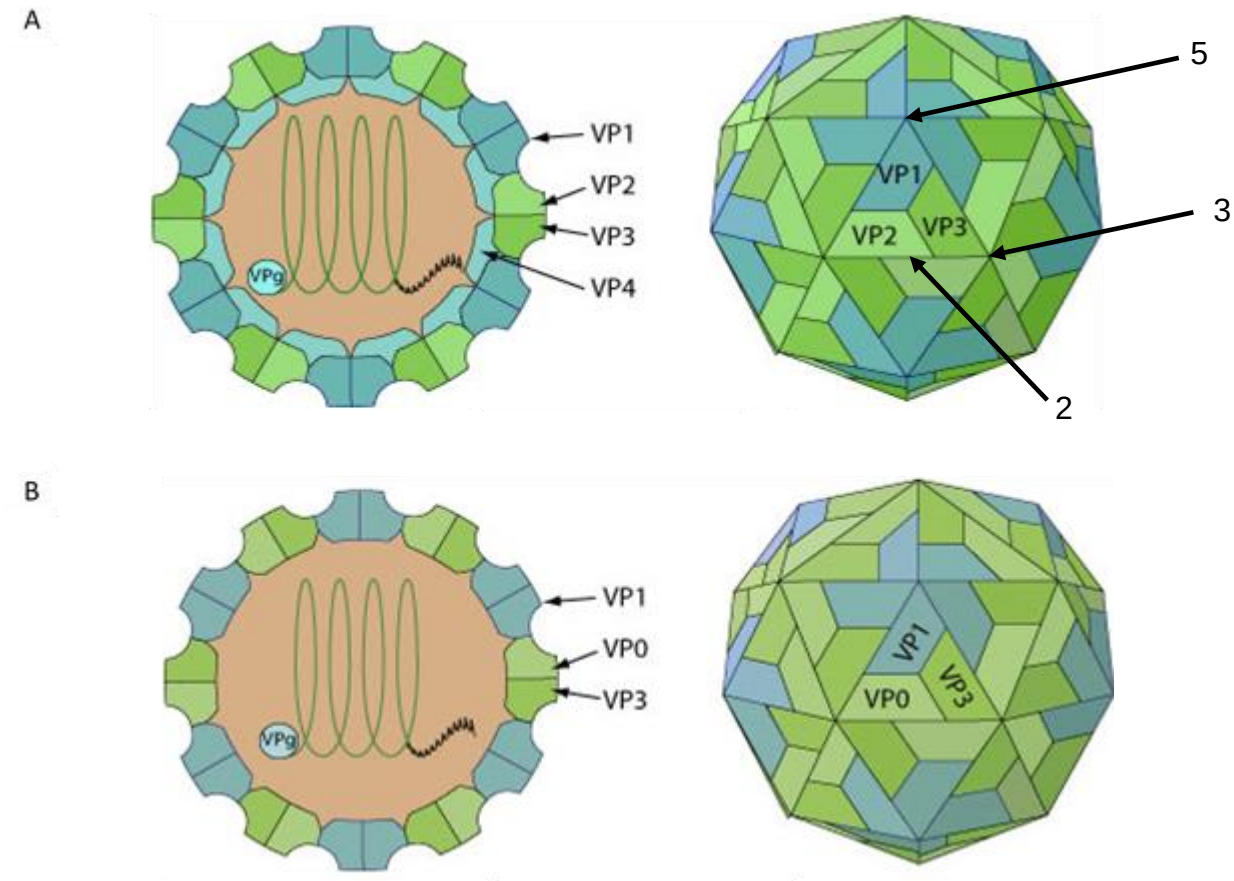


Figure 1.6: Schematic representation of the icosahedral capsid of an A: *Enterovirus*, made up of 60 copies of the 4 polypeptides; VP1, VP2, VP3 and VP4. VP4 is located on the internal side of the capsid and B: *Parechovirus*, also made up of 60 copies of each 3 proteins encoded by the virus; VP0, VP1 and VP3. Taken from http://viralzone.expasy.org/all_by_protein/97.html and http://viralzone.expasy.org/all_by_protein/657.html for *Enterovirus* and *Parechovirus* respectively.

The 5-, 3- and 2- fold axis of symmetry of the icosahedral particle are also marked.

1.3.4.3 P2 and P3 proteins and their functions

In enteroviruses 2A and 3C proteases are necessary during viral replication to cleave the polyprotein into its major components P1- P3. In addition, they cleave host cell proteins preventing host cell's transcription and translation (Bushell & Sarnow, 2002; Strauss & Wuttke, 2007; Lin *et al.*, 2009). The 2A in all other picornaviruses does not function as a protease. In several though, a distinctive sequence including the motif NPGP prevents peptide bond formation at the C-terminus of 2A, meaning that the polyprotein is made in two pieces (Donnelly *et al.*, 2001). Other picornaviruses, including parechoviruses, have a 2A which is a relative of a family of proteins involved in the control of cell proliferation (Hughes and Stanway, 2000). 2BC is the precursor protein of 2B and 2C proteins. 2B is responsible for increasing cell membrane permeability (Barco & Carrasco, 1995). 2C has an NTPase activity and is considered essential for synthesis of viral negative sense RNA from positive sense RNA during the replication cycle (Banerjee & Dasgupta, 2001).

3A inhibits intracellular transport between the endoplasmic reticulum (ER) and the Golgi apparatus (Doedens & Kirkegaard, 1995; Wessels *et al.*, 2005). In some viruses, 2BC is also involved in this inhibition (Moffat *et al.*, 2005). 3A modifies the structure of the cell membrane of the host cell (Suhy *et al.*, 2000) and 3A from FMDV has been reported to inhibit the interferon beta signaling pathway (Li *et al.*, 2016). 3AB stimulates the cleavage of $3CD^{pro}$ and helps 3CD to form a complex with a clover-leaf structure at the 5' end of the *Enterovirus* 5' UTR (Harris *et al.*, 1994; Molla *et al.*, 1994). It also helps in the uridylylation of VPg (Pathak *et al.*, 2007).

The picornavirus 3B, viral protein genome (VPg), is made up of about 20 - 25 amino acids (Lin *et al.*, 2009). VPg is a soluble product of the cleavage of 3AB and it serves as a primer

for the synthesis of both positive and negative sense RNA during genome replication, following uridylylation by $3D^{pol}$ (Strauss & Wuttke, 2007). It is linked covalently to the 5' end of the viral genome via a hydroxyl group of a tyrosine residue at position 3 on the VPg (Pathak *et al.*, 2007).

$3CD^{pro}$ displays protease action on the P1 region of the polyprotein during processing (Ypma-Wong *et al.*, 1988; Strauss & Wuttke, 2007; Liu *et al.*, 2010). It also contributes to specificity of cre binding to VPg/VPgpUpU$_{OH}$ (Shen *et al.*, 2007). Although it is the precursor for $3C^{pro}$ and $3D^{pol}$, it does not directly polymerize the viral genome. Instead, it determines the cellular distribution of $3D^{pol}$ which is then released from 3CD by cleavage which in turn aids replication. $3C^{pro}$ is the main protease in all picornaviruses and performs all the P2 and P3 cleavages (Ryan & Flint, 1997; Lin *et al.*, 2009).

$3D^{pol}$ is the RNA-dependent RNA polymerase and it is one of the most important elements among other cellular and viral proteins involved in the RNA replication complex, thus serving as target sites for antiviral drug development (Gong & Peersen, 2010). McBride *et al*. (1996) demonstrated by using a yeast di-hybrid system that the host cell protein Sam68, interacts with $3D^{Pol}$ in poliovirus and Sam68 plays an important role in its replication. Efforts have been made to prevent viral replication by the use of nucleoside analogs whose aim is to terminate or bring about errors during synthesis of new genome by $3D^{pol}$. Sadly, this goal has not been met as the virus becomes resistant to the drugs (Kerkvliet *et al.*, 2010). This is because $3D^{pol}$ is highly prone to error during initiation, elongation or termination, and lacks a proof-reading function so is very likely to produce mutant strains differing from the original viral strain (Barr & Fearns, 2010).

In addition to random incorporation of mutations during replication, picornaviruses also evolve through recombination. This is often identified by taking the genome sequence of two distant regions of the viral genome, VP1 and $3D^{Pol}$. Many picornaviruses have been shown to exhibit frequent recombination and this could be an important factor in pathogenicity (Benschop *et al.*, 2008; Calvert *et al.*, 2010; Lukashev, 2010). For instance recombinants between poliovirus vaccine strains and an unknown enterovirus led to cases of poliomyelitis in Hispaniola (Kew *et al.*, 2002; Martín *et al.*, 2002).

1.3.5 **Picornavirus receptors**

Viral receptors are specific molecules on the host cell surface that fit perfectly like a lock-and-key system to the attachment site of the virion (Rieder & Wimmer, 2002). Receptors partly determine virus host range, susceptibility and pathogenicity and constitute potential targets for antiviral therapy (Schneider-Schaulies, 2000). To infect a cell, a virus must first bind to a receptor, uncoat (which is triggered by destabilization) and then release the viral genome. This means that not every viral-receptor interaction leads to infection. For example, Coxsackievirus A21 (CVA21), an enterovirus, binds to Chinese Hamster Ovary (CHO) cells which expresses human decay-accelerating factor (DAF), a receptor, but no infection occurs in the cell as the virus requires the receptor ICAM-1 to enable uncoating (Shafren *et al.*, 1997). DAF has also been implicated as a receptor for coxsackie B viruses (CVBs), several echoviruses and enterovirus 70 (E70) (Rieder & Wimmer, 2002). Again, other receptors are needed by some of these viruses to give an infection.

The receptors for picornaviruses have been extensively studied and several belong to the integrin receptor family or the immunoglobulin superfamily (IgSF) (De Palma *et al.*, 2008). Studies show that closely related picornaviruses can use different receptors. For example some rhinoviruses interact with the low-density lipoprotein (LDL) receptor but most with intercellular adhesion molecule-1 (ICAM-1) receptor. Viruses from different genera can utilize the same receptor e.g., FMDV, Coxsackievirus A9 (CVA9), and HPeV 1 use αvβ6 integrin as a receptor (Jackson *et al.*, 2000; Williams *et al.*, 2009; Seitsonen *et al.*, 2012). Many integrin receptors recognize the arginine-glycine-aspartic acid (RGD) motif serving as the receptor binding site for FMDV, HPeV1 and CVA9. However, the receptor for echovirus 1 is integrin α2β1 and this does not recognise RGD motifs and so echovirus 1 binds in an RGD-independent manner (Rieder & Wimmer, 2002).

Poliovirus recognises CD155, CVBs identifies coxsackie-adenovirus receptor (CAR), Hepatitis A virus recognizes Hepatitis A virus cellular receptor-1 (HAVcr-1). Heparan sulphate has been implicated as a receptor for certain strains of FMDV, TMEV CVA9 and Coxsackievirus B3 (CVB3) and it is thought to act as a low-affinity and low-specificity anchor to hold the virus onto the cell surface before it is recognised by a more specific receptor (De Palma *et al.*, 2008; McLeish *et al.*, 2012).

Some viruses require or use additional cellular molecules known as co-receptors, secondary receptors or accessory factors to bring about an infection in the cell (Carter & Saunders, 2007). CVA9 and several echoviruses have been reported to strongly require β2-microglobulin to infect different cell cultures such as green monkey kidney cells (Rieder and Wimmer, 2002). As already mentioned, CVA21 uses both DAF and ICAM-1 for entry, while CVBs use both DAF and CAR. Figure 1.7 shows a schematic representation of some of the picornaviruses receptors.

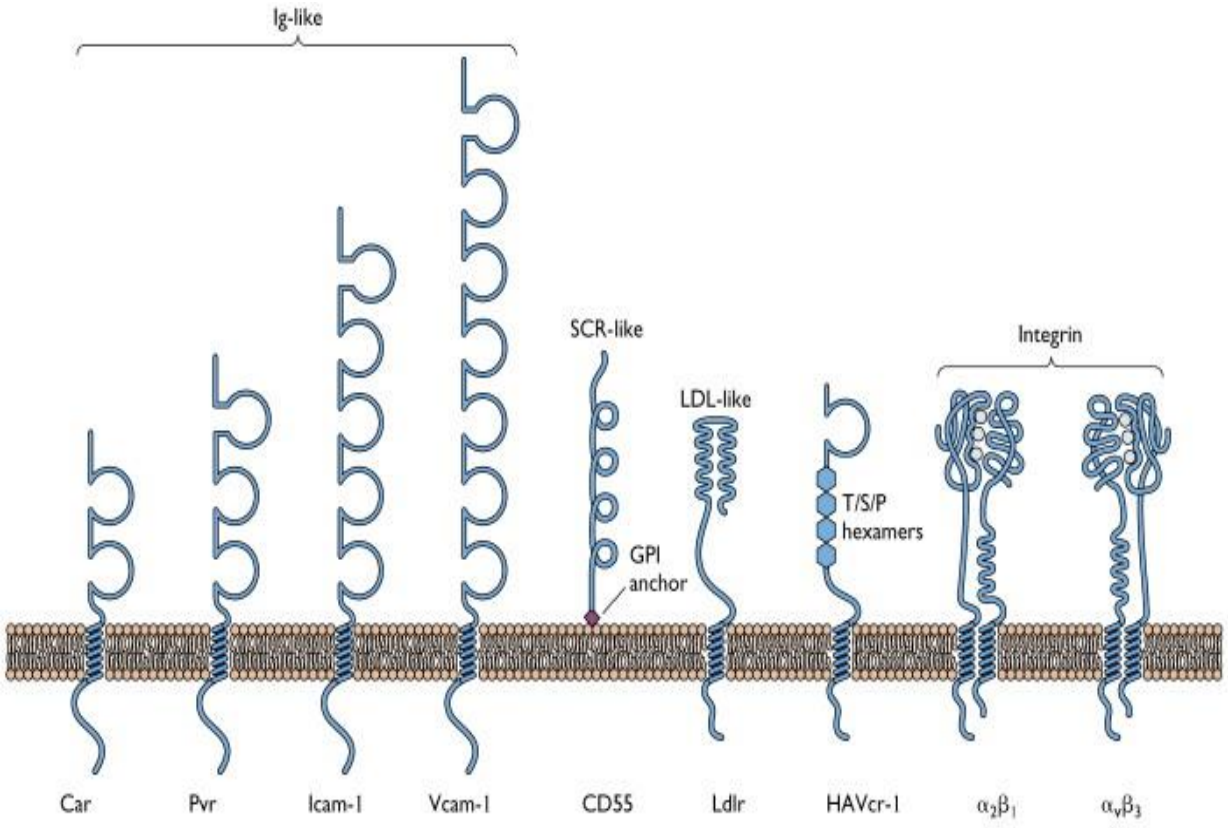


Figure 1.7: Schematic representation of picornaviruses receptors. Taken from Flint et al., 2009.

- Ig- Immunoglobulin superfamily
- SCR- Short consensus repeat
- LDL-Low density lipoprotein
- T/S/P- Threonine/serine/proline
- Car- Coxsackievirus and adenovirus receptor
- PVR- Poliovirus receptor
- Icam-1- Intercellular adhesion molecule 1
- Vcam-1- Vascular cell adhesion molecule 1
- CD55- Cluster of differentiation which is also known as DAF- decay acceleration factor
- Ldlr- Low-density lipoprotein receptor
- HAVcr-1- Hepatitis A virus cellular receptor 1

1.4 Determination and Prediction of RNA structures

RNA structures play an important role in picornavirus molecular biology, for instance in the IRES, which has a complex structure, the cre needed for VPg uridylylation and structures seen in the virus 3' UTR. To understand the role of the picornavirus structures, as well as RNA structures in other viruses and cellular RNAs, a number of methods have been used to determine and predict RNA structures.

1.4.1 Determination of RNA structures

1.4.1.1 Chemical and enzymatic reactivity

Chemicals or enzymes can show a difference between paired and unpaired nucleotides in an RNA sequence as they may recognise these nucleotides differently. This allows structural information to be inferred. The information is then usually compared with the results from structure predictions using thermodynamic folding algorithms to determine their reliability (Wilkinson *et al.*, 2008).

Evstafieva *et al*., (1991), studied a conserved part of the 5' UTR RNA structure of the EMCV by using the enzymes RNase T2, which is specific to RNA sites that are single-stranded, and RNase V1 (made from cobra venom) which cleaves double stranded RNA. The RNA was cleaved partially and the cleaved sites were detected by using a primer to prime reverse transcription and comparing the lengths of the products with an RNA sequencing ladder.

An example of a chemical method used in the determination of the structure of RNA is dimethyl sulfate (DMS). This is normally used for RNAs that are less than or equal to 500 nucleotides in length. DMS modifies bases in the RNA by methylation and modified bases stop the primer extension reaction. Modification is prevented by base-pairing in the RNA structure and so base-paired sites can be recognised (Tijerina *et al.*, 2007).

The most widely used chemical method to probe RNA molecules is selective 2' hydroxyl acylation analysed by primer extension (SHAPE).

In SHAPE, the RNA is treated with reagents such as N-methylisotoic anhydride (NMIA) or 1-methyl-7-nitroisatoic anhydride (1M7) which selectively reacts to the 2'-hydroxyl group to form adducts which is then identified by primer extension (Wilkinson *et al.*, 2008; Deigan *et al.*, 2009). Again the reverse transcriptase cannot continue copying passed the modified nucleotide and so these sites can be recognised. The different reactivity of structured and unstructured nucleotides means that each nucleotide can be examined and compared with predicted structures (Low & Weeks, 2010). SHAPE analysis was used to study the most recently discovered IRES type, type V (Sweeney *et al.*, 2012) and also the magnesium dependent folding of the FMDV IRES (Lozano *et al.*, 2014).

1.4.1.2 Physical method

Electron Microscopy (EM)

EM can be used to determine the structure of macromolecules at nanometer resolution. It was originally difficult to get a good image contrast for very small biomacromolecules (<200 kDa) until development in cryogenic electron microscopy (Cryo-EM) and cryogenic electron tomography (Cryo-ET) (Gong *et al.*, 2015). In the Cryo-ET technique, images that

are projected from waves are aligned by sections to generate a tomogram (three-dimensional reconstruction) of the image. This technique gives the highest three-dimensional resolution of imaged objects (Gan & Jensen, 2012).

In Cryo-EM, liquid nitrogen is used to keep samples that are analysed at cryogenic temperature. Electron beams are directed on a specimen and these are either scattered or allowed to pass through the specimen thereby giving details of its structural arrangement (Kühlbrandt, 2014).

Nuclear Magnetic resonance (NMR)

NMR is useful as it can be performed under physiological conditions and allows the three-dimensional structure of protein and nucleic acids could be obtained at atomic resolution (Lindon *et al.*, 2000).

Although it is difficult to give the structure of large RNAs (≥ 50 kDa) by NMR alone, NMR has been used to effectively determine the RNA structures of more than 40% of the total RNA structures stored in the nucleic acid database (Gong *et al.*, 2015). One example is part of the 5’ terminal clover-leaf structure in rhinoviruses (Headey *et al.*, 2007). The rhinovirus 14 cre structure was also determined in this way (Thiviyanathan *et al.*, 2004). It would be useful to combine the use of EM and solution NMR to effectively refine the structures of RNAs.

X-ray diffraction analysis

The biological function of an RNA is directly linked to its three-dimensional structure. X-ray diffraction analysis can be used to determine the structure of RNA molecules, to show

molecules that are identical as well as to show changes that were brought about due to biochemical activities in the RNA molecules (Ke & Doudna, 2004).

X-ray crystallography identifies molecular and atomic structures of crystals when a beam of incident X-rays diffracts into different specific directions, and then, the angles and intensities of the diffraction are measured to produce a three-dimensional picture of its electron density (Bird, 1997). The limitation to this method is that crystals of large RNA molecules are less likely to be arranged in an orderly manner because for effective X-ray diffraction analysis, the crystal should be of a pure composition and a regular structure without any imperfections (Ferre-D'Amare & Doudna, 2001).

1.4.2 Prediction of RNA structures

1.4.2.1 Thermodynamic method

The information encoded by a nucleic acid sequence determines its function in relationship to its structure (Hofacker, 2003). It is more difficult to determine the function of the RNA structures experimentally than to predict the encoded structures in the first instance (Zuker *et al.*, 1991).

Free energy minimization is used to restrict the number of secondary structures that is considered in a system of sequences, hence, an algorithm which relies on this measure to generate a range of possible RNA folding of all base pairs close to the minimum free energy was developed. This Zuker software is known as; "MFOLD" (Zuker *et al.*, 1991) where "M" denotes "Multiple" (Zuker, 2003). The MFOLD software used for folding RNA has been in use since the late 1980's (Zuker, 1989). One potential problem with this

approach is that the correct structure may not be the one predicted to have the minimum free energy. However, the program can give a number of predicted structures and each of them could be analysed by other methods e.g. phylogenic conservations.

Another software, known as “The Vienna RNA secondary structure server” predicts RNA secondary structure from a single sequence, a set of aligned sequence and finally, it designs sequences that would give a particular already defined structure. There are therefore three functions; RNAfold, RNAalifold and RNAinverse command line programs respectively (Hofacker, 2003). The use of more than one sequence is potentially useful as key structures may tend to be conserved in related molecules.

1.4.2.2 Phylogenetic analysis and mutational analysis

DNA and RNA sequence of existing structures can be aligned to acquire information which shows their similarities and differences. The information obtained from the analysis could be used to infer possible DNA or RNA structures in the sequence and this relationship could be expressed in a phylogenetic tree (Katoh & Standley, 2013; Somarowthu *et al.*, 2015).

Covariance is a phenomenon which is also used to predict DNA or RNA structures. Here it is assumed that a functional structure is present in all closely-related viruses. When two bases that are predicted to base pair are found to co-vary in a specific sequence, this is considered to be evidence that the two bases do participate in a structure e.g. a G-C base pair in sequence could be replaced by A-U/T or there may be a single difference, G-U and vice versa (Al-sunaidi *et al.*, 2007; Williams *et al.*, 2009). Figure 1.8 shows a diagram

illustrating covariance/mutation in a structure found in the $3D^{pol}$-encoding region of HPeV. The structure is made up of three stem-loops and a tertiary structure (kissing) interaction between two of the loops (Williams *et al.*, 2009).

Mutational analysis has also been a very useful tool in recent studies to predict the importance of an RNA structure in the genome (Mirmomeni *et al.*, 1997). Here nucleotides predicted to form a base pair are mutated. Mutation of one nucleotide should weaken the structure and give a virus with a particular phenotype. A paired change allowing a different base pair may give a normal phenotype. Obtaining a revertant or pseudo-revertant with a normal phenotype and with a restored base-pair is also evidence that the predicted base-pair exists.

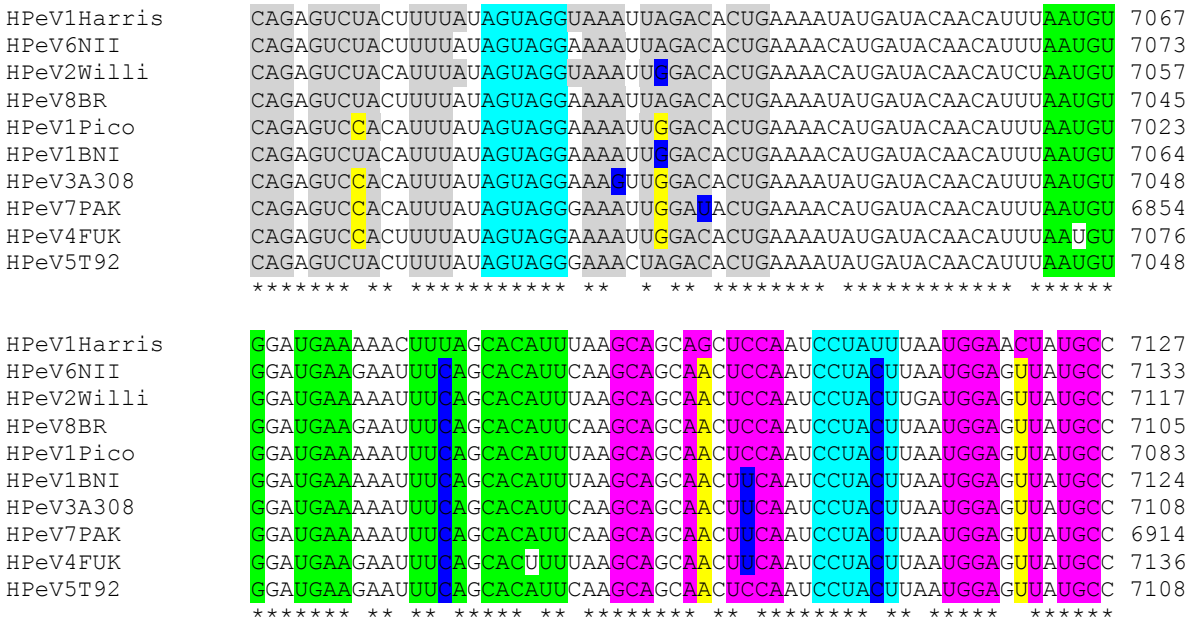


Figure 1.8: Alignment of the 3D structure in 10 completely sequenced strains of HPeVs. The three stem-loops are highlighted in grey, green and pink with the tertiary structure (kissing) interaction between two of the loops highlighted in light blue. The dark blue designates a difference in only one nucleotide of a base pair whose structure is still maintained via a G-C/U interaction. Yellow colour shows an interaction when both nucleotides of a base pair are different (covariant) while the unshaded regions indicate where there is no interaction between nucleotides.

1.5 General Overview of RNA Structures in Viral Genomes

Over a couple of decades, studies have shown that some RNA could form structures that could have significant biological functions. It is now an advantage to study viral RNA molecules having very small sizes to know the molecular structure, what molecules they interact with and ultimately discovering the function of an RNA molecular structure (Witwer *et al.*, 2001).

Functional RNA secondary structures could be very difficult to identify, as the structures could not be far much different from structures that are formed by random RNA sequence (Hofacker *et al.*, 2004). Very few functionally relevant RNA structures could be found in the whole viral genome. They could be detected using computer software such as mfold (Zuker, 2003) or co-variation analysis. They occur mostly at the 5' UTR and 3'UTR, however, it is possible to find also find them in the ORF (Hofacker *et al.*, 2004). It is possible for sequences however, to contain common structural motifs but very different structures and it is difficult to determine the exact function of conserved RNA structures, however, deletion studies can be used to predict its function (Witwer *et al.*, 2001).

1.5.1 **Phages**

Leviviridae are ssRNA phages without an envelope and tail but possessing two structural virion proteins. The genome length ranges between 3466–4276 nt depending on the type of strain. The replication cycle does not include a DNA stage. There are four and sometimes three partly overlapping genes in *Leviviridae* however, there is a remarkably short GC-rich hairpin followed by an unpaired GGG sequence in the 5' UTR. The GGG

sequence is believed to be the target for replication enzyme and thus needed for the replication of the virus (Hofacker *et al.*, 2004).

1.5.2 Negative stranded RNA Viruses

Single panhandle structures have been identified in most negative stranded RNA viruses. Some of the families include; *Paramyxoviridae* (genome size is about 15-16 Kb), *Bunyaviridea*, *Orthomyxoviridae* and *Arenaviridae*. Some other families; *Rhabdovirdae* (The genome is made up of about 12 kb) and *Filoviridae* do not have a specific recognised panhandle structure, however, there are some conserved structural features within the genome and these are also present in the families which have a panhandle structure (Hofacker *et al.*, 2004).

1.5.3 Positive Strand RNA Viruses

Dicistroviridae is a family of viruses that contain two IRES regions; one located at the 5' UTR and the other at an untranslated intergenic region (IGR) made up of about 190 nt. The IGR-IRES is sandwiched between 2 ORF's and it is also structurally well conserved (Jang & Jan, 2010). This IGR-IRES is smaller than most other IRESs found to date and can bind both 40S ribosomal subunits and 80s ribosomes. It occupies the P-site of the ribosome and allows a non-AUG codon to occupy the A-site. In all the *Dicistroviridae*, the IGR-IRES is mainly made up of 3 pseudoknots.

A conserved structure, having a high number of compensatory mutations, in three genera; *Comovirus*, *Fabavirus* and *Nepovirus* of the plant viruses *Comoviridae* shows a great resemblance to the cre structure of the *Picornaviridae* and it is a well conserved stem with

a large A/U-rich loop area. Potyvirus are also mainly plant pathogens and have a genome of about 10 kb. They have not been studied extensively because of the high diversity of sequences, however, conserved structures have been identified in the 3' UTR (Hofacker *et al.*, 2004).

Cardiovirus 3' UTR is made up of 3 (I, II and III) conserved stem-loop motifs and the deletion of III, II and I could lead to either death, inability to recover infectious virus from RNA transfected to cells and no effect on viral growth respectively. It has also been reported that part of the 3' UTR interacts with the Poly (A) tail to aid viral RNA synthesis. Also, some of the structures in the 3' UTR of Enterovirus are non-conserved within the genus. The 3' UTR of Poliovirus, *Enterovirus C* and *Enterovirus A* are conserved, however, that of *Enterovirus B* is not conserved (Witwer *et al.*, 2001).

1.5.4 ***Retroviridae***

Retroviridae show a large number of functional RNA secondary structures, however, there are little similarities in the motifs used by different genera. It is difficult to compare between the genera because of poor alignments due to the low similarity of the sequences. The *Lentivirus* HIV, is very well characterized both experimentally and computationally in terms of RNA structures. Such structures include the TAR hairpin at the 5' end of the genome which binds to the tat protein and regulates the virus promoter (Bannwarth & Gatignol, 2005), the gag/pol frameshift hairpin (also present in Mammalian Type B Retrovirus) which is involved in the regulation of read through by ribosomes into the pol region by allowing a frame-shift and the RRE structure that regulates the nuclear export of HIV RNAs (Hofacker *et al.*, 2004).

1.6 RNA structures in *Picornaviridae*

1.6.1 Non-coding regions

The genome of RNA viruses is capable of forming secondary structures. There are a number of predicted structures in the 5' UTR and 3' UTR of positive sense RNA viral genomes that have role to play in translation, replication (Mirmomeni *et al.*, 1997), transcription and encapsidation (Sasaki & Taniguchi, 2003a; Liu *et al.*, 2009a). In some cases, these allow interactions between the genome and host cell proteins to bring about replication (Lin *et al.*, 2009).

1.6.1.1 3' UTR

Although studies by deletion mutagenesis show that the 3' UTR is not necessary for virus replication, the 3' UTR of studied picornaviruses have RNA secondary and tertiary structures which seem to be necessary for efficient synthesis of new RNA (Todd & Semler, 1996; Mirmomeni *et al.*, 1997).

The 3' UTR of the *Parechovirus B* (LV) has two stem loop domains; I and II (Tolf *et al.*, 2009) and the HPeVs consists of highly conserved repeats of the sequence; AUUAGAACACUAAUUUG arranged in tandem. Unlike the LV, the 3' UTR of HPeVs is predicted to be made up of a single stem loop in which its U-rich 5' part extensively base pairs with the poly (A) tail (Al-sunaidi *et al.*, 2007).

A highly conserved structure was discovered in the 3' UTR of three different families of +RNA viruses; *Astroviridae*, *Coronaviridae* and *Picornavididae* (*Erbovirus* genus)

suggesting a common ancestral origin and thus usefulness of the structure (Jonassen *et al.*, 1998a; Robertson *et al.*, 2004).

1.6.1.2 5’ UTR

The 5’ UTR of picornaviruses contains a large (usually around 400 nt) structure, the IRES (Pelletier & Sonenberg, 1988). This is made up of several secondary structures necessary for cap-independent translation initiation (Figure 1.9). 5 classes of IRES (I-V) are found in picornaviruses with members of each genus usually containing a specific type (Liu *et al.*, 2009a; Sweeney *et al.*, 2012). Enteroviruses have the Type I IRES, aphthoviruses and cardioviruses have the type II IRES, hepatoviruses have the type III IRES while teschovirus have the type IV (Nateri *et al.*, 2002; Lin *et al.*, 2009). Type V IRESs are seen in most kobuviruses (Sweeney *et al.*, 2012).

The 5’ UTR is made up of several domains and the key IRES domains are usually close to the AUG which initiates the open-reading frame, at the 3’ end of the 5’UTR. For instance the 5’ UTR of HPeV1 has stem-loops A-L (Ghazi *et al.*, 1998) and the important domains for IRES activity are I, J and K (Nateri *et al.*, 2002). There are some similarities between the different IRES types, particularly I, II and V. These all have a GNRA tetraloop motif which is necessary for function (Nateri *et al.*, 2002) and the structure of the stem with the tetraloop is similar. There is also some conservation of nucleotide sequence between type II and V IRESs in one of the stems (boxed in red in Figure 1.9, (Sweeney *et al.*, 2012). The type I-III IRESs have a short polypyrimidine motif (8-12 nucleotides) a gap of about 18 nucleotides then an AUG. In type II and III IRESs this AUG is the one that initiates the open-reading frame and ribosomes are probably directly recruited to this site. In the type I

IRES, this site recruits ribosomes, but these then scan to the AUG which does initiate the open-reading frame, which is up to 140 nucleotides downstream (Belsham, 2009).

The IRES plays a key role in the life cycle of the virus. In addition, mutations in the 5'UTR can be important determinants of disease. The best-studied example is the oral poliovirus vaccine, made up of attenuated strains of each of the three poliovirus types. In all three strains a mutation in the 5' UTR, weakens one of the secondary structures making up the IRES and this affects the efficiency of translation in a cell type-dependent way. Reversion of this mutation is part of the pathway giving virulent viruses in vaccine recipients, which can lead to disease (Macadam *et al.*, 1994).

The 5' UTR also has other RNA structures. The 5' UTR in *Enterovirus* has a stable clover leaf-like structure domain at the 5' end. The domain is used as a cis-acting element for RNA replication which is formed from an RNA/protein complex between the virus and cellular proteins (Herold & Andino, 2000). Several picornaviruses have a 5' terminal stem-loop and this seems to be also involved in RNA replication (Nateri *et al.*, 2002). The 5' terminal stem-loop has also been reported to be needed for RNA encapsidation in Aichivirus (Sasaki & Taniguchi, 2003a).

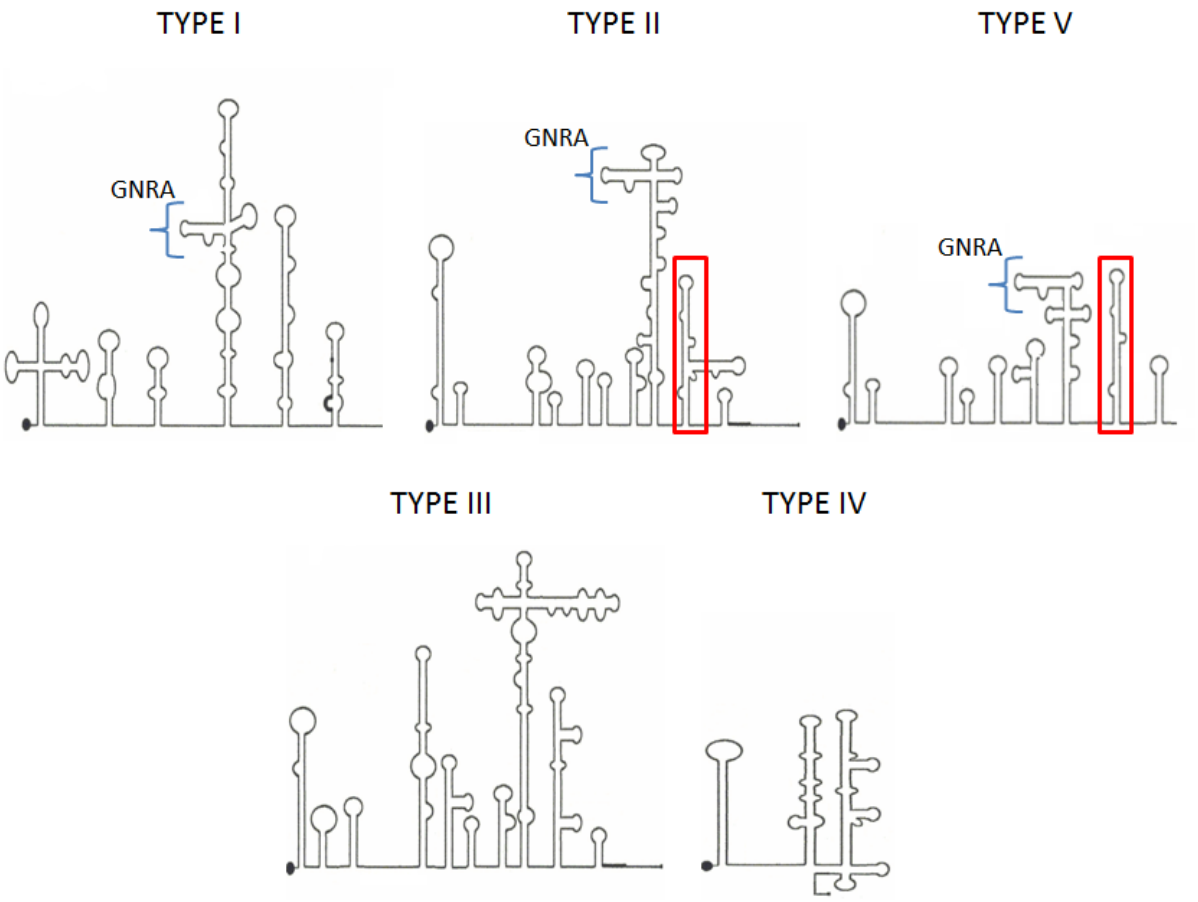


Figure 1.9: Schematic diagram of the 5'UTR structures of picornaviruses with different IRES types, I-V (Sweeney *et al*., 2012). Types 1, 2 and 5 have a GNRA tetraloop which is important for function (Nateri *et al*., 2000). The red boxes highlight a conserved structure and sequences in IRES types II and V. A polypyrimidine tract in IRES types I-III is shown as a thicker line and a downstream AUG is shown as a dot. Image provided by Professor Glyn Stanway.

1.6.2 **Cis-acting replication element (cre)**

In addition to the RNA structures found in the 5' and 3'UTRs, it has become clear that functional RNA structures can also be found within the polyprotein-encoding region of picornaviruses. In picornaviruses, the first to be discovered was the cre (McKnight & Lemon, 1998). This was identified as a structure within the coding region (VP1-encoding) that was needed for virus replication. Similar structures needed for virus replication were then found in cardioviruses (Lobert *et al.*, 1999) in the region encoding VP2 and enteroviruses in the region encoding 2C (Goodfellow *et al.*, 2000). As these were all located in different regions of the genome, it was clear that the cre does not have a fixed location in picornaviruses and this has been confirmed by other studies. Other known and predicted cre locations are 2A and VP2 in RV-A and RV-B (Cordey *et al.*, 2008), 3D in hepatitis A (Yang *et al.*, 2008) and VP0 and VPg in the species *Human parechovirus* and *Ljungan virus* (Al-sunaidi *et al.*, 2007). Unusually, the cre in FMDV seems to be in the 5' UTR (Mason *et al.*, 2002).

They are important in key processes in the viral RNA replication. Some depend mainly on the sequence while others depend mainly on the structure (Liu *et al.*, 2009a).

The function of the cre is to act as a template for $3D^{Pol}$ to add two uridine residues to the position 3 tyrosine residue of VPg to give $VPgpUpU_{OH}$. Uridylation activates the replication process by acting as a primer for synthesis of new genomes via -RNA (Pathak *et al.*, 2007; Shen *et al.*, 2007; Steil & Barton, 2009). The cre in different picornaviruses are diverse in sequence but all single stem-loops. In the loop, there are 2A residues involved in the template actions and these are after, but not always, in the larger motif CAAAC (Al-sunaidi *et al.*, 2007).

1.6.3 **Other conserved elements**

In addition to the cre, there is growing evidence that the coding region of RNA viruses can contain important RNA elements. For example, two important functionally redundant RNA structures were also identified in the region encoding the RNA polymerase of hepatitis C virus. The structures function independently and one has been recorded to be involved in two different RNA/RNA interactions (Diviney *et al.*, 2008).

Using alignments and thermodynamic folding predictions, several elements were predicted in some of the picornaviruses known at the time (Witwer *et al.*, 2001). In addition to the cre, they found 5 conserved elements in *Cardiovirus*, 1 in *Parechovirus*, 6 in *Teschovirus*, 3 in *Hepatovirus*, 2 in *Enterovirus*, 25 in *Aphthovirus*, 1 in Rhinovirus. There were apparently other secondary structures in RV-A and RV-B that are not necessarily conserved between the two species, 4 in RV-A and one of these is also conserved in RV-B (Witwer *et al.*, 2001). The authors did not speculate about the large number of conserved structures in Aphthovirus but this could presumably be related to the high GC content of these viruses compared to most picornaviruses (Stanway, 1990). Several of the predicted conserved structures are in the 3D-encoding region of picornaviruses. This may be due to the encoded protein being relatively highly conserved, but it has also been observed that this part of the genome shows restricted codon variability in many picornaviruses (Hershan, 2012). Such conserved sequences can be seen in the coding region of other RNA viruses.

Some members of *Enterovirus C* including all the 3 types of poliovirus and the coxsackie A viruses, excluding CAV1, CAV19 and CAV22 possess a phylogenetically conserved structure in the region encoding 3C. This structure was reported to have prevented the

cleavage of their RNA by RNase L, an enzyme that can cleave viral RNA after the dinucleotides UA and UU in infected cells. RNase L was reported to cleave the RNA of the other *Enterovirus* species because the conserved structure was absent in these (Han *et al.*, 2007).

Two functional RNA structure elements in the coding region of poliovirus (PV) were recognised by the use of synonymous codon mutation in large segments of the genome. In this case, the PV could still replicate but with reduced efficiency, and this was found to be due to changes in the RNA structure elements due to the mutations caused by the introduction of the different synonymous codons. The RNA structures α and β, were also identified, located in the 3D-encoding sequence, and are about 75 nt long separated by about 150 nt. The functional domain of α was narrowed down to a 48 nt long highly conserved segment while β is a stable 37 nt bases that is a highly conserved hairpin in all *Enterovirus* genus. The exact function of α and β was not established (Song *et al.*, 2012). Another group independently identified an overlapping part of the 3D-encoding region to be important in poliovirus replication by a similar synonymous codon mutation strategy (Burrill *et al.*, 2013).

An analysis of the 3D-encoding region in the HPeV genome showed that some codons are perfectly or almost perfectly conserved (low synonymous codon variation) suggesting that the RNA sequence is important, not just the sequence of the encoded protein. It was suggested that this region folds into an RNA secondary structure made up of three stem-loops and a tertiary structure “kissing” interaction (Williams *et al.*, 2009). Figure 1.10 shows the conserved structure in the $3D^{pol}$-encoding region of HPeV1.

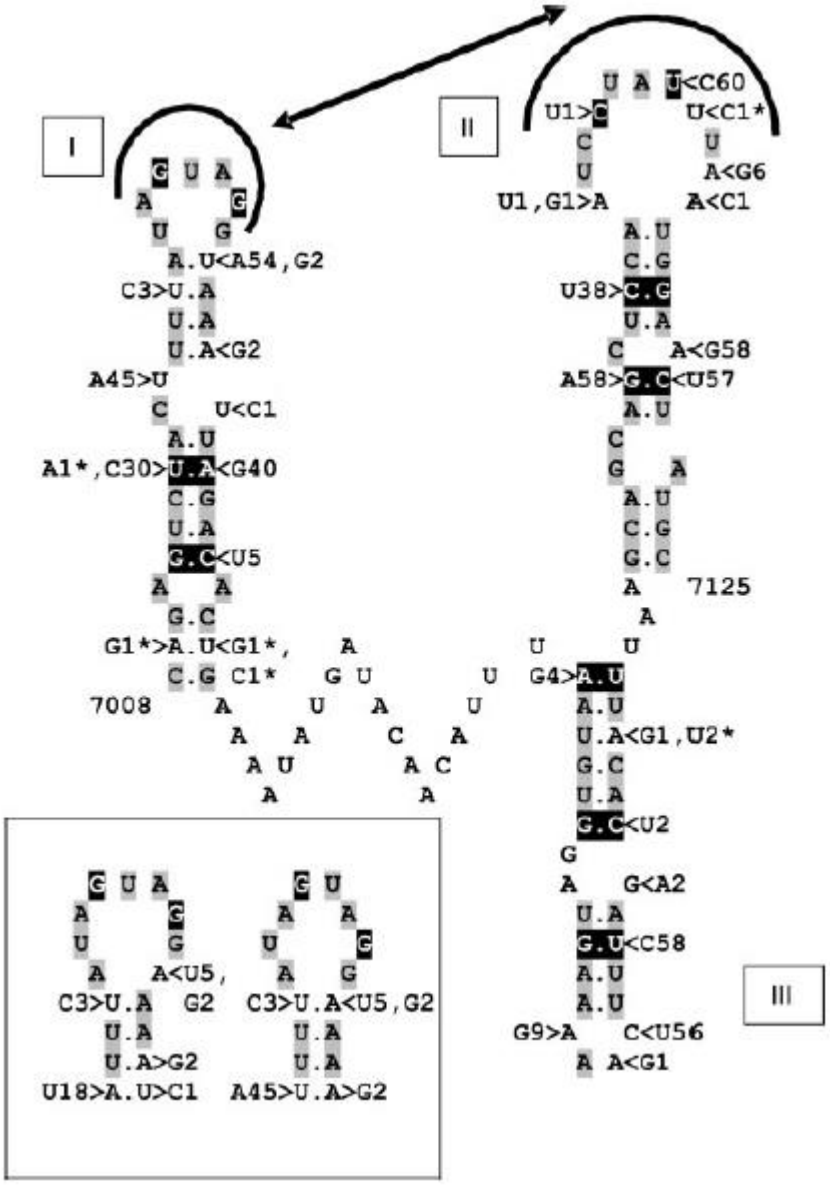


Figure 1.10: Predicted RNA structure in the 3D encoding region of HPeV1 Harris strain (Williams *et al* 2009). The structure is found at positon 7008-7126 and contains 119 nt. It was identified using mfold and also covariance analysis on 61 strains. The structure is made up of three stem loops with a tertiary structure kissing interaction between two of the loops shown by the arrow. The areas shown in white and black are covariant while the gray areas are conserved in all the HPeVs. Nucleotides followed by numbers e.g. A45, is the number of strain where that nucleotide is found rather than the nucleotide in HPeV1 Harris. Some non-covariant nucleotides are indicated with an asterisk (*). The top of domain I could have slightly different forms in different isolates as shown in the panel.

1.7 Aims of present study

HPeVs are being increasingly recognised as important human pathogens but have been little studied compared to enteroviruses and there are no drugs or vaccines to control these viruses.

The conserved structure identified in the 3D-encoding region of HPeVs have not yet been studied to a large extent. The aim of this work was to increase the knowledge and understanding of how this structure works to form the basis for further research in the future and even the development of antiviral agents or vaccines.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Agarose gel electrophoresis

- Molecular biology grade agarose (Fisher)
- Clear loading dye (5 X):

  100 ml of clear loading dye was made by mixing 50 ml Glycerol (Fisher), 200 µl 0.5 M EDTA pH 8 (Fisher), and 49.8 ml dH_2O. The ratio of the loading dye to DNA added was 1:5.

- ELFO buffer (50 X):

  50 X stock of ELFO running buffer was made with 242 g Tris (Fisher) and 100 ml 0.5 M EDTA, pH 8 (Fisher). The pH was normalized to 7.7 using Glacial Acetic Acid (Fisher) and then topped up to 1 l with dH_2O. The final working solution (1 X) was made by adding 100 ml 50 X ELFO buffer to 4900 ml dH_2O.

- GeneRuler 1Kb DNA ladder (0.5 µg/µl, Thermo Scientific):

  The loading mixture was prepared by adding 1 µl of the DNA ladder to 1 µl 6X DNA loading dye and 4 µl nuclease free H_2O (Thermo Scientific). The DNA ladder used is shown in Figure 2.1.

- Safe view nucleic acid stain (NBS biological)

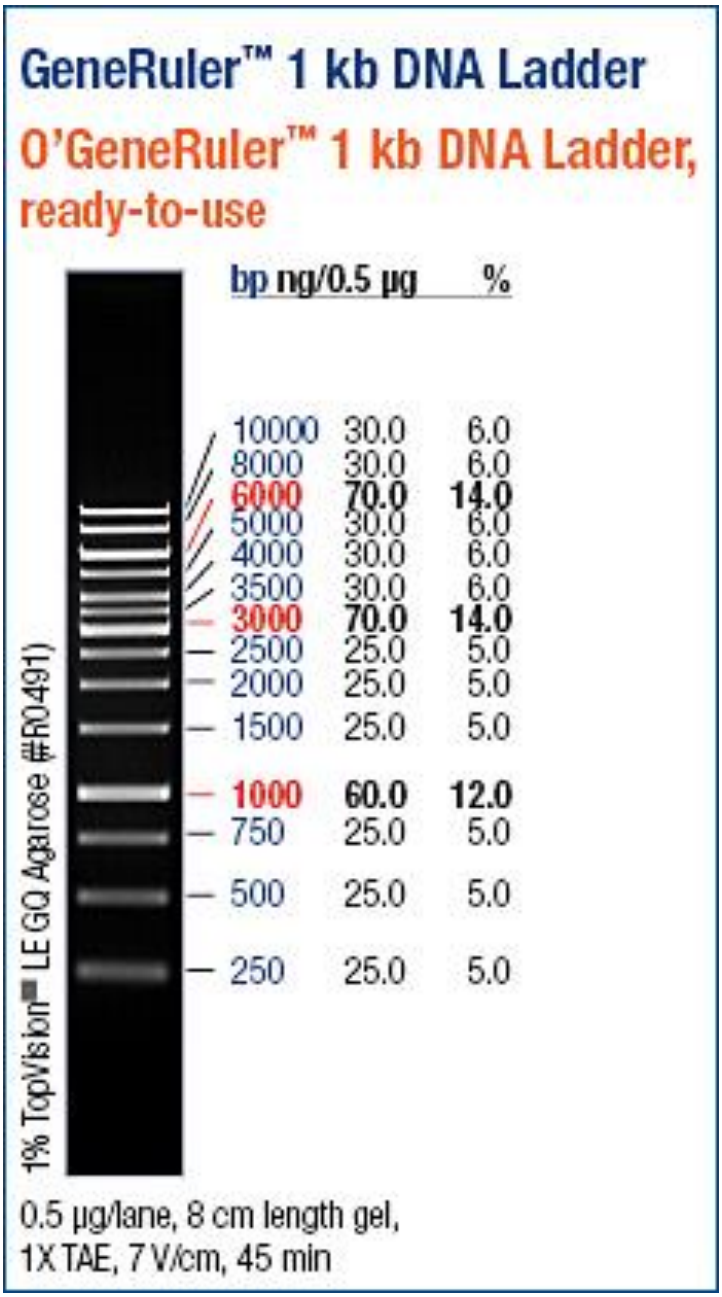


Figure 2.1: 1 Kb DNA ladder showing fragment sizes and amount of DNA (ng) in each band. 0.5 µg/µl of the Thermo Scientific ladder was loaded.

2.1.2 Bacterial cell culture media

- Luria broth (LB):

  Lennox L-Broth (Melford Laboratories) was prepared by adding 20 g of the powder per 1 l dH_2O and autoclaving at 121^oC for 15 min to give a final solution containing Tryptone (10 g/l), yeast (5 g/l), sodium chloride (5 g/l) and Tris-HCl buffer (pH 8.5) (1.5 g/l). This was used to grow *Escherichia coli* (*E. coli*) bacteria.

- LB agar:

  100 ml LB agar was made by adding 1.48 g Oxoid Bacteriological agar (Fisher) to 100 ml LB.

- Antibiotics:

  - A stock solution of 100 mg/ml Ampicillin (Sigma), 50 mg/ml Kanamycin (Sigma) and 10 mg/ml Tetracycline (Sigma) were prepared by dissolving the antibiotics in sterile dH_2O.
  - 5-Bromo-4-Chloro-3-Indolyl-β-D-Galactoside (X-Gal) (Promega) (80 µg/ml):

    The working solution was obtained by dissolving in dimethyl formamide (Sigma) and then covered with aluminum foil and stored at -20^oC in 1 ml aliquots.
  - Isopropyl-β-D-thiogalactopyranoside (IPTG) (Promega):

    A 0.5 mM solution was prepared in dH_2O. This was stored at -20^oC in aliquots of 1 ml.

2.1.3 **Bacterial strain**

The Top 10 strain of *E. coli* which has the genotype: F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G (Invitrogen) was used at its exponential phase of growth to propagate recombinant DNA.

2.1.4 **Cloning vectors**

CT-GFP Fusion Topo TA expression kit (Invitrogen), pBSK (Epoch), pGEM-T Easy Vector System (Promega), pEGFP-C1 (Clontech), pmCherry-C1 (Clontech) and pGL4.73 [hRluc/SV40] (Promega) were used to engineer constructs in this research.

2.1.5 **Ligation reagent**

T4 DNA ligase provided with 10 X reaction buffer (Promega)

2.1.6 ***In vitro* transcription**

RNase inhibitor, 40 U/μl (Fermentas)

Ribonucleotide Solution Set (NEW ENGLAND BioLabs)

T7 RNA polymerase provided with 100 mM DTT and transcription optimized 5 X buffer (Promega)

2.1.7 **Mammalian cell culture**

- Accutase (Sigma)
- Foetal bovine serum FBS- research grade (Sigma)
- 50 mg/ml Gentamicin (PAA)
- Non-Essential Amino-acids 100 X (PAA)
- Penicillin/streptomycin 100 X (Sigma)
- Trypsin-EDTA (0.25% 1 X, Gibco)
- Dulbecco's modified Eagle's medium (DMEM) (Sigma):

  DMEM was supplemented with 10 % FBS, 1 ml of gentamicin solution or 5 ml of penicillin/streptomycin solution and 5 ml of Non-Essential Amino-acids to give DMEM growth medium.
- McCoy (Gibco):

  The McCoy medium was supplemented with 10 % FBS and 5 ml of penicillin/streptomycin solution to give McCoy growth medium.
- Roswell park memorial institute (RPMI) (Sigma):

  The RPMI was supplemented with 10 % FBS and 5 ml penicillin/streptomycin solution to give RPMI growth medium.
- Cell lines:

  The cell lines used are summarized in Table 2.1

Table 2.1. Names, origins and suppliers of the cell lines used for the experiments in this report including the summary of methods of maintenance of the cells.

Cell line	Name / Origin	Supplier	Flask type	Digestion Enzyme	Growth Medium
A-549	Human lung carcinoma	Prof. Glyn Stanway	Non-filtered	Trypsin	DMEM
CAMA-1	Pleural effusion of human breast cancer	Prof. Nelson Fernández	Filtered	Accutase	RPMI
GMK	Green monkey kidney	Dr. Merja Roivainen, Finland	Non-filtered	Trypsin	DMEM
HT-29	Human colon adenocarcinoma	Dr. Andrea Mohr	Filtered	Trypsin	McCoy
MCF-7	Michigan Cancer Foundation-breast cancer	Prof. Nelson Fernández	Filtered	Accutase	DMEM
MD-MBA-231	M.D. Anderson - metastatic breast cancer	Prof. Nelson Fernández	Filtered	Trypsin	DMEM
PC3	Prostate cancer	Prof. Elena Klenova	Filtered	Accutase	RPMI
Vero	African green monkey kidney	Prof. Glyn Stanway	Non-filtered	Trypsin	DMEM

2.1.8 **Mammalian cell over-lay**

- Carboxymethyl cellulose (CMC) (Sigma):

  A 2% CMC gel was made by mixing 2 g CMC and 100 ml dH_2O and then autoclaving.

- CMC-Agarose:

  2 g CMC and 1 g Agarose (Fisher) were mixed 100 ml dH_2O and autoclaving.

CMC and CMC-Agarose were melted in a microwave or boiling water-bath and then mixed with mammalian cell growth media (in a water-bath with temperature maintained at 50^oC) at a 1:3 ratio.

2.1.9 **Mammalian cell preparation for bioimaging**

- Hoechst stain (33342) (Invitrogen) was provided as a 16.2 mM solution.
- Mowiol mounting medium:

  6 ml of dH_2O was added to a stirred mixture of 2.4 g mowiol (CalBiochem) and 6 g glycerol (Fisher). This was incubated at room temperature (RT) for 2 hr. 12 ml of 2 M Tris-HCl buffer (pH 8.5) was then added and incubated in a water bath between 50/60^oC for 10 min until the mowiol dissolved. This media was centrifuged at 5000 rpm for 15 min to remove any undissolved solid and the media was then aliquoted to 2 ml Eppendorf tubes wrapped in aluminum foil and stored at -20^oC.

2.1.10 **Oligonucleotides**

Designed PCR primers were synthesized commercially by Invitrogen (Purity-desalt, scale of synthesis: 50 nmole) and were dissolved in water to give a final working concentration of 100 µM unless otherwise stated. The oligonucleotides used are shown in Tables 2.2 – 2.5.

2.1.11 **Plasmid DNA preparation**

- Miniprep kit (Qiagen): Small scale preparation of DNA
- Midiprep kit (Qiagen): Large scale preparation of DNA
- Isolation of DNA from agarose gels:

  Gel extraction kit (Qiagen) (DNA isolation from Agarose gel electrophoresis)

2.1.12 **PCR reagents**

- dNTP Mix (10 mM, Thermo Scientific)
- *Pfu* DNA Polymerase (Thermo Scientific)
- *Taq* DNA Polymerase (recombinant) (Thermo Scientific)

Table 2.2: Oligonucleotides used to generate DNA fragments that were ligated to Topo vector.
The underlined nucleotides are XhoI and BamHI restriction sites added to enable insertion into the vector. In fusion oligonucleotides, the dsRed sequence is highlighted in red.

Primer No.	Direction	Origin	Position	Sequence (5' to 3')
OL1749	Forward	HPeV1 (5' UTR)	43-65	GAAGCTTAAACCCCGACTTGCTGAGCTTCTCT
OL2093	Reverse	dsRed/ HPeV1 ($3D^{pol}$)	-/ 6782-6796	GTAGTGTAAATACAC/ TCCGCGGTTACAGGAACAGGT
OL2092	Forward	dsRed/ HPeV1 ($3D^{pol}$)	-/ 6782-6812	TCCTGTAACCGCGGA/ GTGTATTTACACTACAAATTTAATCAGTCCT
OL2064	Reverse	HPeV1 ($3D^{pol}$)	7314-7339	GGATCCGTCATGTCCAATGTTCCAAGTTAGTG
OL2110	Reverse	HPeV1 ($3D^{Pol}$)	7229-7250	ATCAAACACCATGGGCATCAAC
OL2108	Forward (sequencing primer)	HPeV1 (5' UTR)	404-427	GCCAGCGGAACAACATCTGGTAAC
OL2109	Reverse (sequencing primer)	dsRed	-	GGGCCTTGTGGATCTCGCCCT
OL2242	Forward (sequencing primer)	dsRed	-	GAAGTTCATCGGCGTGAACTTCCCCT
OL1855	Forward (sequencing primer)	GFP TopoVector	-	CGACACAATCTGCCCTTTCG
OL1854	Reverse (sequencing primer)	GFP TopoVector	-	GGGTAAGCTTTCCGTATGTAGC

Table 2.3: Oligonucleotides used to generate and analyse DNA fragments that were ligated to the C-terminus of cut pEGFP-C1 vector. The underlined nucleotides are complete XhoI and BamHI restriction sites or overhangs added to enable insertion into the vector.

Primer No.	Direction	Origin	Position	Sequence (5' to 3')
OL2065	Forward	Random sequence	Nil	TCGAGCTTAACGGTAGGCTGTCC
OL2066	Reverse	Random sequence	Nil	GATCGGACAGCCTACCGTTAAGC
OL2063	Forward	HPeV1	6823-6846	CTCGAGCTTAAGTCTACCAATCGTATATGGCGATG
OL2064	Reverse	HPeV1	7314-7339	GGATCCGTCATGTCCAATGTTCCAAGTTAGTG
OL2154	Reverse	HPeV1	7228-7249	GGATCCATCAAACACCATGGGCATCAAC
OL2151	Forward	HPeV1	7250-7284	CTCGAGCTTAATAAGATTAATGTTTTGTTTTTCTTTTGCTATGGAC
OL2155	Forward	HPeV1	7008-7030	CTCGAGCTTAACAGAGTCTACTTTTATAGTAGGT
OL2156	Reverse	HPeV1	7096-7126	GGATCCGCATAGTTCCATTAAATAGGATTGGAGCTGC
OL2149	Forward	HPeV1	2351-2376	CTCGAGCTTAACAGATGGATTTAACCGACCCTCTTTG
OL2150	Reverse	HPeV1	2843-2868	GGATCCAGACTATTGTTATCTCTGACAGTTCT
OL2183	Forward	HPeV1	1869-1900	ACTCGAGCTTAAGATATTTCAAGTGGTCAGCAACAACAGCCCCCC
OL2184	Forward	CAV9	6638-6666	ACTCGAGCTTAACGGTCACCTCATAGCCTTTGATTACTCTG
OL2185	Reverse	CAV9	7131-7155	AGGATCCGTATGTCTTTCATAGGCATAACAGG
OL2186	Forward	CAV9	7347-7369	ACTCGAGCTTAAGTAAATTAGAGACAATTTGAACTG
OL2187	Reverse	CAV9	7433-7452	AGGATCCCCTCCGCACCGAATGCGGAG
OL2188	Forward	HRV 1B	6072-6100	ACTCGAGCTTAACT ATGTTACAA TGGGCATCAA AAAGAGAGAC
OL2189	Reverse	HRV 1B	6557-6589	AGGATCCAGTCCCAGAACATCCAGAAGGAACTCCACCTGC
OL2190	Forward	HRV 1B	5715-5744	ACTCGAGCTTAA GTCAGA TTAAAATCTC CAAACATGCT AATG
OL2191	Reverse	HRV 1B	6682-6714	AGGATCCTTGTAAGAAAAAATCACATCATCACCATATGCG
OL2192	Forward	HRV 1B	7094-7133	TCGA GCTTAGGTAGATAT AGAAATAATG AATGAATGAT TCTTTAATTC TAT
OL2193	Reverse	HRV 1B	7094-7133	GATCATAGAATTAAAGAATCATTCATTCATTATTTCTATATCTACCTAAGC
OL2196	Forward	HPeV1	6920-6956	G AAGTGACTGG GTCAAGGAAA GATGAACCTC CTTCAC

OL2197	Reverse	HPeV1	6920-6956	GTGAAGGAGGTTCATCTTTCCTTGACCCAGTCACTTC
OL2198	Forward	HPeV1	6969-7010	TG GAGGTTGAAT TCTTAAAGAG GAAACCTGGT TATTTCCCAG
OL2199	Reverse	HPeV1	6969-7010	CTGGGAAATAACCAGGTTTCCTCTTTAAGAATTCAACCTCCA
OL2200	Forward 1	Aichi virus A	8102-8134	TCGAGCTTAATCCATGGTGATATAAAGACCACCCACTTCCTTC
OL2201	Forward 2	Aichi virus A	8135-8156	GGGTGAGCCCCTAAGCCATGGT
OL2202	Reverse 1	Aichi virus A	8121-8156	GATCACCATGGCTTAGGGGCTCACCCGAAGGAAGTGGGTG
OL2203	Reverse 2	Aichi virus A	8102-8120	GTCTTTATATCACCATGGATTAAGC
OL2204	Forward	HPeV1	7008-7034	ACTCGAGCTTAA CAG AAAGCACTTTCATTGTAGGA AAGT
OL2205	Reverse	HPeV1	7096-7126	ACTCGAGCTTAA CAG AAAGCACTTTCATTGTAGGA AAGT
OL1637	Forward (sequencing primer)	EGFP	-	CTGGAGTACAACTACAACAGCCA
OL2027	Reverse (sequencing primer)	EGFP	-	TGATCAGTTATCTAGATCCGGTG
OL1983	Forward (sequencing primer)	mCherry	-	TCCTCCGAGCGGATGTACCCCGAG

* CTCGAG (XhoI) and GGATCC (BamHI) are complete sites. TCGA (XhoI) and GATC (BamHI) are overhang given when pairs of the oligonucleotides are annealed

Table 2.4: Oligonucleotides used to generate and analyse DNA fragments that were used in the production of the Luciferase replicon.

Primer No.	Direction	Origin	Position	Sequence (5' to 3')
OL2259	Forward	HPeV1 (5' UTR)	1-32	GAGCTCTTTGAAAGGGGTCTCCTAGAGAGCTTGGCCGT
OL2244	Reverse	HPeV1 (5' UTR)/ Luc	680-710/ 499-512	ACCTTGGAAGCCAT/ AATGTCAATATTAACAATAAAAGGAAACCAG
OL2245	Forward	HPeV1 (5' UTR)/ Luc	696-710/ 499-524	GTTAATATTGACATT/ ATGGCTTCCAAGGTGTACGACCCCGA
OL2260	Reverse	Luc + NheI + HPeV1 (cre)	1401-1431/ 2316-2322	CAGAACC GCTAGC/ CTGCTCGTTCTTCAGCACGCGCTCCACGAAG
OL2261	Forward	Luc + NheI + HPeV1 (cre)	1419-1431/ 2316-2342	GAAGAACGAGCAG GCTAGC/ GGTTCTGTTGTTACATTCCAGAATTCA
OL2262	Reverse	HPeV1 (*Cre*)	2343-2359	CTGCAG TCCAA CTTAAG TCCATCTGTGAACCCCA
OL2248	Reverse (sequencing primer)	HPeV1 (5' UTR)	192-216	TGCTTTAAGCATCCTTGTTCCACTC
OL2249	Forward (sequencing primer)	Luc	579-603	GAACGTGCTGGACTCCTTCATCAAC
OL2258	Reverse	pHPeV1 Vector	-	GTGAGTTAGCTCACTCATTAGGCAC
OL2247	Forward	T7		GTAATACGACTCACTATAGGGCGA
CreFseqpri	Forward (sequencing primer)	Luc	1226-1248	ACCTTCGGGCCAGCGACGATCTG
CreRseqpri	Reverse (sequencing primer)	HPeV1 (VP3)	2107-2132	AGCATTGTCCAAGGTGGAAGTTGAG
GeneralFcre	Forward	HPeV1 (VP0)	-	CTTCGTGGAGCGCGTGCTGAAGAACGA
GeneralRcre	Reverse	HPeV1 (VP0)	-	AGTCTTAAGTACATCTGTGAACCCCATG
AngF	Forward	HPeV1 (VP0)	-	GTTGCTGACACTTTCTATGTCAAGACCGA
AngR	Reverse	HPeV1 (VP0)	-	TCGGTCTTGACATAGAAAGTGTCAGCAAC
LouF	Forward	HPeV1 (VP0)	-	AGACCGATTCATCCGACTTAGGGCAG
LouR	Reverse	HPeV1 (VP0)	-	CTGCCCTAAGTCGGATGAATCGGTCT

* In fusion oligonucleotides, the grey and yellow highlights show HPeV1 and Luciferase sequence respectively.
* Restriction enzymes are highlighted in purple. Table 2.6 show the names and sequence of the restriction enzymes used.

Table 2.5: Oligonucleotides used to generate DNA fragments containing mutations in the Luciferase replicon.

Primer No.	Direction	Origin	Position	Sequence (5' to 3')
OL2263	Forward	HPeV1 ($3D^{pol}$)	5766-5797	GTATGCATATTGCTGGTAATGGTGAAATGGGA
OL2264	Reverse	HPeV1 ($3D^{pol}$)	6913-6938	TGCGTGACCCAGTCACTTCGGCTCCA
OL2265	Forward	HPeV1 ($3D^{pol}$)	6822-6849	GTCTACCAATCGTATATGGCGATGATGT
OL2266	Reverse	HPeV1 (3' UTR)	7310-7339	GTCATGTCCAATGTTCCAAGTTAGTGTCCT
OL2267	Forward	HPeV1 ($3D^{pol}$)	7224-7248	CTAAGTTGATGCCCATGGTGTTTGA
OL2268	Reverse	pHPeV1 Vector	-	GACTGGAAAGCGGGCAGTGAGCGCA
OL2269	Forward	HPeV1 ($3D^{pol}$)	6830-6858	ATCGTATATGGCGCTGCTGTTATTCTTTC
OL2270	Reverse	HPeV1 ($3D^{pol}$)	6830-6858	GAAAGAATAACAGCAGCGCCATATACGAT
OL2274	Forward (sequencing primer)	HPeV1 (3C)	5628-5653	GAGTGCACCATTCAGGTGGCATCAAG
OL2275	Forward (sequencing primer)	HPeV1 ($3D^{pol}$)	6341-6369	GATGAGTTAAGAAAGCTAGATAAGATTGC

2.1.13 **Reagents for determining cell's stress response**

- tert-Butyl hydrogen peroxide 70 wt % in H_2O (Aldrich):

  A 1 mM stock solution was prepared in DMEM. Final working solutions of 500 µM and 50 µM were also prepared in DMEM.

- Sodium chloride (Fisher):

  A 500 mM stock solution was prepared in DMEM. Final working solutions of 400 mM and 140 mM were also prepared in DMEM.

- Thapsigargin (Sigma):

  Thapsigargin was dissolved in DMSO to give a stock solution of 250 µM. Final working solutions of 1 µM and 0.3 µM were prepared in DMEM.

2.1.14 **Restriction Enzymes**

All restriction enzymes were provided by Themo Scientific. Table 2.6 shows the restriction enzymes used in this research and also the sequence recognized by each of the enzymes.

2.1.15 **Sequencing**

8-10 µl of final products of DNA with concentration ≥100 ng/µl were sequenced commercially by GATC Biotech or Source Bioscience.

2.1.16 **Transfection reagents**

- Opti-MEM I Reduced Serum Medium (1 X), liquid - with GlutaMax (Gibco)

- Lipofectin reagent (Invitrogen)
- Actinomycin-D (Fisher):

  Actinomycin-D was dissolved in DMSO to give a stock solution of 10 mg/ml. Final working solutions of 0.0001 µg/ml – 1000 µg/ml were prepared in DMEM.

2.1.17 **Viruses**

CAV9 (Griggs strain) was recovered from the infectious cDNA clone pCAV9 (Hughes et al., 1994) and propagated in GMK cells.

HPeV1 (Harris strain) sucrose purified virus was kindly provided by Dr. Sisko Tauriainen. Stocks were also made by propagating on HT29 cells.

2.1.18 **Virus cDNA**

- pCAV9 (Hughes *et al.*, 1995)
- pHPeV1 (Nateri *et al.*, 2002)
- pHRV1B (Hughes and Stanway, unpublished)

Table 2.6: Restriction enzymes used

Prototype Restriction enzyme	Isoschizomers	sequence
BspTI	AflII	C^TTAAG
Cfr421	SacII	CCGC^GG
EcoRI	-	G^AATTC
MluI	-	A^CGCGT
Mph1130I	-	ATGCA^T
NheI	-	G^CTAGC
BamHI	-	G^GATCC
XhoI	-	C^TCGAG
SacI	-	GAGCT^C
PstI	-	CTGCA^G

* ^ shows the cutting position

2.1.19 **Luciferase assay:**

Renilla Luciferase Assay System (Promega)

2.1.20 **Cell viability assay:**

Wst 1 (Abcam)

2.1.21 **General purpose materials**

- Absolute ethanol (Fisher):

  70% ethanol (7 volume absolute ethanol + 3 volume dH20)

- Acetic acid glacial (BDH laboratory):

  10 % acetic acid (36 ml dH2O + 4 ml acetic acid)

- Calcium chloride (Sigma):

  A 0.1 M stock solution was made by dissolving 14.7 g in 100 ml of dH_2O and then autoclaving.

- Cell fixation

- Formaldehyde (Fisher):

  2% formaldehyde (2 ml formaldehyde (37%) + 35 ml dH_2O)

- Paraformaldehyde (Fisher):

  4% Paraformaldehyde (4g + 100 ml PBS).

  The powder was weighed in a fume hood and transferred to a conical flask where the correct amount of PBS was added. This was then stirred on a heated stirrer until the boiling mixture became clear. The mixture was left to

cool in the fume hood, aliquoted into 20 ml universal tubes and stored at -20^{o}C. Portions to be used were thawed in the fridge (4^{o}C).

- Crystal Violet (CV) (ACROS ORGANICS):

  CV stain for plaque assay (500 ml) (0.5 g + 495 ml ddH2O + 5 ml absolute ethanol)

  CV stain for CV assay (0.08 %) (32 mg crystal violet + 40 ml dH20)

- Dimethyl sulfoxide (DMSO) (Sigma-Aldrich)
- Iso-propanol (Fisher)
- Water:

  Water used throughout this work was purified by reverse osmosis (dH_2O). For some applications, nuclease-free water was purchased from Thermo Scientific.

- Phosphate buffer saline (PBS) (Fisher)

  1 PBS tablet was dissolved in 100 ml of ddH2O to make 1X PBS. This was then autoclaved before use.

2.2 Methods

2.2.1 Cell Culture

2.2.1.1 Cell propagation

Cells were split at a ratio of 1:4 from cells growing in a 25 cm^2 flask (Table 2.1). Cell splitting was carried out in a level 2 Biosafety Cabinet (triplered LABORATORY TECHNOLOGY) in a level 2 tissue culture laboratory. The medium was aspirated from already growing cells, washed twice with Phosphate-Buffered Saline (PBS), treated with 350 µl trypsin-EDTA or Accutase and then placed on a rocking table till about 80% of cells were detached from the wall of the flask. 20 ml of fresh growth medium was added to the flask containing detached cells then one-quarter of the mixture was deposited into each new 25 cm^2 flask. Split cells were incubated at 37^oC and 5% CO_2 for 3-5 days to allow them to become confluent before being split again to maintain the viability of the cells. For transfection in a 6-well plate, 2 ml of the mixture from the flask being split was deposited into each well of the plate already containing a sterile cover slip for microscopy or no cover slip for flow cytometry. Cell viability assay using the Illuminometer (BMG LABTECH) was carried out in 24-well plate. 0.5 ml of the mixture from the flask being split was deposited into each well of the plate. The 6 and 24-well plates were kept in a sealed polythene bag containing moist tissue paper. A Summary of cell maintenance parameters is shown in Table 2.1.

2.2.1.2 Cell storage by freezing

Confluent cells were washed twice with 1X PBS and then detached using 200 µl Trypsin-EDTA and left on a rocking table for about 80% of cells to detach. 1 ml of FBS was added to the detached cells in the flask, mixed thoroughly and then 1 ml of the mixture (containing

cells and FBS) was transferred to a fresh Cryotube before adding 111 µl of dimethylsulphoxide (DMSO) to give a 10 % mixture. Cells in the Cryotube were wrapped in cotton wool, placed in a storage and kept at -80^{o}C for slow freezing of the cells to take place.

Frozen cells were recovered by thawing in ice and transferring the content in the Cryotube to a fresh flask containing 4 ml of complete growth media. The growth media, now containing dead / unattached cells was aspirated the following day and then fresh media was added to the flask to enable already attached growing cells to continue to grow. This was then left for cells to be confluent before propagation.

2.2.2 Cell biology techniques

2.2.2.1 Transfection

DNA/RNA were transfected into 60% confluent cells in a 6-well plate using Lipofectin (Invitrogen). 100 µl of Opti-MEM media was added to tubes labeled ‘A’ and ‘B’. 5 µg of DNA/RNA was pipetted into tube ‘A’ while 10 µl of Lipofectin was added to tube ‘B’. The contents of tube ‘B’ was added to tube ‘A’ after incubation for 1hr and then left for 20 min at RT. The cells were washed twice with Opti-MEM media and 800 µl of the media with the mixed content of tubes ‘A and B’ were added to each well containing cells. This was then left for 24 hr in an incubator (37^{o}C, 5 % CO_2). The media was changed the following day by adding 1 ml of fresh media and incubated for another 16-24 hr (37^{o}C, 5 % CO_2).

Transfection involving the stress experiment (Chapter 4) was carried out for 6 hrs before the media was changed. The stress element was then introduced 14 hr after the media

was changed and it was left to act on the cells for another 6 hr before cells were harvested and analysed using the FACS.

Cells were also stressed by infection with HPeV1 or CAV9 (2.49 x 10^4 pfu/ml and 1.26 x 10^5 pfu/ml respectively), for 8 hr on transfected cells after the transfection in the cells was expressed for 14 hr prior to the infection.

The concentration of all DNAs used for transfection were measured by NanoDrop spectrophotometry (NanoDrop™ 2000/2000c Spectrophotometers) and brought to a concentration of 380 ±20 ng/ µl by dilution using the formula: ((X/Y)*Z)-Z, where X= initial DNA concentration, Y= Expected final DNA concentration and Z= Volume of initial DNA concentration to be diluted

2.2.2.2 Fixation

GMK cells that were infected after transfection were fixed before FACS analysis to deactivate the virus used for the infection. To retain detached cells, the cells were washed with 200 µl PBS and treated with 200 µl trypsin for about 5 min to detach the cells from the plates. 80-100 % of the cells were harvested by flushing the wells with the corresponding aspirated media and pipetting the products from each labelled well to its corresponding tube. Harvested cells in a microcentrifuge tube were centrifuged at 1300 rpm for 5 min at room temperature. The supernatant was aspirated and the pellets were resuspended in 250 µl of chilled 4% paraformaldehyde. This was left in ice for 30 min for cell fixation to occur.

2.2.2.3 Immunofluorescence staining

The transfected cells were washed twice with PBS. A stock of 1:2000 ratio of 16.2 mM Hoechst stain to media was made and incubated for 10 min at 37^{o}C. 1 ml of the media/Hoechst stain was pipetted to each of the wells in a 6-well plate containing transfected cells and incubated at 37^{o}C for 20 min. Media was aspirated and the plates were then washed twice with PBS. Mounting media was added to a fresh slide and the cover slips containing cells were then placed downwards to make contact with the mounting media and left to dry in the fridge (4^{o}C) for about 10 min before analysis with a BX 41 fluorescence microscope using standard techniques.

2.2.2.4 Flow cytometry analysis with the Fluorescence Activated Cell Sorter (FACS)

After 46±2 hr of transfection in 6-well plates, media was aspirated from each well containing transfected cells and pipetted to individually labeled FACS tubes. To retain detached cells, the cells were washed with 200 µl PBS and treated with 200 µl trypsin for about 5 min to detach the cells from the plates. 80-100 % of the cells were harvested by flushing the wells with the corresponding aspirated media and pipetting the products from each labelled well to its corresponding tube. Analysis with an Accuri FACS machine was carried out immediately after harvest to prevent degradation/ death of live cells.

2.2.3 Annealing complementary oligonucleotides

Complementary oligonucleotides (Table 2.7) giving XhoI and BamHI overhangs necessary for ligation to already cut vectors were incubated for 10 min at 99^{o}C and cooled down to

70^{o}C for 7 min in a PCR machine. The block was then taken out of the machine and left to cool further at room temperature before ligation of the product or storage at -20^{o}C.

Table 2.7: Characteristics of each oligonucleotide used for annealing

Annealed Oligonucleotides	Concentration of oligonucleotides (µM)	Amount of each used (µl)	Construct in EGFP
OL2065 & OL2066	100	15	pEGFPeC
OL2192 & OL2193	10	15	pEGFPHRV(3UTR)
OL2200, OL2201, OL2202 & OL2203	3	10	pEGFPAichi(3UTR)

* The concentration of oligonucleotides and the amount of each used are variable because the table contains the correct concentration and amount of each annealed oligonucleotides that was used for ligation to a cut DNA plasmid. Several attempts to ligate the annealed products that had other different concentrations and amount of the oligonucleotides, to a cut DNA plasmid failed to anneal.

2.2.4 **PCR**

2.2.4.1 PCR mutagenesis

PCR reactions were performed consisting of 35 cycles of denaturing (1 min at 95^{o}C), annealing (1 min at 50^{o}C) and extension (2 min per 1kb at 72^{o}C). Preparation included adding 1X PCR buffer, 0.2 mM dNTP, 10 ng cDNA, 100 nmoles of each primer (except otherwise stated), 2.5 U *Pfu* polymerase and H_2O (Nuclease free) to give a total volume of 50 µl.

There was a final incubation at 72^{o}C for 15 min after the addition of 1.5 µl (5 U/µl) of *Taq* DNA polymerase to samples that were to be cloned in pGEMT-easy vector (because it is a TA cloning vector). PCR products were run on a 1 % agarose gel to confirm accuracy of band formed and then purified using a PCR purification kit (Qiagen), according to the manufacturer's instructions.

2.2.4.2 Colony PCR

Colony PCR was carried out using *Taq* DNA polymerase to screen for the presence of inserts in plasmids in colonies of transformed cells. Each colony was touched with a toothpick which was briefly placed in the PCR mix. The mixture contained 1 X PCR buffer, 3 µl of 25 mM $MgCl_2$, 1 µl (0.2 mM) of dNTP, 100 nmoles of each primer, 1 µl (5 U/µl) of *Taq* polymerase and H_2O (Nuclease free) to give a total volume of 50 µl. The reaction performed consisted of 20 cycles of denaturing (1 min at 95^{o}C), annealing (1 min at 50^{o}C) and extension

(2 min per 1 kb at 72^{o}C). The PCR product was loaded into an agarose gel to find if the insert in the plasmid vector was in the correct orientation.

2.2.4.3 Overlap PCR

Overlap PCR was used to join two pieces of DNA together without the use of restriction enzymes or DNA ligase (Bryksin & Matsumura, 2010). Table 2.8 summarises how overlap PCR was carried out. The method uses a pair of general primers (F1 and R2) to define the ends of the joined fragment and two complementary primers (R1 and F2) containing sequences at the junction to be made.

Table 2.8: Summary of how overlap PCR was carried out

Template	**Primer**	**Product**
cDNA	F1 + R1	Fragment 1
cDNA	F2 + R2	Fragment 2
Fragment 1 + fragment 2	F1 + R2	Joined fragments

2.2.5 Cloning DNA fragments into plasmid vectors

2.2.5.1 Digestion with restriction enzymes

1-5 µg of plasmid DNA was digested with restriction enzymes according to the manufacture's instruction. Single digestion with individual restriction enzyme was carried out with its buffer provided while double digestion was carried by using the information in the "Fermentas Restriction Endonucleases Activity" chart for the five buffers system produced by this company www.fermentas.com/doubledigest. 20 µl reactions were set up for analysis while 50 µl reactions were set up for DNA fragment isolation.

2.2.5.2 Linearization of pEGFP-C1 and pmCherry-C1

pEGFP-C1 and pmCherry vectors were linearized by digestion using XhoI and BamHI by adding 1 µg of vector plasmid, 1 X green buffer, 20 U XhoI, 10 U BamHI and 36 µl H_2O (nuclease free) then incubating for 1 hr at 37^oC. The product was run in an agarose gel, cut and purified using a gel purification kit (Qiagen) according to the manufacturer's instructions. The DNA was eluted by adding 50 µl of nuclease free H_2O.

2.2.5.3 Ligation of annealed oligonucleotides/DNA fragments to cut vector

Cut vectors were ligated with the annealed primers or other DNA fragments containing the same restriction site overhangs to form different constructs. Ligation was performed by adding 1 X ligation buffer, 2 µl pEGFP-C1 or pmCherry (cut with XhoI and BamHI), 4 µl annealed primers or 1:1 DNA insert to vector ratio, 1 µl T4 DNA ligase (5 Weiss U/µl) and H_2O (nuclease free) to give a total volume of 20 µl. This was then left overnight at 4^oC

before transformation to give a high possibility of transformants with the correct recombinant plasmid.

2.2.5.4 Ligation to TOPO Vector

Ligation of a PCR fragment to NT-GFP fusion TOPO (Invitrogen), a TA expression kit, was performed by adding 2 µl fresh PCR product, 1 µl salt solution (1.2 M $NaCl_2$, 0.06 M $MgCl_2$), 1 µl TOPO vector, and 2 µl H_2O (nuclease free) and left for 5 min at RT before transformation.

The competent cells (top 10 strain of *E. coli*) used for the transformation of the insert ligated to TOPO vector was provided by Invitrogen with the kit.

2.2.5.5 Recombinant plasmid construction from primary vector (pGEM-T Easy)

PCR products were ligated into pGEMT-easy vector using the materials and procedures provided by the kit (pGEM®-T Easy Vector Systems) and then transformed into 100 µl of competent *E. coli* cells. White colonies were grown and plasmid DNA was prepared and sequenced to confirm the correct construct was present. The DNA was then double digested, the inserts were purified using a gel extraction kit (Qiagen) and ligated to the already cut pEGFP-C1 and pMCherry. Ligation products were transformed into 100 µl of competent *E. coli* cells and the colonies were picked, analysed by colony PCR and then, colonies assumed to be correct were propagated and plasmid DNA was purified before being sequenced.

2.2.6 **Construction of mutant plasmids**

21 silent mutations were introduced to the conserved RNA structure in the $3D^{pol}$ encoding region of HPeV1. The designed sequence was synthesized commercially by Epoch Life Science, INC. and the plasmid received was cut with XhoI and BamHI prior to ligation into vectors cut with the same enzymes.

2.2.7 **Transformation process**

Competent cells were prepared from the Top 10 strain of *E. coli*. 200 µl of an overnight culture was added to 10 ml of fresh Luria Broth (LB) and incubated in a shaker for 2 hr at 37^{o}C. 5 ml of ice-cold 0.1 M $CaCl_2$ was then added to the sample after it was centrifuged (4000 rpm, 10 min, 4^{o}C) and the supernatant decanted. It was then left in ice for 1 hr and centrifuged again at the same conditions listed above. 500 µl of 0.1 M $CaCl_2$ was used to resuspend the pellet and the sample was kept in ice for at least 30 min before use.

100 µl of competent cells were incubated with 5 µl of the DNA in ice for 30 min, after which the samples were directly heat shocked at 42^{o}C in a water bath for 60 sec and immediately incubated in ice for 2 min. 250 µl LB was added to the samples in different 1.5 ml microcentrifuge tubes and incubated in a shaker (200 rpm, 37^{o}C, 1 hr). 100 µl of transformed cells were then spread on ampicillin plates (for constructs in TOPO vector) or kanamycin plates (for EGFP and MCherry constructs) and the samples were incubated at 37^{o}C for 16-24 hr.

2.2.8 Nucleic acid extraction and purification

2.2.8.1 Small-scale DNA extraction

Colonies were picked from each plate using a toothpick and inoculated in 10 ml LB containing ampicillin (100 µg/ml) or kanamycin (50 µg/ml). All samples were incubated at 37^{o}C for 24 hr and then plasmid DNA purification was carried out using a Miniprep kit (Qiagen) according to the instructions from the manufacturer.

2.2.8.2 Large-scale DNA extraction

After 4-6 hr incubation of colonies in 10 ml LB containing the necessary antibiotic, each sample was transferred to 200 ml LB (with corresponding antibiotic). Samples were incubated at 37^{o}C for 24 hr and centrifuged (6000 rpm, 15 min). The supernatant was discarded and bottles were inverted on a piece of tissue to dry completely. Plasmid DNA purification was carried out using a Midiprep kit (Qiagen) according to the instructions from the manufacturer.

2.2.8.3 DNA isolation from agarose gel

PCR products and digested DNA were recovered from the gel by carefully cutting the required band with a scalpel. The cut gel was placed in a 1.5 ml microcentrifuge tube and then purified with a gel extraction kit (Qiagen) according to the manufacture's instruction. Purified DNA was eluted using 30-50 µl nuclease free water.

2.2.9 **Agarose gel electrophoresis**

Usually a 1% agarose gel was made by dissolving 0.5 g of agarose in 50 ml ELFO buffer. However, when very small DNA fragments (≤200 bp) were to be analysed or purified, a 2% agarose gel was used. A homogenous mixture was obtained by bringing the mixture to a boil in a microwave. This was then left to cool at room temperature before 3 µl of Safe view nucleic acid stain was added and then the mixture was poured into a gel tray with a comb and sealed with tape at the open ends. The gel was left for 20-30 min to solidify. The gel was placed into an electrophoresis chamber containing enough ELFO buffer to cover it. The comb was gently removed to create slots in the gel for DNA to be loaded.

Electrophoresis of 5 µl 1kb DNA ladder solution was carried out alongside DNA samples (5 µl of the sample for analysis and 50 µl of the sample for DNA fragment isolation) mixed with loading dye and put in a different individual slot per sample. The loading dye to DNA mixture was a 1:5 ratio. Samples were run at 100 V for about 20 min before being viewed under blue light and an image taken (Illuminator SYNGENE) for future reference.

2.2.10 **Plaque assay**

Infection with CAV9 and HPeV1 was carried out using GMK and HT29 cells respectively. Serial dilution in serum free media was used to get 10^{-3} and 10^{-4} dilution of the viral load. A control of cells containing no virus was also done during each assay.

4 ml of the 5 ml media in the 25 cm^2 flask containing confluent cells was pipetted and discarded then 100 µl of each virus dilution was inoculated and kept on a rocking table (24 rpm, 1 hr). 1 ml of 2 % carboxymethyl cellulose (CMC)-agarose mixed with 1 ml of DMEM

(with serum) and 1 ml of DMEM (without serum) was added to the infected monolayer and allowed to solidify for 10 min at RT. The flasks were then incubated for 4-6 days.

PBS washing buffer was gently added to the sides of the flasks with the agarose gel overlay. Infected cells were stained with 0.1% CV for 15 min and washed with PBS to display plaques formed due to infection.

2.2.11 **Crystal violet growth assay/Wst 1 assay**

A crystal violet (CV) assay was performed to measure the amount of GMK cells still present in 24-well plates after the cells were treated with Actinomycin-D, H_2O_2, NaCl or Thapsigargin so as to note the optimum concentration of each chemical to be used on the cells based on the experiment. The assay was carried out by adding 500 µl of 2% formaldehyde per well and left at RT for 1 hr. Plates were then washed thrice with PBS and tapped to remove as much liquid as possible after every wash. They were left to dry and then 200 µl of 0.08% CV was added per well. This was left for another 1 hr at room temperature before being washed thrice with dH_2O and tapped again to remove as much liquid as possible after every wash. This was left to dry before adding 200 µl of 10% acetic acid to each well and then left on the shaker for 1hr at room temperature. Plates were then read at 490 nm wavelength using a microplate reader (BMG LABTECH).

A Wst 1 assay was also used to measure the number of viability of cells. 2 ml aliquots of Wst 1 were made, wrapped in aluminum foil and stored at -20^{o}C. 300 µl of media containing already treated cells was aspirated and discarded to leave 200 µl of media in each well. 20 µl of the Wst 1 was then added directly to each well, incubated at 37^{o}C in

optimal cell growth conditions for 1 hr and then read at 440 nm wavelength of absorbance using a microplate reader (BMG LABTECH).

2.2.12*In vitro* transcription

DNA to be used for in vitro transcription was first linearized using MluI restriction enzyme. The DNA product was then purified using a gel purification kit (Qiagen) according to the manufacturer's instruction and the DNA was eluted using nuclease free H_2O. 100 µl reaction was set up: 20 µl 5X transcription buffer, 10 µl 100 mM DTT, 20 µl rNTP mix (2.5 mM of each of the rNTP solution), 100 U Recombinant RNasin ribonuclease inhibitor, 1 µg digested DNA, 40 U T7 RNA polymerase and nuclease free water. This was incubated at 37^{o}C for 120 min and stored at -80^{o}C until use.

2.2.13Luciferase assay

1.8 µg of each DNA/RNA sample was transfected in a 24-well plate using the protocol of transfection in section 2.2.2.1. The media was changed the following day by adding fresh media and then incubated for another 16-24 hr (37^{o}C, 5% CO_2). To analyse the transfection, cells were washed twice with PBS and lysed using 60 µl of lysate buffer (1 X). This was then kept for not less than 20 min in -80^{o}C to completely freeze the cells. 20 µl of each lysate was then added to 20 µl of 1 X *Renilla* Luciferase assay substrate (diluted according to the manufacture's instruction) in a white 96-well plate. This was left on a flat surface and covered with foil paper for 15-20 min at room temperature. The plate was then read at 450 nm wavelength using an Illuminometer (BMG LABTECH).

2.2.14 **Computer analysis of DNA sequences**

Different online tools were used to analyse sequences throughout this research. They include:

Basic Local Alignment Search Tool (BLAST):

The NCBI BLAST: http://blast.ncbi.nlm.nih.gov/Blast.cgi was used in this research to compare input DNA sequence(s) with already existing sequence nucleotide sequence in the database (Altschul *et al*., 1990). The default parameters were used.

CLUSTAL-W:

This was carried out using http://www.genome.jp/tools/clustalw/ for aligning multiple DNA or amino acid sequence to compare similarities and differences of the sample queries (Larkin *et al.*, 2007).

MEGA5:

Phylogenetic tree was constructed with MEGA5.2 from http://www.megasoftware.net/ using the neighbour-joining tree option (Tamura *et al.*, 2011). This was done after carrying out alignment of all sequences to be used for generating the tree using Clustal W.

RNA-Folding programme (mfold):

RNA folding form from the mfold server http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form was used to predict RNA secondary structures using the default parameters of the form (Zuker, 2003).

The Sequence Manipulation Suite:

The design of reverse primers or analysing sequenced results in reverse direction was done by Reverse complement of DNA sequence from http://www.bioinformatics.org/sms/index.html (Stothard, 2000).

Translate:

This is an option from ExPASy: http://web.expasy.org/translate/ which enables the translation of DNA or RNA sequence to amino acids (Gasteiger *et al.*, 2003).

Web-cutter:

Sites of restriction enzymes within a sequence of DNA were identified using Webcutter 2.0 at http://rna.lundberg.gu.se/cutter2/

2.2.15 **Highlighting of sequences**

All highlighting in different colours were carried out manually throughout this thesis.

Chapter 3

Examining the role of an RNA structure in the 3D^{pol}-encoding region of HPeV1 on RNA stability

3.1 Introduction

The picornavirus genome contains several regions where RNA secondary structures are important. At the 5'end of the genome is an untranslated region (5' UTR). This region contains the highly structured Internal Ribosome Entry Site (IRES) which initiates the translation process into a single polyprotein, as well as structures involved in RNA replication (Nateri *et al.*, 2002; Han *et al.*, 2007; Lin *et al.*, 2009; Song *et al.*, 2012)l. The 3' end of the genome also has a UTR and this is followed by a poly (A) tail. The poly (A) tail after the 3' UTR varies in length from one picornavirus to another and the 3' UTR itself is variable in length with the shortest being in rhinoviruses (~35 nb) and the longest in *Megrivirus* (~654 nb) (Farkas *et al.*, 2012). This region is not essential for virus replication as poliovirus mutants lacking the 3' UTR can be produced (Rohll *et al.*, 1995; Todd & Semler, 1996), but it has been reported to aid viral RNA stability and efficient RNA replication (Bedard & Semler, 2004). The 3' UTR often contains RNA secondary structures, for instance a single stem-loop in rhinoviruses (Rohll *et al.*, 1995) and two stem-loops which interacts to form a kissing interaction in enteroviruses (Mirmomeni *et al.*, 1997). A highly conserved structure was discovered in the 3' UTR of three different families of positive sense RNA viruses; *Astrovirus*, *Coronavirus* and Equine rhinovirus suggesting a common ancestral origin and usefulness of the structure (Jonassen *et al.*, 1998b; Robertson *et al.*, 2004). Another structure is found in the 3' UTR of several different picornavirues (Boros *et al.*, 2012). The $3D^{pol}$ is encoded in the ORF just before the 3' UTR and it is responsible for the replication of the RNA genome via a negative sense RNA intermediate (Kerkvliet *et al.*, 2010). It also performs VPg uridylylation in conjunction with the cre. The cre is found in all picornaviruses and it is needed for VPg uridylylation during

RNA replication (Paul *et al.*, 2003). It is a stem-loop structure containing a loop with an A residue needed as a template for the uridylylation. The cre is responsible for the initiation of viral replication as VPgpUpU is the primer for RNA replication (Paul *et al.*, 2003; Bedard & Semler, 2004). The cre is found in different places in different picornaviruses, usually within the coding region (Goodfellow *et al.*, 2000; Liu *et al.*, 2009a; Tolf *et al.*, 2009) but the *Aphthovirus* cre is in the 5' UTR (Mason *et al.*, 2002). More RNA structures have been found within the coding region of poliovirus. These have been reported to be involved in RNA replication (Song *et al.*, 2012) or binding to RNase L to overcome host cell defences (Han *et al.*, 2007; Townsend *et al.*, 2008). Another RNA structure was predicted in the 3D-encoding region of HPeVs (Williams *et al.*, 2009) but its function is unknown. The aim of the work described in this chapter is to find if the structure found in 3D influences RNA stability.

3.2 Investigation of the 3D conserved structure

3.2.1 Confirmation of the structure

To investigate if the previously reported RNA structure (position 7008 to 7126 in HPeV1 Harris) is conserved in HPeVs (Williams *et al.*, 2009), a BLAST search and alignment was performed to find all the sequences in the database which contain all or most of the region being studied. 254 HPeV sequences were found, 212 complete and 42 with nucleotides missing from the 5' and/or 3' ends of the region. The incomplete sequences were included to maximise potential variability. The alignment is shown in Figure 3.2. The region is highly conserved throughout the sequences with few positions where there is variation. Many of these are positions predicted to not be base-paired. The first stem loop is supported by two pairs of covariance, although in a few of the sequences non-covariant changes are seen at these positions and weaken the structure. Other changes that disrupt the structure are mainly in the CAG/CUG predicted to make up the bottom of the stem and this part of the structure may not form. The second stem-loop is mainly supported by changes which maintain G-C/U or U-G/A interactions. In some of the sequences there is a change which affects the top base pair and this increase the size of the loop in these sequences. The third stem-loop is supported by one covariant change and by changes which keep G-C/U or U-G/A interactions. The predicted kissing interaction is highly conserved with just a U/C difference that maintains the interaction with a G. In summary the structure is mostly well-supported by the many virus sequences analysed.

To investigate if the structure is maintained in related viruses, alignments with Lungan viruses (*Parechovirus B* specie), Sebokele virus and Ferret parechovirus (proposed as “Parechovirus C” and “D” respectively) were performed. There was no evidence of a related structure in *Parechovirus B* or “D” (data not shown), but a similar type of structure could be predicted for “Parechovirus C”, Sebokele virus (Figure 3.3). The sequences making up the structure are very different between HPeV and Sebokele virus, but the same pattern of three stem-loops and a kissing interaction between the loops of the first and third stem-loops could be identified. As there is only one available Sebokele virus sequence this could not be studied by covariance.

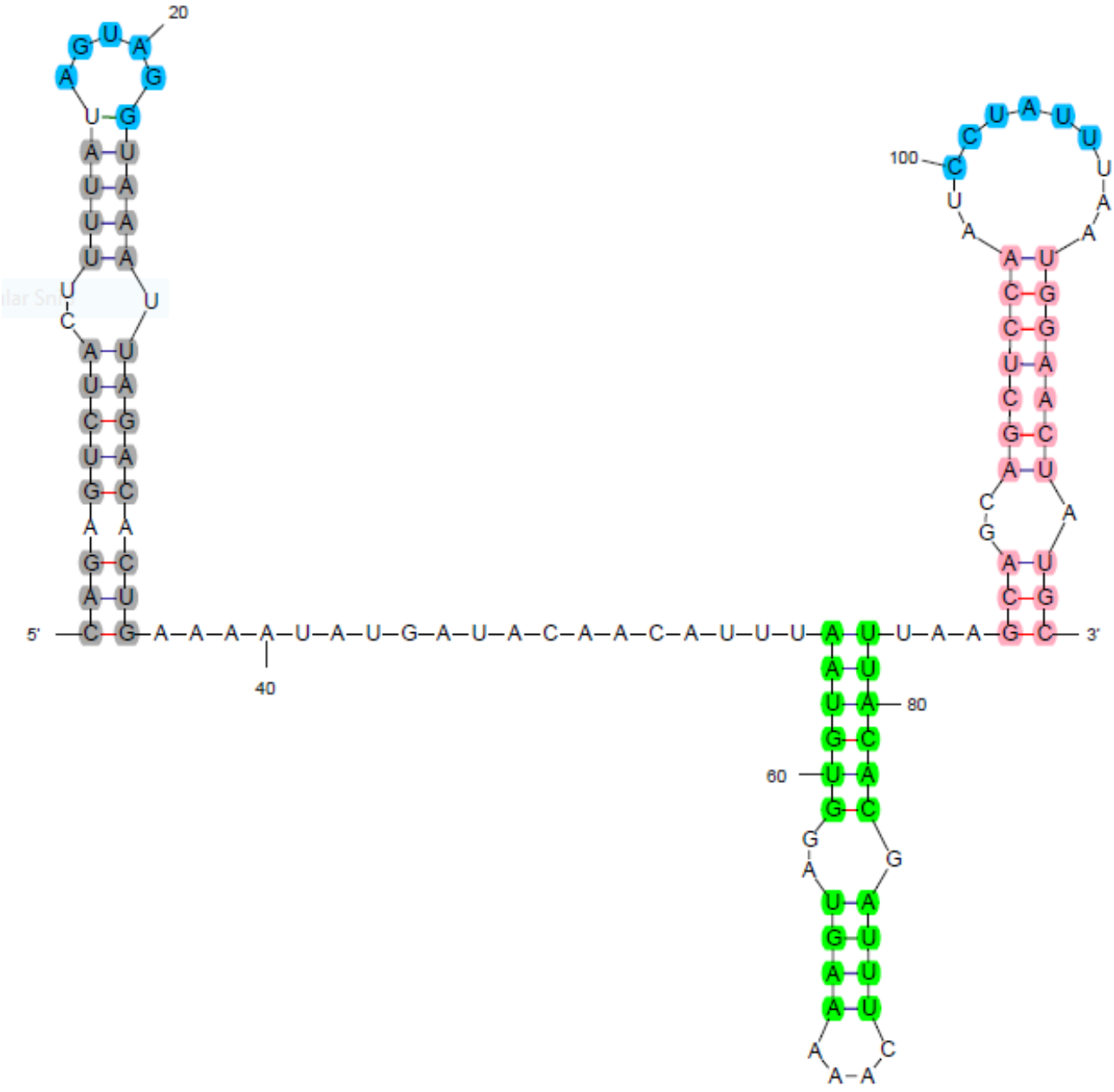
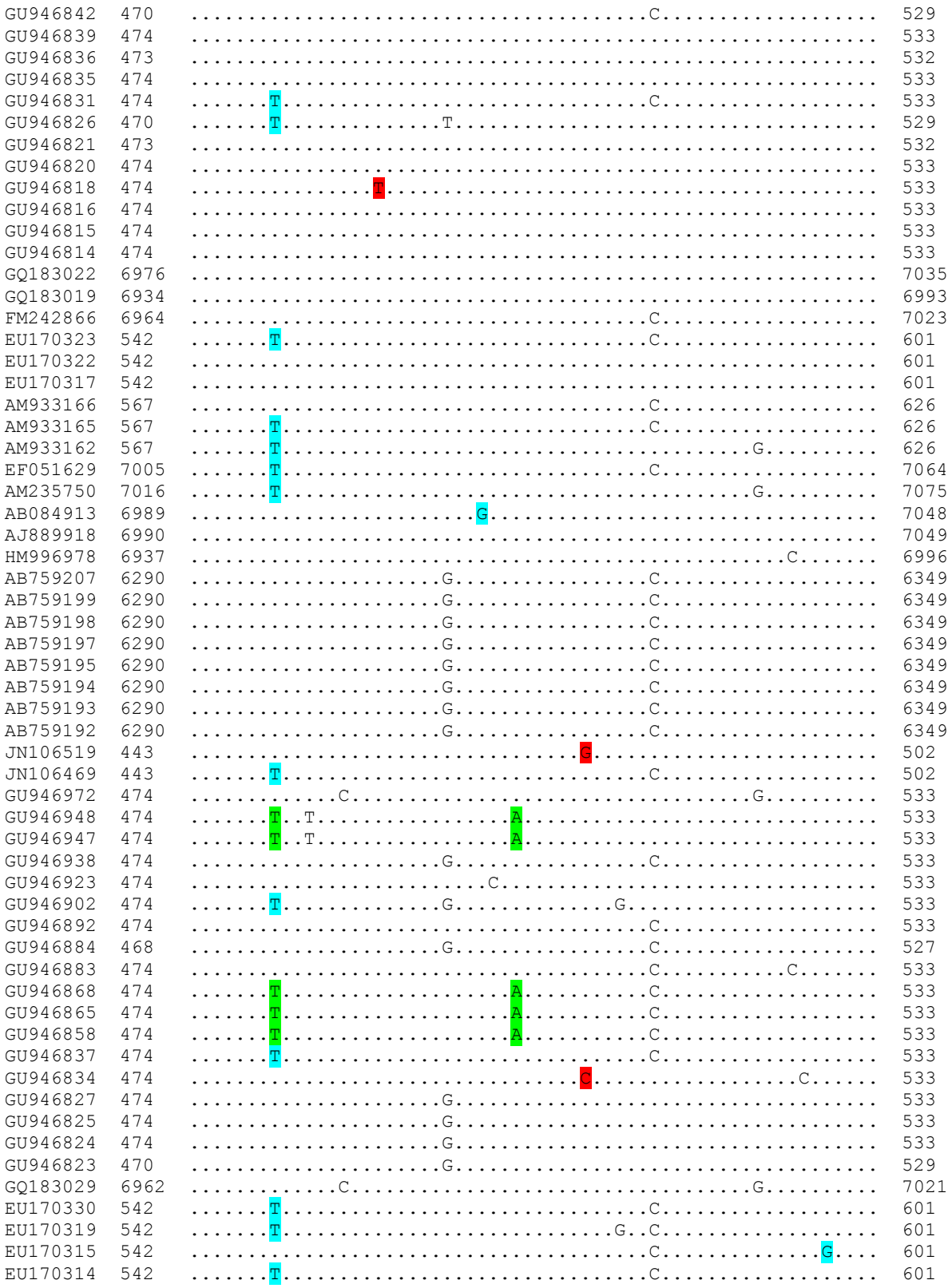
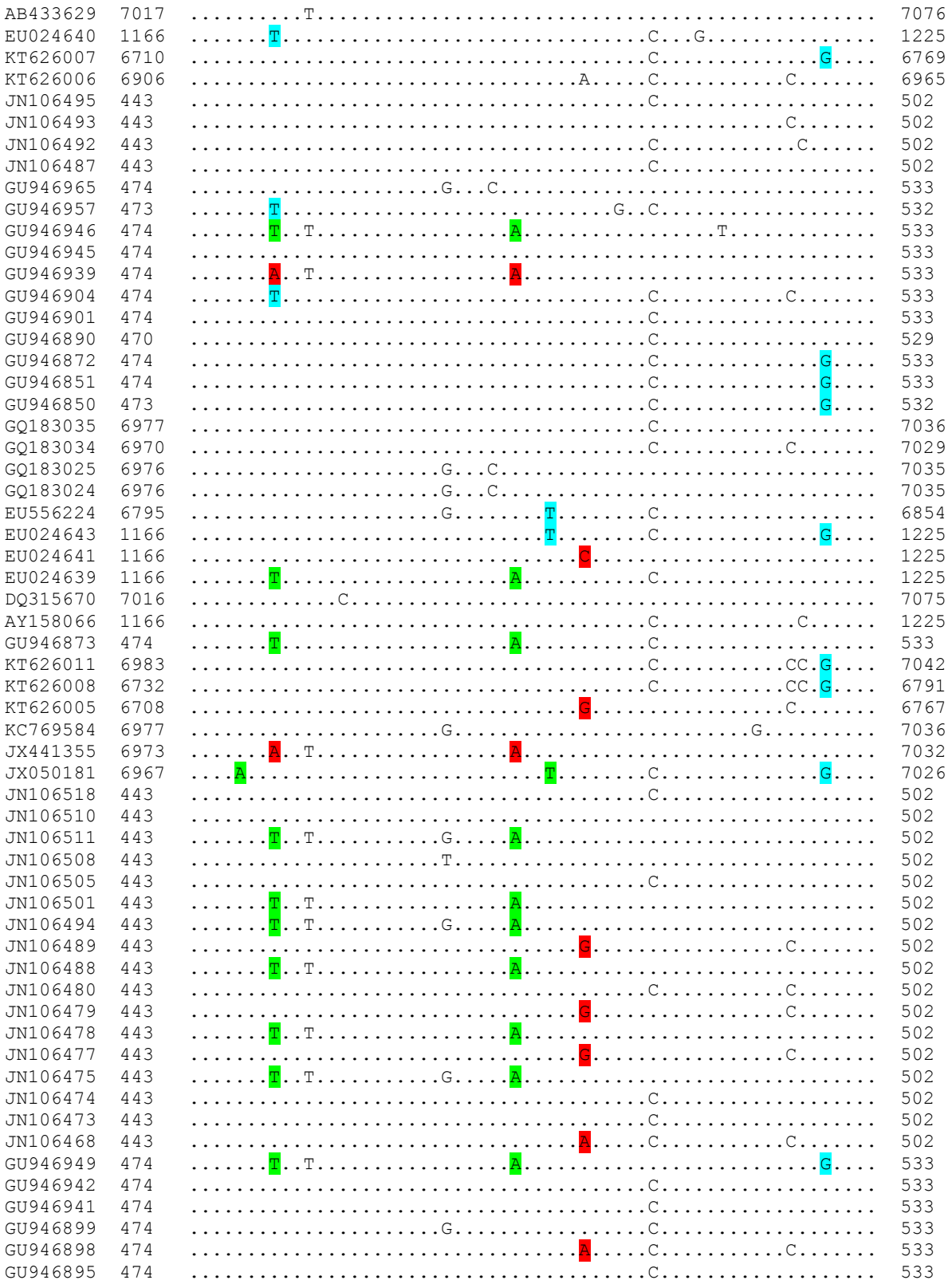


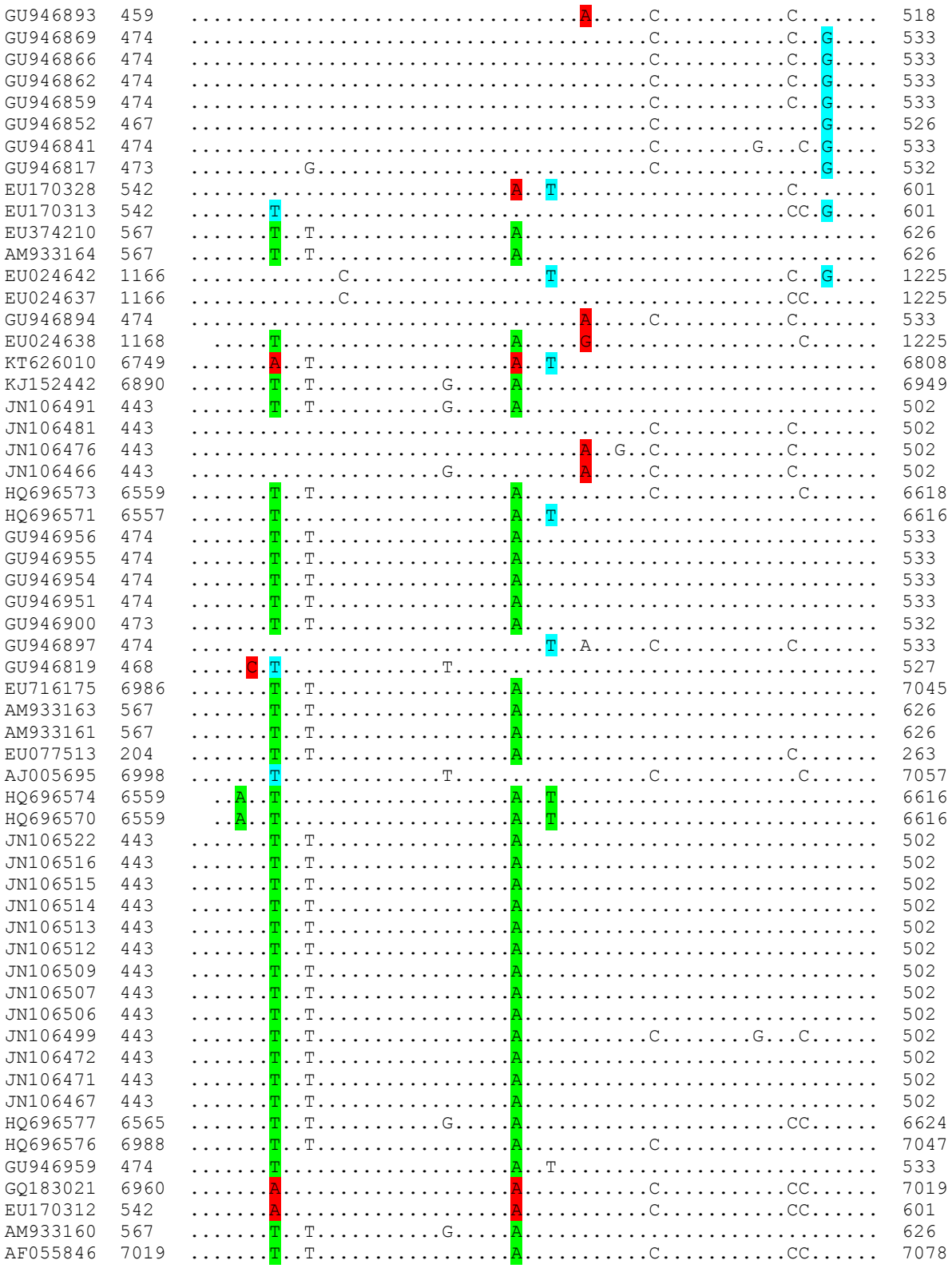
Figure 3.1: Schematic representation of the predicted RNA secondary structure in the 3D encoding region of HPeV1. The three stem loops are highlighted in grey, green and pink while the tertiary structure kissing interaction between two of the loops is highlighted in blue. dG = -20.40 [initially -20.40]. The schematic figure was generated using the default parameters of the RNA folding form for Mfold.

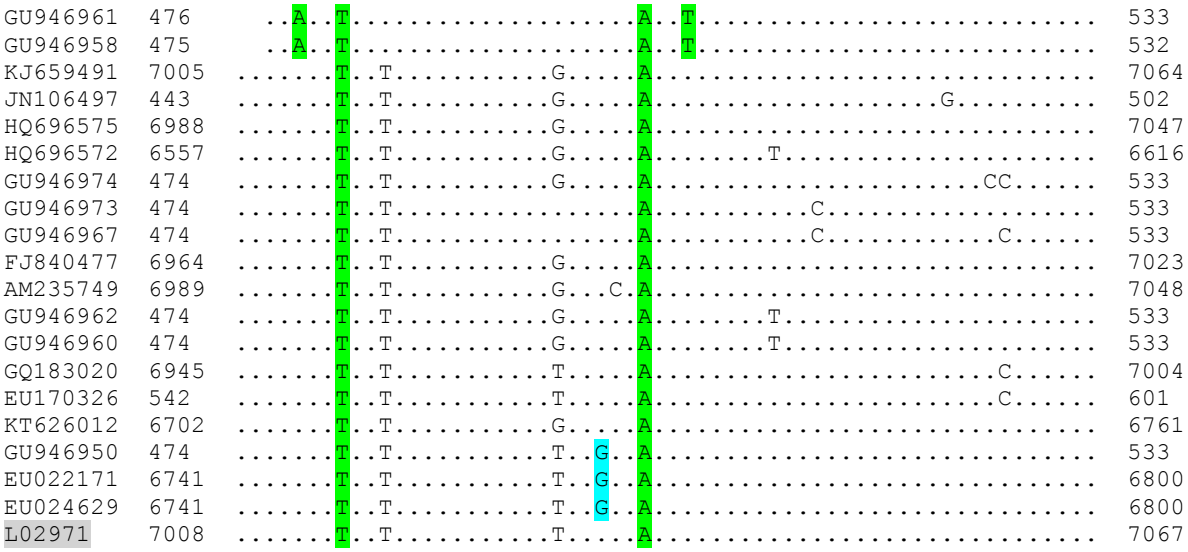
```
Query     1     CAGAGTCCACATTTATAGTAGGAAAATTGGACACTGAAAATATGATACAACATTTAATGT  60
GU946968  474   ............................................................  533
GU946964  474   ............................................................  533
GU946921  474   ............................................................  533
GU946870  474   ............................................................  533
GU946840  474   ............................................................  533
GQ183032  6963  ............................................................  7022
GQ183030  6963  ............................................................  7022
GQ183023  6976  ............................................................  7035
LC027514  2953  ..........................................C.................  3012
LC027475  2949  ..........................................C.................  3008
LC027474  2954  ..........................................C.................  3013
LC027513  2952  ..........................................C.................  3011
LC027512  2953  ..........................................C.................  3012
LC027502  2935  ..........................................C.................  2994
LC027501  2950  ..........................................C.................  3009
KM986843  6997  ..........................................C.................  7056
KJ659490  6997  ..........................................C.................  7056
JX826607  6833  ..........................................C.................  6892
AB759205  6290  ............................................................  6349
GU946971  474   ............................................................  533
GU946969  474   ............................................................  533
GU946943  474   ............................................................  533
GU946937  474   ..........................................C.................  533
GU946936  468   ............................................................  527
GU946934  470   ............................................................  529
GU946933  468   ............................................................  527
GU946932  470   ............................................................  529
GU946931  470   ..........................................C.................  529
GU946928  470   ............................................................  529
GU946927  467   ............................................................  526
GU946926  473   ............................................................  532
GU946919  470   ............................................................  529
GU946918  470   ..........................................C.................  529
GU946917  470   ............................................................  529
GU946912  474   ............................................................  533
GU946911  474   ............................................................  533
GU946910  468   ............................................................  527
GU946909  474   ............................................................  533
GU946908  474   ............................................................  533
GU946907  474   ............................................................  533
GU946906  474   ............................................................  533
GU946905  474   ............................................................  533
GU946903  468   ............................................................  527
GU946856  474   ............................................................  533
GU946855  474   ............................................................  533
GU946853  465   ............................................................  524
GU946847  473   ............................................................  532
GU946844  474   ............................................................  533
GU946838  474   ............................................................  533
GU946833  468   ............................................................  527
GU946829  474   ............................................................  533
GU946828  474   ............................................................  533
GU946822  474   ............................................................  533
GQ183033  6963  ............................................................  7022
GQ183031  6962  ............................................................  7021
GQ183028  6933  ............................................................  6992
GQ183027  6933  ............................................................  6992
GQ183026  6939  ............................................................  6998
```

```
FM178558  7004  ..........................................................  7063
EU170340  542   ..........................................................  601
EU170339  542   ..........................................................  601
EU170338  542   ..........................................................  601
EU170335  542   ..........................................................  601
EU170334  542   ..........................................................  601
EU170333  542   ..........................................................  601
EU170332  542   ..........................................................  601
EU170327  542   ..........................................................  601
EU170324  542   ..........................................................  601
EU170321  542   ..........................................................  601
EU170320  542   .......T..................................................  601
EU170318  542   ..........................................................  601
GU946913  476     ........................................................  533
KX068679  6991  ....................................C.....................  7050
KT626009  6725  ....................................C.....................  6784
LC027487  2953  ....................................C.....................  3012
LC027486  2935  ....................................C.....................  2994
JX575746  7003  .................................C.....C.....................  7062
AB759206  6290  ..........................................................  6349
AB759204  6290  ..........................................................  6349
AB759203  6290  ..........................................................  6349
AB759202  6290  ..........................................................  6349
AB759201  6290  ..........................................................  6349
AB759200  6290  ..........................................................  6349
AB759191  6290  ....................................C.....................  6349
AB759190  6290  ....................................C.....................  6349
AB759189  6290  ....................................C.....................  6349
AB759187  6290  ....................................C.....................  6349
AB759186  6290  ....................................C.....................  6349
AB759185  6290  ....................................C.....................  6349
JN106498  443   ....................................C.....................  502
JN106490  443   ....................................C.....................  502
JN106486  443   ................................T.........................  502
AB668033  6290  ....................................C.....................  6349
AB668032  6290  ....................................C.....................  6349
AB668031  6290  ....................................C.....................  6349
AB668030  6290  ....................................C.....................  6349
AB668029  6395  ....................................C.....................  6454
GU946963  474   ..........................................................  533
GU946944  474   ....................................C.....................  533
GU946940  474   ....................G...............C.....................  533
GU946935  473   ..........................................................  532
GU946930  469   ..........................................................  528
GU946929  467   ..........................................................  526
GU946925  474   ..........................................................  533
GU946924  474   ..........................................................  533
GU946922  470   ..........................................................  529
GU946920  470   ..........................................................  529
GU946916  470   ..........................................................  529
GU946915  470   ..........................................................  529
GU946914  474   ..........................................................  533
GU946896  474   ..........T...............................................  533
GU946891  474   ..................................................C.......  533
GU946885  474   ....................G.....................................  533
GU946882  473   ....................G.....................................  532
GU946879  473   ....................G.....................................  532
GU946878  474   ..........................................................  533
GU946874  474   ....................G.....................................  533
```


GU946842 470 529
GU946839 474 533
GU946836 473 532
GU946835 474 533
GU946831 474 533
GU946826 470 529
GU946821 473 532
GU946820 474 533
GU946818 474 533
GU946816 474 533
GU946815 474 533
GU946814 474 533
GQ183022 6976 7035
GQ183019 6934 6993
FM242866 6964 7023
EU170323 542 601
EU170322 542 601
EU170317 542 601
AM933166 567 626
AM933165 567 626
AM933162 567 626
EF051629 7005 7064
AM235750 7016 7075
AB084913 6989 7048
AJ889918 6990 7049
HM996978 6937 6996
AB759207 6290 6349
AB759199 6290 6349
AB759198 6290 6349
AB759197 6290 6349
AB759195 6290 6349
AB759194 6290 6349
AB759193 6290 6349
AB759192 6290 6349
JN106519 443 502
JN106469 443 502
GU946972 474 533
GU946948 474 533
GU946947 474 533
GU946938 474 533
GU946923 474 533
GU946902 474 533
GU946892 474 533
GU946884 468 527
GU946883 474 533
GU946868 474 533
GU946865 474 533
GU946858 474 533
GU946837 474 533
GU946834 474 533
GU946827 474 533
GU946825 474 533
GU946824 474 533
GU946823 470 529
GQ183029 6962 7021
EU170330 542 601
EU170319 542 601
EU170315 542 601
EU170314 542 601

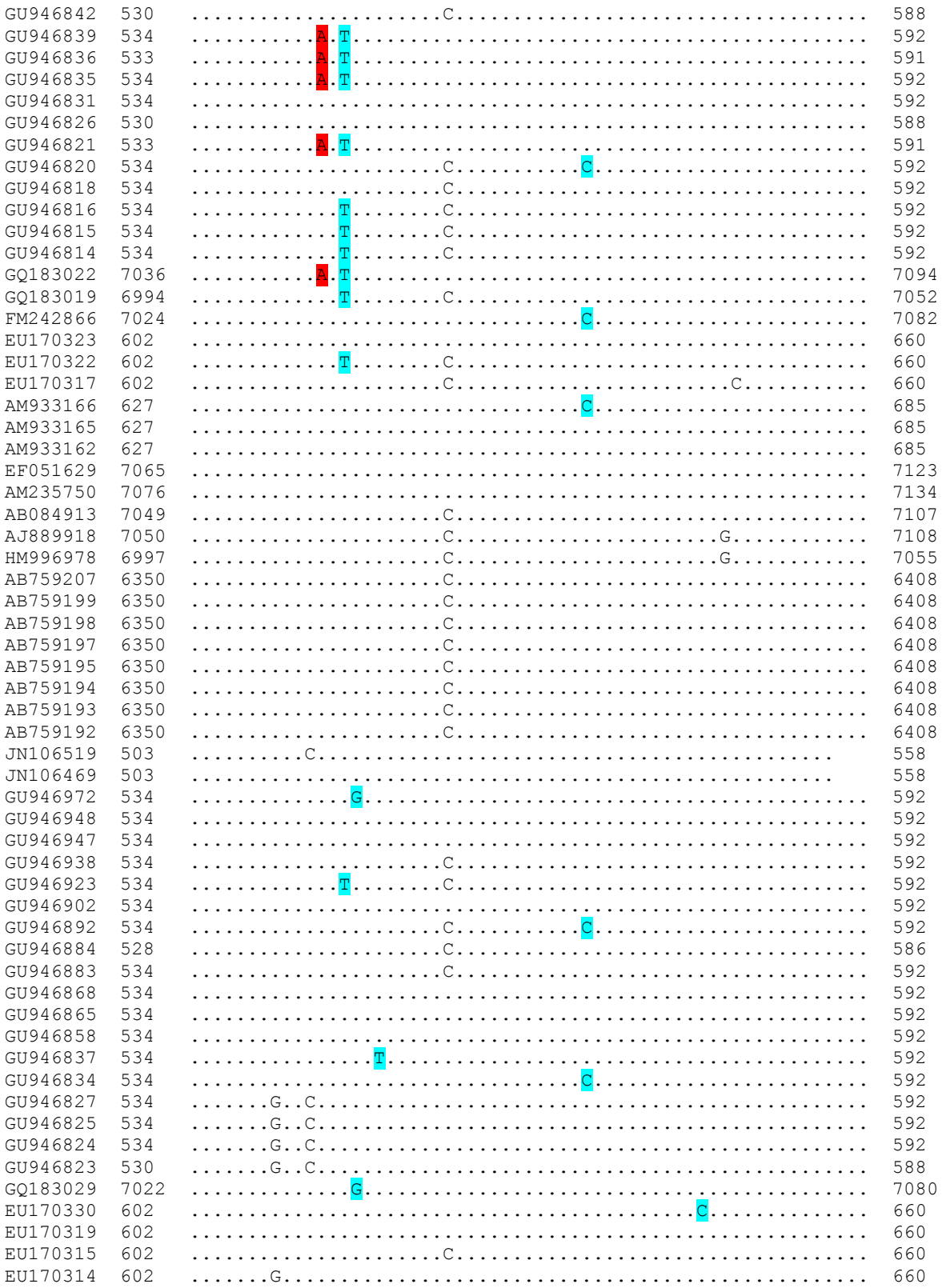

AB433629 7017 ....T.......... 7076
EU024640 1166 .......T.....C...G..... 1225
KT626007 6710 ..........C.........G.... 6769
KT626006 6906 ..........A.....C.....C..... 6965
JN106495 443 ....C.......... 502
JN106493 443 ..........C.....C..... 502
JN106492 443 ..........C.....C..... 502
JN106487 443 ....C.......... 502
GU946965 474 ..........G...C.......... 533
GU946957 473 .......T.....G..C.......... 532
GU946946 474 .......T..T.....A.....T..... 533
GU946945 474 533
GU946939 474 .......A..T.....A.......... 533
GU946904 474 .......T..........C.....C..... 533
GU946901 474 ....C.......... 533
GU946890 470 ....C.......... 529
GU946872 474 ..........C.........G.... 533
GU946851 474 ..........C.........G.... 533
GU946850 473 ..........C.........G.... 532
GQ183035 6977 ....C.......... 7036
GQ183034 6970 ..........C.....C..... 7029
GQ183025 6976 ..........G...C.......... 7035
GQ183024 6976 ..........G...C.......... 7035
EU556224 6795 ..........G.....T.....C..... 6854
EU024643 1166 .........T.....C.........G.... 1225
EU024641 1166 ....C.......... 1225
EU024639 1166 .......T.....A.....C..... 1225
DQ315670 7016 ....C.......... 7075
AY158066 1166 ..........C.....C..... 1225
GU946873 474 .......T.....A.....C..... 533
KT626011 6983 ..........C.....CC.G.... 7042
KT626008 6732 ..........C.....CC.G.... 6791
KT626005 6708 ..........G.....C..... 6767
KC769584 6977 ..........G.....G..... 7036
JX441355 6973 .......A..T.....A.......... 7032
JX050181 6967 ....A.....T.....C.........G.... 7026
JN106518 443 ....C.......... 502
JN106510 443 502
JN106511 443 .......T..T.....G.....A..... 502
JN106508 443 .....T.......... 502
JN106505 443 ....C.......... 502
JN106501 443 .......T..T.....A.......... 502
JN106494 443 .......T..T.....G.....A..... 502
JN106489 443 ..........G.....C..... 502
JN106488 443 .......T..T.....A.......... 502
JN106480 443 ..........C.....C..... 502
JN106479 443 ..........G.....C..... 502
JN106478 443 .......T..T.....A.......... 502
JN106477 443 ..........G.....C..... 502
JN106475 443 .......T..T.....G.....A..... 502
JN106474 443 ....C.......... 502
JN106473 443 ....C.......... 502
JN106468 443 ..........A.....C.....C..... 502
GU946949 474 .......T..T.....A.........G.... 533
GU946942 474 ....C.......... 533
GU946941 474 ....C.......... 533
GU946899 474 .....G.....C.......... 533
GU946898 474 ..........A.....C.....C..... 533
GU946895 474 ....C.......... 533

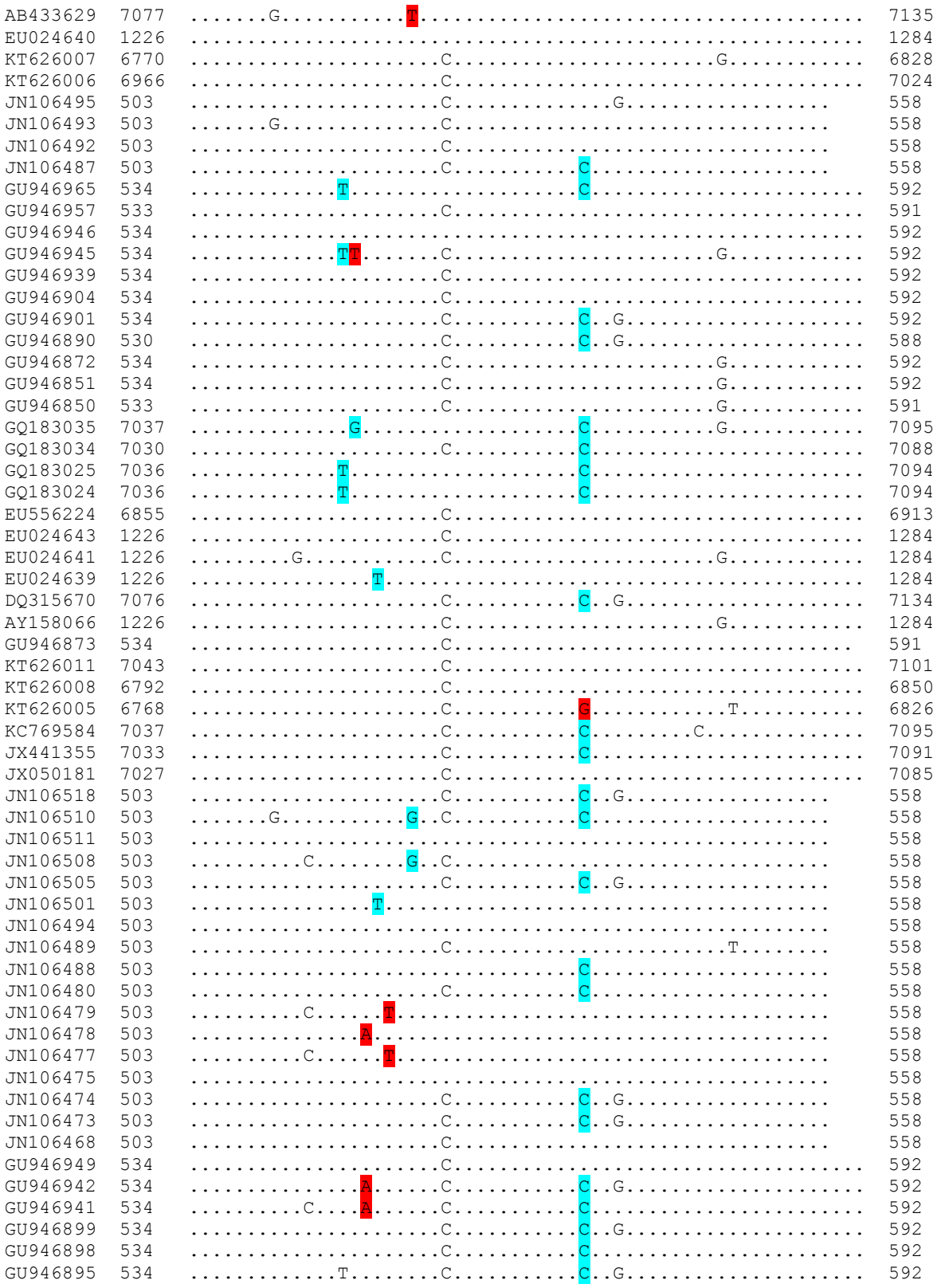

GU946893 459 518
GU946869 474 533
GU946866 474 533
GU946862 474 533
GU946859 474 533
GU946852 467 526
GU946841 474 533
GU946817 473 532
EU170328 542 601
EU170313 542 601
EU374210 567 626
AM933164 567 626
EU024642 1166 1225
EU024637 1166 1225
GU946894 474 533
EU024638 1168 1225
KT626010 6749 6808
KJ152442 6890 6949
JN106491 443 502
JN106481 443 502
JN106476 443 502
JN106466 443 502
HQ696573 6559 6618
HQ696571 6557 6616
GU946956 474 533
GU946955 474 533
GU946954 474 533
GU946951 474 533
GU946900 473 532
GU946897 474 533
GU946819 468 527
EU716175 6986 7045
AM933163 567 626
AM933161 567 626
EU077513 204 263
AJ005695 6998 7057
HQ696574 6559 6616
HQ696570 6559 6616
JN106522 443 502
JN106516 443 502
JN106515 443 502
JN106514 443 502
JN106513 443 502
JN106512 443 502
JN106509 443 502
JN106507 443 502
JN106506 443 502
JN106499 443 502
JN106472 443 502
JN106471 443 502
JN106467 443 502
HQ696577 6565 6624
HQ696576 6988 7047
GU946959 474 533
GQ183021 6960 7019
EU170312 542 601
AM933160 567 626
AF055846 7019 7078

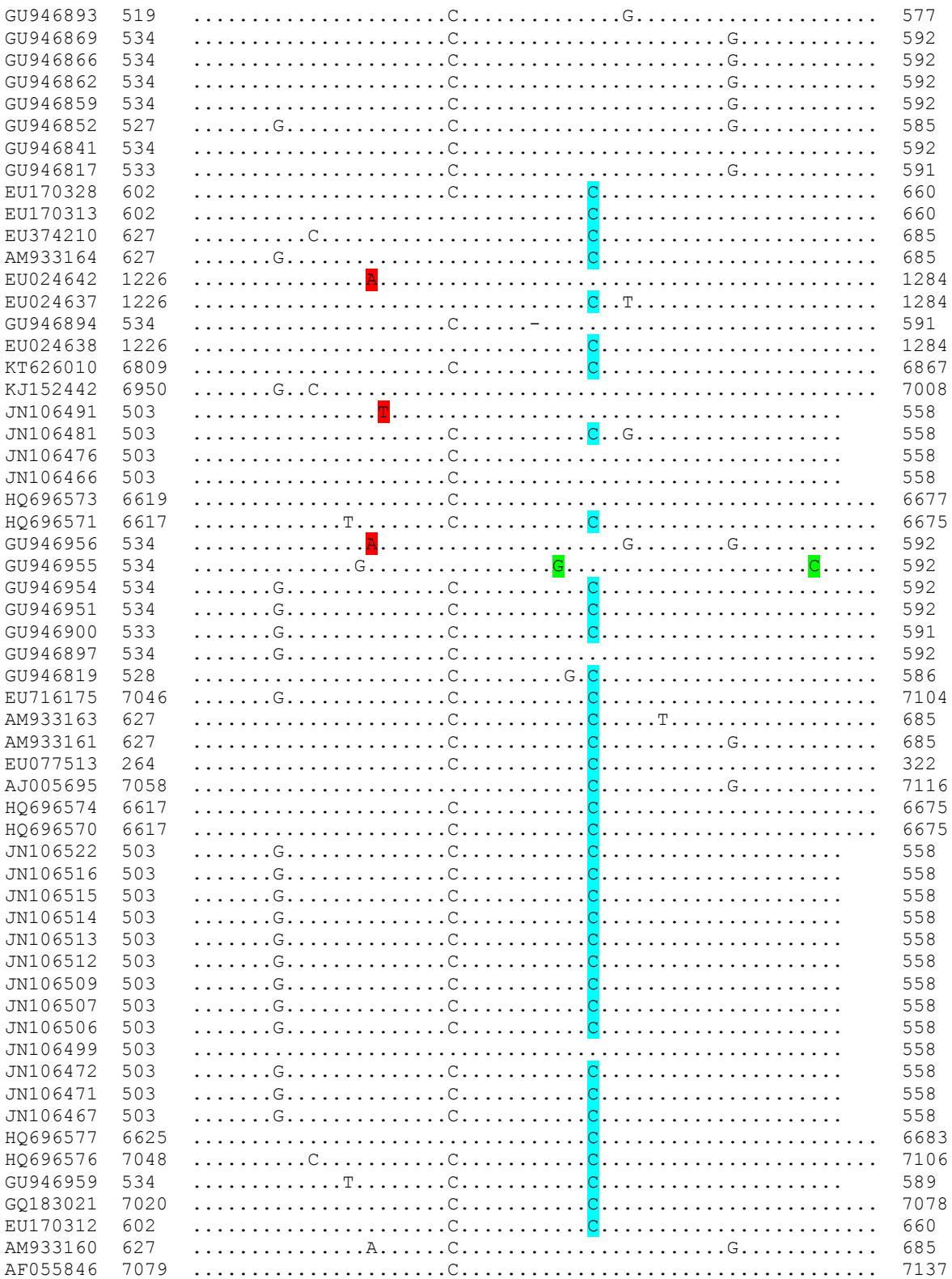

GU946961 476 ..A..T.....A..T..... 533
GU946958 475 ..A..T.....A..T..... 532
KJ659491 7005 .....T..T.....G.....A..... 7064
JN106497 443 .....T..T.....G.....A.....G..... 502
HQ696575 6988 .....T..T.....G.....A..... 7047
HQ696572 6557 .....T..T.....G.....A.....T..... 6616
GU946974 474 .....T..T.....G.....A.....CC..... 533
GU946973 474 .....T..T.....A.....C..... 533
GU946967 474 .....T..T.....A.....C.....C..... 533
FJ840477 6964 .....T..T.....G.....A..... 7023
AM235749 6989 .....T..T.....G...C.A..... 7048
GU946962 474 .....T..T.....G.....A.....T..... 533
GU946960 474 .....T..T.....G.....A.....T..... 533
GQ183020 6945 .....T..T.....T.....A.....C..... 7004
EU170326 542 .....T..T.....T.....A.....C..... 601
KT626012 6702 .....T..T.....G.....A..... 6761
GU946950 474 .....T..T.....T..G..A..... 533
EU022171 6741 .....T..T.....T..G..A..... 6800
EU024629 6741 .....T..T.....T..G..A..... 6800
L02971 7008 .....T..T.....T.....A..... 7067

```
Query      61    GGATGAAAAATTTCAGCACATTTAAGCAGCAACTTCAATCCTACTTAATGGAGTTATGC  119
GU946968   534   ............................................................  592
GU946964   534   ............................................................  592
GU946921   534   ............................................................  592
GU946870   534   ............................................................  592
GU946840   534   ............................................................  592
GQ183032   7023  ............................................................  7081
GQ183030   7023  ............................................................  7081
GQ183023   7036  ............................................................  7094
LC027514   3013  ............................................................  3071
LC027475   3009  ............................................................  3067
LC027474   3014  ............................................................  3072
LC027513   3012  ............................................................  3070
LC027512   3013  ............................................................  3071
LC027502   2995  ............................................................  3053
LC027501   3010  ............................................................  3068
KM986843   7057  ............................................................  7115
KJ659490   7057  ............................................................  7115
JX826607   6893  ............................................................  6951
AB759205   6350  ......................C.....................................  6408
GU946971   534   ......................C.....................................  592
GU946969   534   ......................C.....................................  592
GU946943   534   ..................................C.........................  592
GU946937   534   ............................................................  592
GU946936   528   ......................C.....................................  586
GU946934   530   ......................C.....................................  588
GU946933   528   ......................C.....................................  586
GU946932   530   ......................C.....................................  588
GU946931   530   ............................................................  588
GU946928   530   ......................C.....................................  588
GU946927   527   ......................C.....................................  585
GU946926   533   ......................C.....................................  591
GU946919   530   ......................C.....................................  588
GU946918   530   ............................................................  588
GU946917   530   ......................C.....................................  588
GU946912   534   ......................C.....................................  592
GU946911   534   ......................C.....................................  592
GU946910   528   ......................C.....................................  586
GU946909   534   ......................C.....................................  592
GU946908   534   ............T.............................................  592
GU946907   534   ............T.............................................  592
GU946906   534   ......................C.....................................  592
GU946905   534   ......................C.....................................  592
GU946903   528   ......................C.....................................  586
GU946856   534   ......................C.....................................  592
GU946855   534   ......................C.....................................  592
GU946853   525   ......................C.....................................  583
GU946847   533   ......................C.....................................  591
GU946844   534   ......................C.....................................  592
GU946838   534   ......................C.....................................  592
GU946833   528   ......................C.....................................  586
GU946829   534   ......................C.....................................  592
GU946828   534   ......................C.....................................  592
GU946822   534   ......................C.....................................  592
GQ183033   7023  ......................C.....................................  7081
GQ183031   7022  ......................C.....................................  7080
GQ183028   6993  ......................C.....................................  7051
GQ183027   6993  ......................C.....................................  7051
GQ183026   6999  ......................C.....................................  7057
```

```
FM178558  7064  ......................C.....................................  7122
EU170340  602   ......................C.....................................  660
EU170339  602   ......................C.....................................  660
EU170338  602   ......................C.....................................  660
EU170335  602   ......................C.....................................  660
EU170334  602   ......................C.....................................  660
EU170333  602   ......................C.....................................  660
EU170332  602   ......................C.....................................  660
EU170327  602   ......................C.....................................  660
EU170324  602   ......................C.....................................  660
EU170321  602   ......................C.....................................  660
EU170320  602   ........................................................................  660
EU170318  602   ......................C.....................................  660
GU946913  534   ............T...........................................................  592
KX068679  7051  ............T...........................................................  7109
KT626009  6785  ..........................................T.............  6843
LC027487  3013  ......................C.....................................  3071
LC027486  2995  ......................C.....................................  3053
JX575746  7063  ........................................................................  7121
AB759206  6350  .........G............C.....................................  6408
AB759204  6350  .........G............C.....................................  6408
AB759203  6350  .........G............C.....................................  6408
AB759202  6350  .........G............C.....................................  6408
AB759201  6350  .........G............C.....................................  6408
AB759200  6350  .........G............C.....................................  6408
AB759191  6350  ......................C.....................................  6408
AB759190  6350  ......................C.....................................  6408
AB759189  6350  ......................C.....................................  6408
AB759187  6350  ......................C.....................................  6408
AB759186  6350  ......................C.....................................  6408
AB759185  6350  ......................C.....................................  6408
JN106498  503   ..........................................................  558
JN106490  503   ..........................................................  558
JN106486  503   ..........................................................  558
AB668033  6350  ......................C.....................................  6408
AB668032  6350  ......................C.....................................  6408
AB668031  6350  ......................C.....................................  6408
AB668030  6350  ......................C.....................................  6408
AB668029  6455  ......................C.....................................  6513
GU946963  534   ...........A.T...........................................................  592
GU946944  534   ......................C.....................................  592
GU946940  534   ........................................................................  592
GU946935  533   ............T........C.....................................  591
GU946930  529   ............T........C.....................................  587
GU946929  527   ............T........C.....................................  585
GU946925  534   ............T........C.....................................  592
GU946924  534   ............T........C.....................................  592
GU946922  530   ............T........C.....................................  588
GU946920  530   ............T........C.....................................  588
GU946916  530   ............T........C.....................................  588
GU946915  530   ............T........C.....................................  588
GU946914  534   .........G............C.....................................  592
GU946896  534   ..............G.........................................................  592
GU946891  534   ......................C.....................................  592
GU946885  534   ......................C.....................................  592
GU946882  533   ......................C.....................................  591
GU946879  533   ......................C.....................................  591
GU946878  534   ......................C....................G.............  592
GU946874  534   ......................C.....................................  592
```


GU946842 530 588
GU946839 534 592
GU946836 533 591
GU946835 534 592
GU946831 534 592
GU946826 530 588
GU946821 533 591
GU946820 534 592
GU946818 534 592
GU946816 534 592
GU946815 534 592
GU946814 534 592
GQ183022 7036 7094
GQ183019 6994 7052
FM242866 7024 7082
EU170323 602 660
EU170322 602 660
EU170317 602 660
AM933166 627 685
AM933165 627 685
AM933162 627 685
EF051629 7065 7123
AM235750 7076 7134
AB084913 7049 7107
AJ889918 7050 7108
HM996978 6997 7055
AB759207 6350 6408
AB759199 6350 6408
AB759198 6350 6408
AB759197 6350 6408
AB759195 6350 6408
AB759194 6350 6408
AB759193 6350 6408
AB759192 6350 6408
JN106519 503 558
JN106469 503 558
GU946972 534 592
GU946948 534 592
GU946947 534 592
GU946938 534 592
GU946923 534 592
GU946902 534 592
GU946892 534 592
GU946884 528 586
GU946883 534 592
GU946868 534 592
GU946865 534 592
GU946858 534 592
GU946837 534 592
GU946834 534 592
GU946827 534 592
GU946825 534 592
GU946824 534 592
GU946823 530 588
GQ183029 7022 7080
EU170330 602 660
EU170319 602 660
EU170315 602 660
EU170314 602 660


AB433629 7077 .......G..........T..... 7135
EU024640 1226 1284
KT626007 6770 .....C.....G..... 6828
KT626006 6966 .....C..... 7024
JN106495 503 .....C.....G..... 558
JN106493 503 .......G.....C..... 558
JN106492 503 .....C..... 558
JN106487 503 .....C.....C..... 558
GU946965 534 ..........T.....C..... 592
GU946957 533 .....C..... 591
GU946946 534 592
GU946945 534 ..........TT.....C.....G..... 592
GU946939 534 .....C..... 592
GU946904 534 .....C..... 592
GU946901 534 .....C.....C..G..... 592
GU946890 530 .....C.....C..G..... 588
GU946872 534 .....C.....G..... 592
GU946851 534 .....C.....G..... 592
GU946850 533 .....C.....G..... 591
GQ183035 7037 ..........G.....C.....G..... 7095
GQ183034 7030 .....C.....C..... 7088
GQ183025 7036 ..........T.....C..... 7094
GQ183024 7036 ..........T.....C..... 7094
EU556224 6855 .....C..... 6913
EU024643 1226 .....C..... 1284
EU024641 1226 .....G.....C.....G..... 1284
EU024639 1226 ....T..... 1284
DQ315670 7076 .....C.....C..G..... 7134
AY158066 1226 .....C.....G..... 1284
GU946873 534 .....C..... 591
KT626011 7043 .....C..... 7101
KT626008 6792 .....C..... 6850
KT626005 6768 .....C.....G.....T..... 6826
KC769584 7037 .....C.....C.....C..... 7095
JX441355 7033 .....C.....C..... 7091
JX050181 7027 .....C..... 7085
JN106518 503 .....C.....C..G..... 558
JN106510 503 .......G.....G..C.....C..... 558
JN106511 503 558
JN106508 503 .....C.....G..C..... 558
JN106505 503 .....C.....C..G..... 558
JN106501 503 ....T..... 558
JN106494 503 558
JN106489 503 .....C.....T..... 558
JN106488 503 .....C..... 558
JN106480 503 .....C.....C..... 558
JN106479 503 .....C.....T..... 558
JN106478 503 ....A..... 558
JN106477 503 .....C.....T..... 558
JN106475 503 558
JN106474 503 .....C.....C..G..... 558
JN106473 503 .....C.....C..G..... 558
JN106468 503 .....C..... 558
GU946949 534 .....C..... 592
GU946942 534 ..........A.....C.....C..G..... 592
GU946941 534 .....C....A.....C.....C..... 592
GU946899 534 .....C.....C..G..... 592
GU946898 534 .....C.....C..... 592
GU946895 534 ..........T.....C.....C..G..... 592


GU946893 519 577
GU946869 534 592
GU946866 534 592
GU946862 534 592
GU946859 534 592
GU946852 527 585
GU946841 534 592
GU946817 533 591
EU170328 602 660
EU170313 602 660
EU374210 627 685
AM933164 627 685
EU024642 1226 1284
EU024637 1226 1284
GU946894 534 591
EU024638 1226 1284
KT626010 6809 6867
KJ152442 6950 7008
JN106491 503 558
JN106481 503 558
JN106476 503 558
JN106466 503 558
HQ696573 6619 6677
HQ696571 6617 6675
GU946956 534 592
GU946955 534 592
GU946954 534 592
GU946951 534 592
GU946900 533 591
GU946897 534 592
GU946819 528 586
EU716175 7046 7104
AM933163 627 685
AM933161 627 685
EU077513 264 322
AJ005695 7058 7116
HQ696574 6617 6675
HQ696570 6617 6675
JN106522 503 558
JN106516 503 558
JN106515 503 558
JN106514 503 558
JN106513 503 558
JN106512 503 558
JN106509 503 558
JN106507 503 558
JN106506 503 558
JN106499 503 558
JN106472 503 558
JN106471 503 558
JN106467 503 558
HQ696577 6625 6683
HQ696576 7048 7106
GU946959 534 589
GQ183021 7020 7078
EU170312 602 660
AM933160 627 685
AF055846 7079 7137

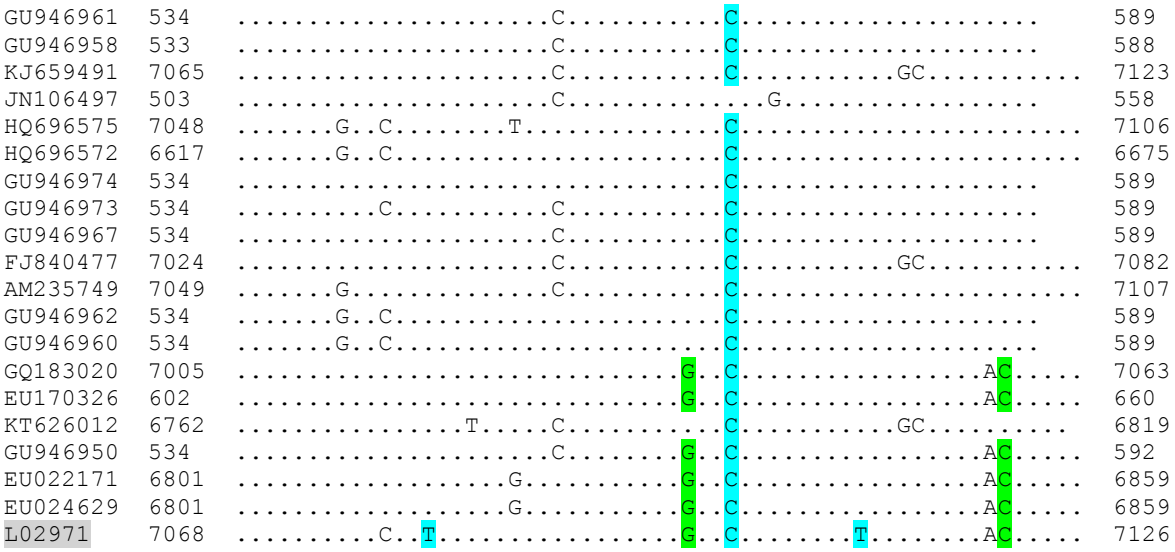


Figure 3.2: Alignment of the nucleotide sequences of 254 HPeV isolates in the region of the predicted structure. As HPeV1 Harris (L02971 and highlighted in grey) is relatively diverse from the other isolates, the sequence with accession number GU946968 was used in a Blast search. The results were reformatted to give the alignment, where dot (.) shows an identical nucleotide to the search sequence and differences are shown as letters. The predicted structure is indicated on the search sequence. Changes in predicted unpaired regions are not highlighted. Changes which maintain base-pairing through G-C/U or U-A/G interactions are highlighted in turquoise, covariant changes in green and changes which disrupt the structure in red. The highlighted regions were carried out manually.

Figure 3.3: Alignment of the sequences of HPeV1 Harris strain and Sebokele virus in the region of the predicted HPeV RNA structure. The predicted stem-loops are shown in grey, green and pink, while the kissing interaction is shown in turquoise. The areas of the predicted structure were highlighted manually.

3.2.2 **A promiscuous RNA structure in the 3'UTR of several picornaviruses**

A stem-loop structure has been identified in the 3'UTR of members of several different virus families with a positive sense RNA genome, including *Astroviridae* and *Coronaviridae* (Jonassen *et al.*, 1998b). The structure was found several years ago in the 3'UTR of the picornavirus equine rhinitis virus B and more recently in several newly discovered picornaviruses. The function of the structure, known as the stem-loop II-like motif (s2m) has not yet been identified. Another element was found to occur in the 3'UTR of more than one picornavirus genus including *Gallivirus* and *Kobuvirus* (Boros *et al.*, 2012). This is made up of a stem-loop where the stem and loop are variable, but two bulges have almost perfectly conserved sequences. To establish the distribution of this structure we used the Aichivirus (*Kobuvirus* genus) sequence to search the nr database at low stringency. This found other viruses with the same structure and the sequences of the structures in each of these was used to search the database to find the set of viruses shown in Figure 3.4. All the structures contain the sequence GATATAAAG in one bulge and CCCTAA in the other, with some more variable nucleotides. The stems are more variable between genera, but show some conservation. In addition to the picornaviruses, the structure was also found in two closely-related members of the caliciviruses. On the other hand, only two of the 6 genera and proposed genera of Kobuvirus have the structure.

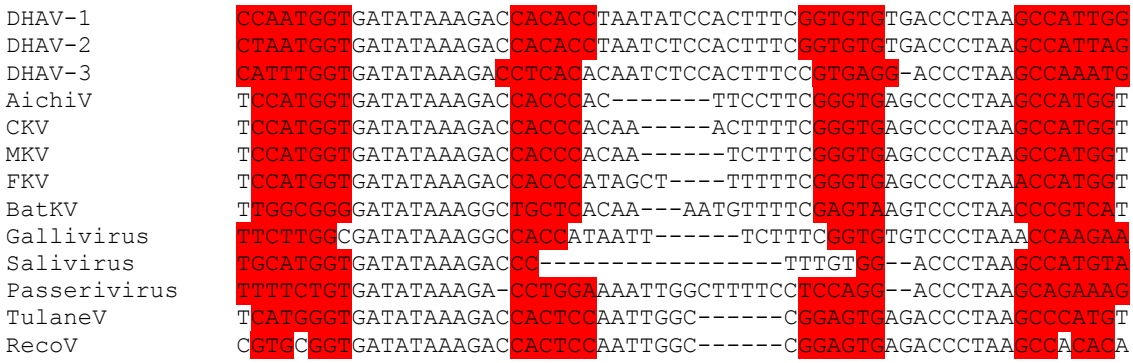


Figure 3.4: A conserved RNA structure found in several diverse picornaviruses and in Tulane virus and Recovirus, members of *Caliciviridae*. AichiV (Aichivirus), CKV (canine kobuvirus), MKV (murine kobuvirus) and FKV (feline kobuvirus) are members of the *Aichivirus A* species of *Kobuvirus*, while BatKV (bat kobuvirus) is a member of the proposed species “Aichivirus F”. DHVA-1, 2, 3 are genotypes of duck hepatitis A virus a member of the *Avihepatovirus* genus. The Aichivirus sequence was used to search the nr database using Blast, at low stringency (E value 100, wordlength 16). Each of the matches identified was used to research the database. The red highlights show predicted base pairs and this was done manually.

3.3 Constructs to investigate the role of the RNA structure in the 3D^{pol}-encoding region

Viral RNA stability was investigated indirectly by fluorescence techniques. A 518 bp region containing the RNA structure (named 3DS) in the 3D^{pol}-encoding region of the HPeV1 genome was ligated into pEGFP-C1 and pmCherry vectors between XhoI and BamHI sites to give pEGFP3DS and pMCh3DS respectively. A stop-codon is placed upstream of the virus cDNA to ensure that no extra amino acids are added to the vector fluorescent proteins (Figure 3.5 and Table 3.1). A pair of oligonucleotides was added to produce control constructs, pEGFPeC and pMCheC, which had 18 nucleotides added. These oligonucleotides are complementary and have XhoI and BamHI ends when annealed and a stop-codon to give the same C-terminus to EGFP and mCherry as that in the pEGFP3DS and pMCh3DS constructs. All the constructs made were sequenced and the results were confirmed to be correct (data not shown).

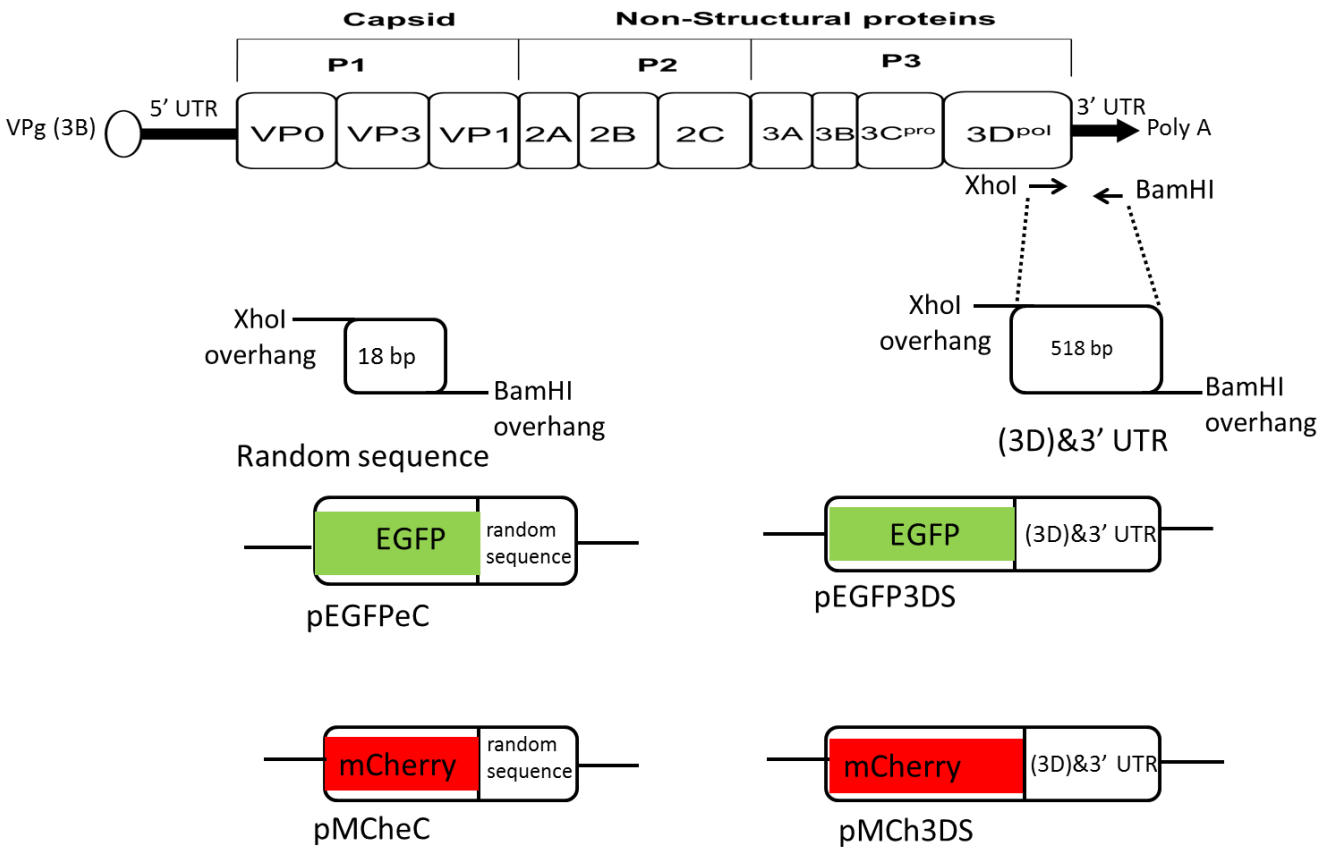


Figure 3.5: Schematic representation of constructs pEGFP3DS and pMCh3DS, which contain the RNA structure, 3DS, in a section of HPeV1 Harris cDNA including part of the $3D^{pol}$-encoding region and the 3' UTR ligated to pEGFP-C1 and pmCherry-C1 respectively. It also shows the controls, pEGFPeC and pMCheC, which gives the same C-terminus to EGFP and mCherry as that in pEGFP3DS and pMCh3DS.

Table 3.1: Description of each construct used for analysis, size and origin of PCR fragment ligated to cut pEGFP-C1 and pmCherry vectors.

Construct	Oligonucleotides	Description	DNA position/ size (bp)	Origin
pEGFPeC	OL2065 & OL2066	Randomly selected nucleotides to generate the same end to the EGFP protein as in the other constructs. Overlapping oligonucleotides were ligated.	0/ 18	Nil
pMCheC	OL2065 & OL2066	Randomly selected nucleotides to generate the same end to the MCherry protein as in the other constructs. Overlapping oligonucleotides were ligated.	0/ 18	Nil
pEGFP3DS	OL2063 & OL2064	Part of the $3D^{Pol}$ sequence (containing the structure to be analyzed) and the complete 3' UTR. A PCR product was ligated	6822 – 7339/ 518	HPeV1
pMch3DS	OL2063 & OL2064	Part of the $3D^{Pol}$ sequence (containing the structure to be analyzed) and the complete 3' UTR. A PCR product was ligated	6822 – 7339/ 518	HPeV1

3.3.1 **Fluorescence microscopy**

The DNA constructs were transfected with Lipofectin (section 2.2.2.1) into different cell lines and fluorescence from different channels were observed with a BX 41 fluorescence microscope (Olympus). The nuclei were stained with Hoechst stain and viewed with the blue filter, EGFP was viewed using the green filter and mCherry was viewed with the red filter.

The cell lines: A-549, Cama-1, MD-MBA-231 and PC3 were not effectively transfected (data not shown) hence results could not be analysed. However, GMK and MCF-7 gave sufficient transfected cells to be usefully observed. Results in different microscope fields from both cells showed similar fluorescence in cells transfected with all the constructs, but unexpectedly, pEGFPeC produced brighter fluorescence than pEGFP3DS (Figure 3.6 and 3.7). The EGFP fluorescence is variable in individual cells in a field which makes interpretation difficult. However, the results imply that the structure 3DS may not have a positive effect on viral RNA stability. If the RNA were more stable, each molecule should be translated more often and so stronger fluorescence should be seen.

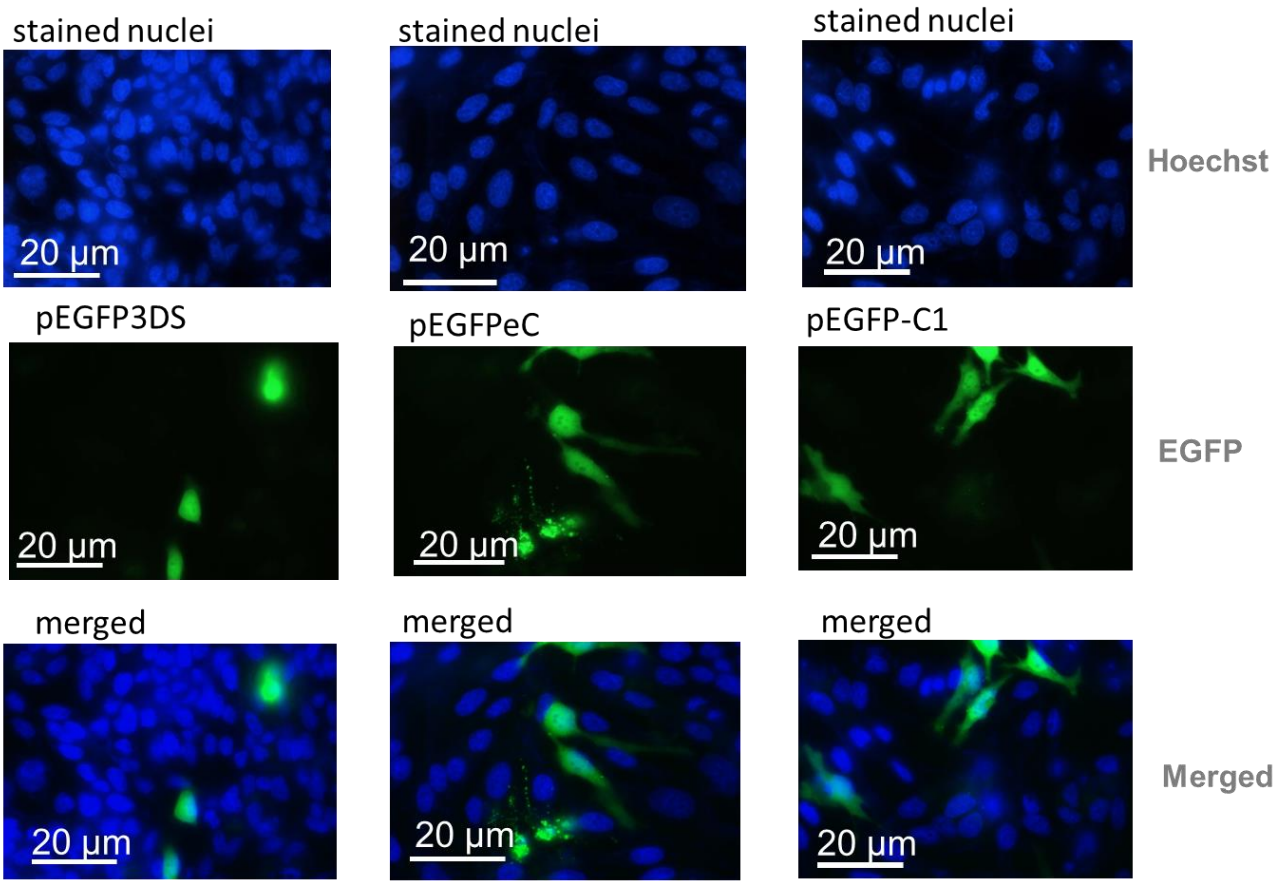


Figure 3.6: MCF-7 cells transfected with pEGFP3DS (green), pEGFPeC (green) and pEGFP-C1 (green) using Lipofectin. Cells were observed unfixed after treated with Hoechst for 30 min, 46 ± 2 hr after transfection, the examined using a BX 41 microscope at an exposure time of 5 ms and 500 ms for the blue (cell nucleus, Hoechst) and green (EGFP) channels respectively. The exposure times were kept constant for each individual channel during microscopy of all the fields. Scale: 20 µm. The selected images show typical results from 10 random fields from each construct.

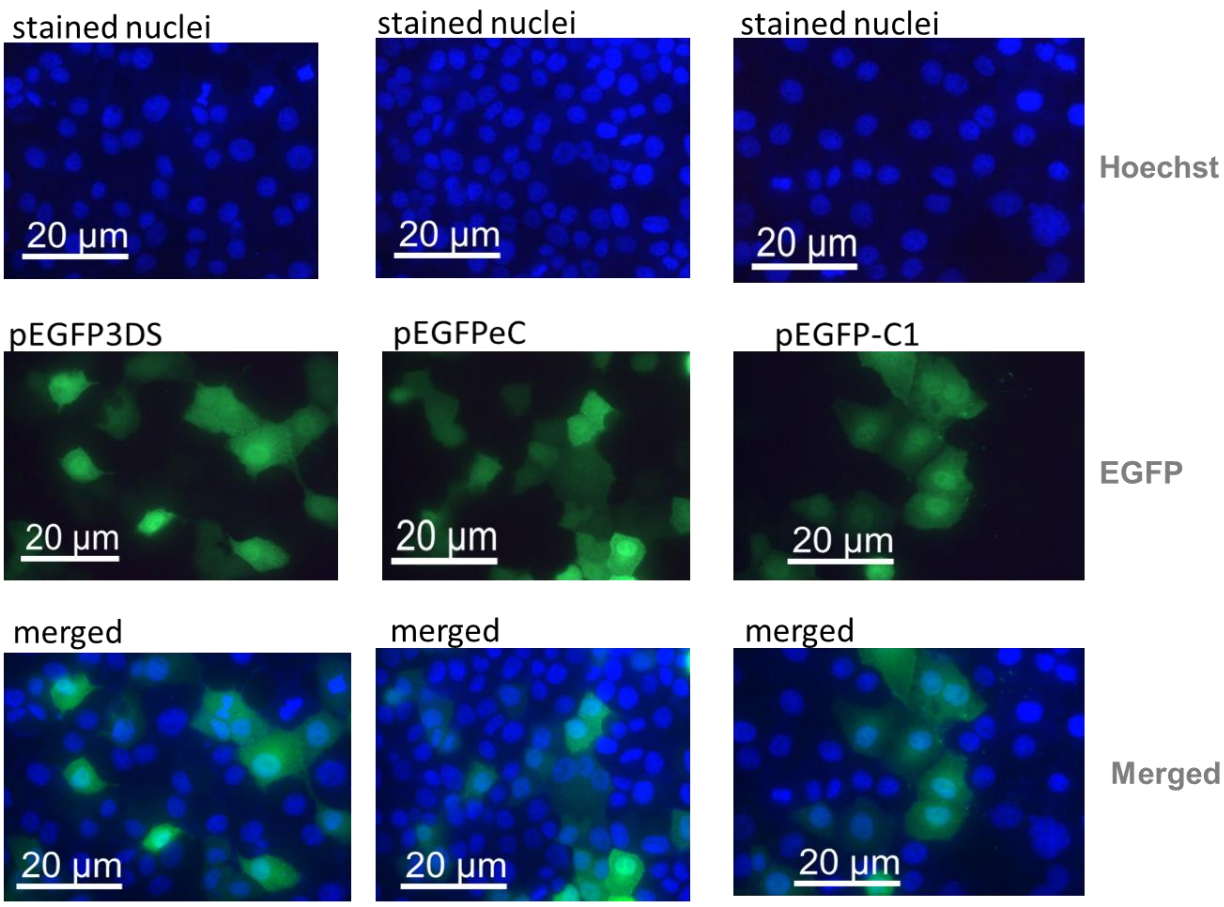


Figure 3.7: GMK cells transfected with pEGFP3DS (green), pEGFPeC (green) and pEGFP-C1 (green) using Lipofectin. Cells were observed unfixed after treated with Hoechst stain for 30 min, 46 ± 2 hr after transfection, the examined using a BX 41 microscope at an exposure time of 1 ms and 200 ms for the blue (cell nucleus, Hoechst) and green (EGFP) channels respectively. The exposure times were kept constant for each individual channel during microscopy of all the fields. Scale: 20 µm. The selected images show typical results from 10 random fields from each construct.

3.3.1.1 Co-transfection of pEGFP3DS and pEGFPeC with pmCherry-C1

To have a standard for comparing the microscopy results obtained from pEGFP3DS and pEGFPeC, pEGFP3DS was co-transfected with pmCherry-C1 and pEGFPeC was also co-transfected with pmCherry-C1.

The results show that pEGFPeC and pEGFP-C1 always produced relatively higher EGFP fluorescence when compared with the mCherry fluorescence from pmCherry-C1 than pEGFP3DS where the mCherry signal was more pronounced than the EGFP fluorescence produced (Figure 3.8).

Generally, the mCherry signal was very poor and this posed a problem during microscopy, also high mCherry signals degraded at a relatively higher rate than EGFP to about half its original fluorescence while the image was being adjusted to the right focus for capture.

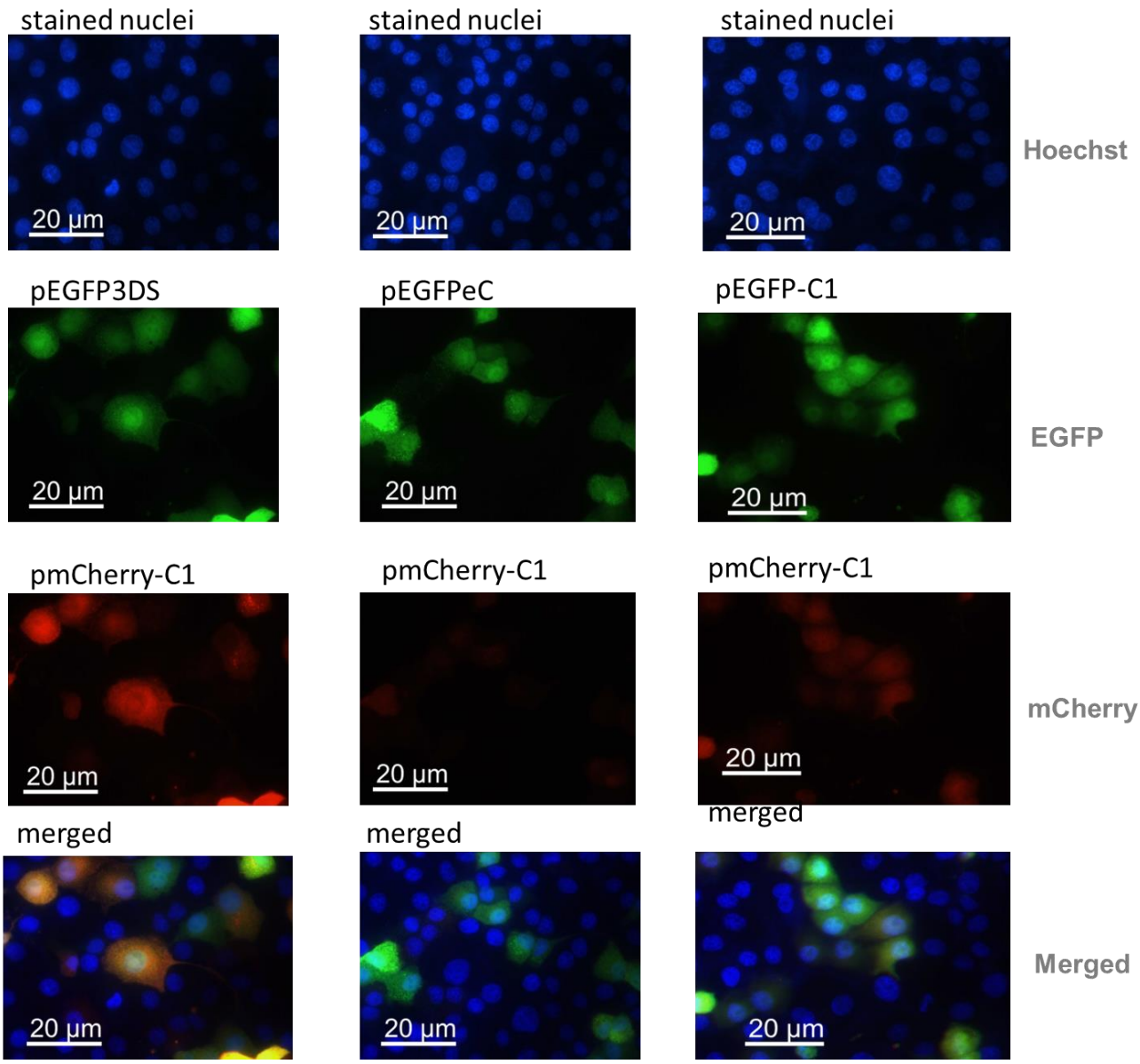


Figure 3.8: Co-transfection of pmCherry-C1 vector (red) with pEGFP3DS (green), pEGFPeC (green) and pEGFP-C1 (green) into GMK cells using Lipofectin. Cells were observed unfixed after treated with Hoechst stain for 30 min, 46 ± 2 hr after transfection. A BX 41 microscope was used at an exposure time of 1 ms, 200 ms and 10 ms for the blue, green and red channels respectively. The exposure times were kept constant for each individual channel during microscopy of all the fields. Scale: 20 µm. The selected images show typical results from 10 random fields from each construct.

3.3.1.2 Co-transfection of pMCh3DS and pMCheC with pEGFP-C1

The effect of the structure, 3DS, was analysed using another set of co-transfections. The constructs pEGFP3DS and pEGFPeC were reproduced in pmCherry-C1 to give pMCh3DS and pMCheC respectively and then co-transfected with pEGFP-C1 (Figure 3.9). Again, if the 3DS has a positive effect on the stability of the RNA, then higher red fluorescence should be produced. However, because of the low red fluorescence generally obtained in the mCherry signal, microscopy of the mCherry constructs itself was a challenge.

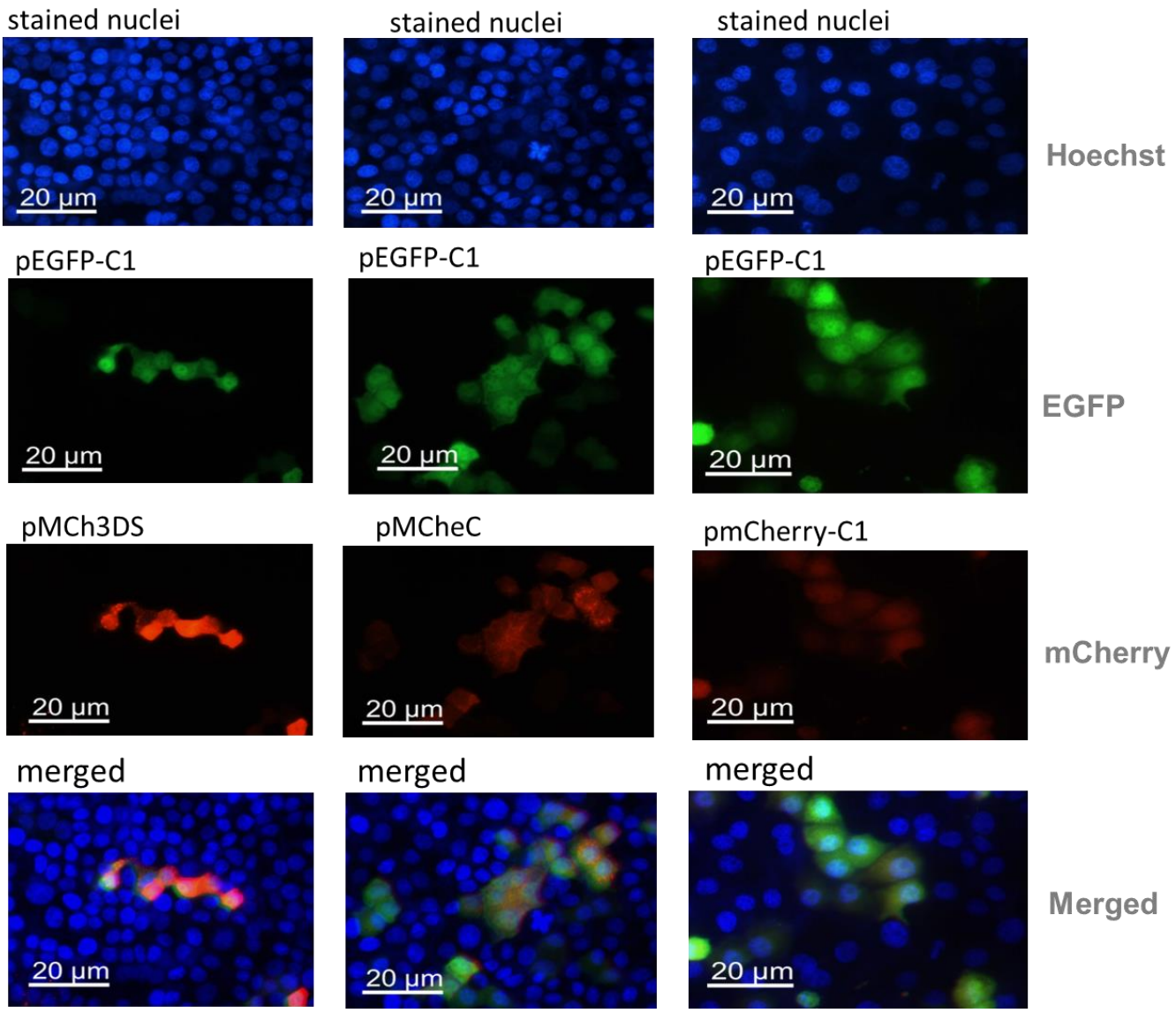


Figure 3.9: Co-transfection of pEGFP-C1 (green) with pMCh3DS (red) and pMCheC (red) into GMK cells using Lipofectin. Cells were observed unfixed after treated with Hoechst stain for 30 min, 46 ± 2 hr after transfection. A BX 41 microscope was used at an exposure time of 1 ms, 200 ms and 10 ms for the blue, green and red channels respectively. The exposure times were kept constant for each individual channel during microscopy of all the fields. Scale: 20 µm. The selected images show typical results from 10 random fields from each construct.

3.3.1.3 Co-transfection of pEGFP3DS with pMCheC and pEGFPeC with pMCh3DS

A final set of co-transfections investigated the fluorescence produced by pEGFP3DS and pEGFPeC when co-transfected with pEGFP3DS with pMCheC and pEGFPeC with pMCh3DS (Figure 3.10). This was done to also compare the fluorescence of the EGFP constructs, pEGFP3DS and pEGFPeC, with the mCherry constructs, pMCheC and pMCh3DS. Again, pEGFPeC produced a relatively high signal when compared to the mCherry fluorescence in pMCh3DS. It was higher than pEGFP3DS when also compared with the mCherry signal in pMCheC. The controls pEGFP-C1 and pmCherry-C1 have relatively the same amount of fluorescence produced (Figure 3.10). Still because of the low red fluorescence generally obtained in the mCherry constructs, microscopy of the co-transfection was quite a challenge.

Generally, the results showed some variation in the expression level of EGFP/MCherry when the RNA includes the 3DS RNA structure.

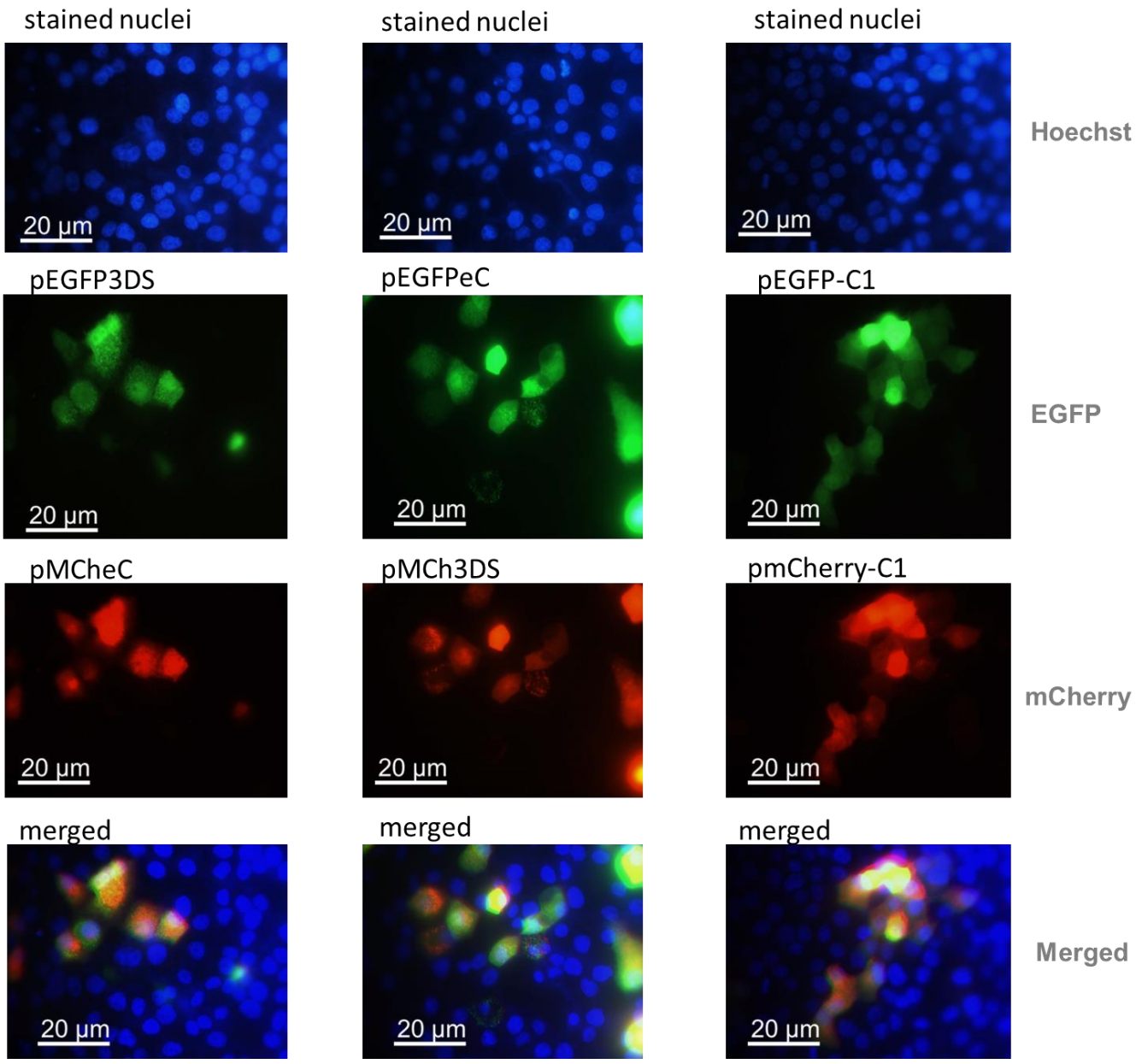


Figure 3.10: Co-transfection of pEGFP3DS (green) with pMCheC (red), pEGFPeC (green) with pMCh3DS (red) and pEGFP-C1 (green) with pmCherry-C1 whole vector into GMK cells using Lipofectin. Cells were observed unfixed after treated with Hoechst stain for 30 min, 46 ± 2 hr after transfection. A BX 41 microscope was used at an exposure time of 1 ms, 20 ms and 10 ms for the blue, green and red channels respectively. The exposure times were kept constant for each individual channel during microscopy of all the fields. Scale: 20 µm. The selected images show typical results from 10 random fields from each construct.

3.3.2 **IRES driven translation experiment**

As picornavirus RNA translation is IRES-dependent, the structure may function in an IRES dependent manner. To study the possible effect of the 3DS on IRES-dependent translation, a bicistronic construct was produced with dsRed as the reporter of IRES-driven translation. This was done by adding an HPeV1-IRES dsRed fusion downstream of the GFP gene in pcDNA3.1/NT-GFP-TOPO, which will be called Topo vector in the rest of this thesis. Two other constructs were made, one included the 3DS region plus the HPeV1 3' UTR. To eliminate any effect of the HPeV1 3' UTR, the other just included the 3DS region.

The three bicistronic constructs, T5dsRed, T5dsRed3DS and T5dsRed3D (Figure 3.11), are made up of the Green Fluorescent Protein (GFP) gene, which is present in Topo vector as the first cistron and the 2nd cistron (dsRed) is controlled by the 5' UTR of HPeV1. It was speculated that if the 3DS structure enhances IRES driven translation, the dsRed signal should be increased (Figure 3.12).

The construct T5dsRed was provided by Dr. Shaia S.R. Almalki, while T5dsRed3DS and T5dsRed3D were made by overlap PCR using specific primers (Table 2.2) to join the 3DS region plus the HPeV1 3' UTR or just the 3DS region to the already made T5dsRed.

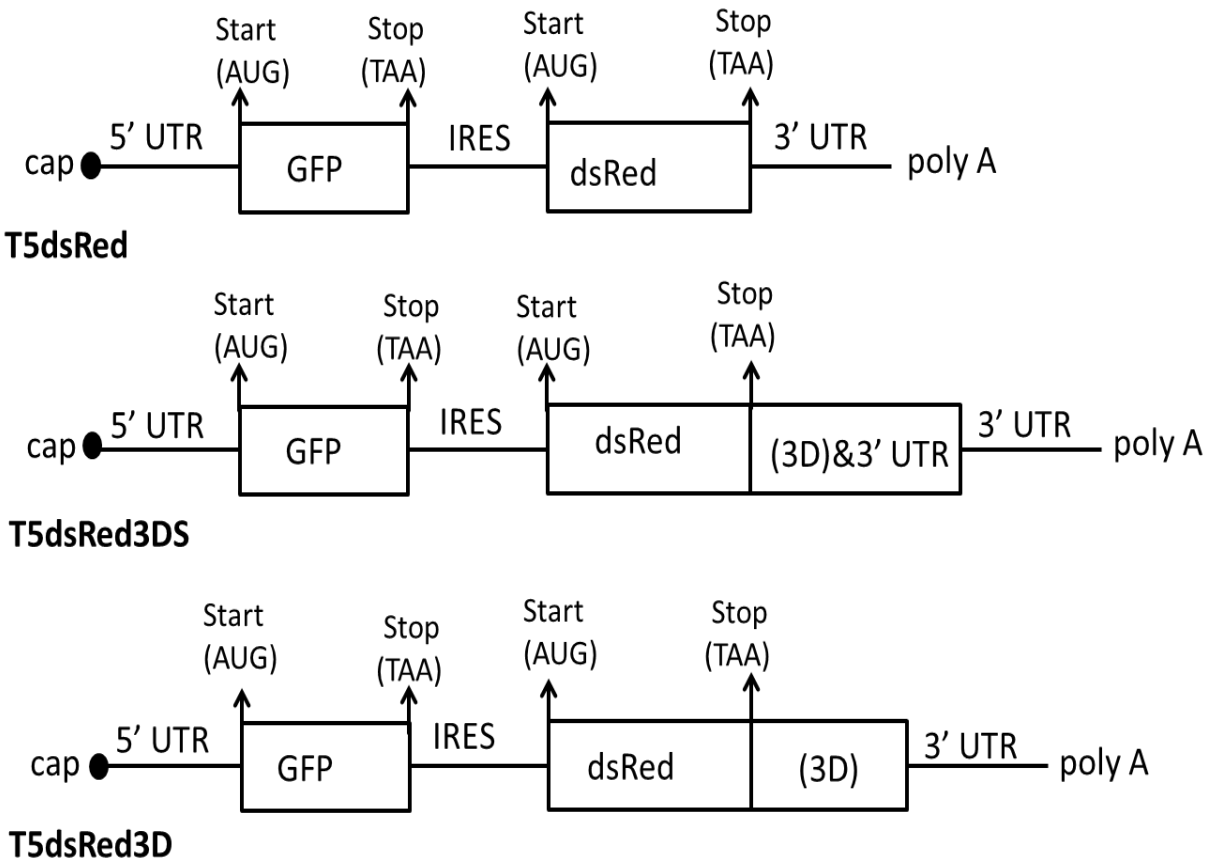


Figure 3.11: Schematic diagram to show the structure of RNA transcribed *in vivo* from constructs ligated to a TOPO GFP vector. T5dsRed is a construct consisting of the 5' UTR of HPeV1 joined to dsRed. T5dsRed3DS is T5dsRed joined to part of the 3D region (containing the 3DS structure) plus the 3' UTR from HPeV1 while T5dsRed3D consist of T5dsRed joined to the 3D region (containing the 3DS structure of interest) from HPeV1 without 3' UTR.

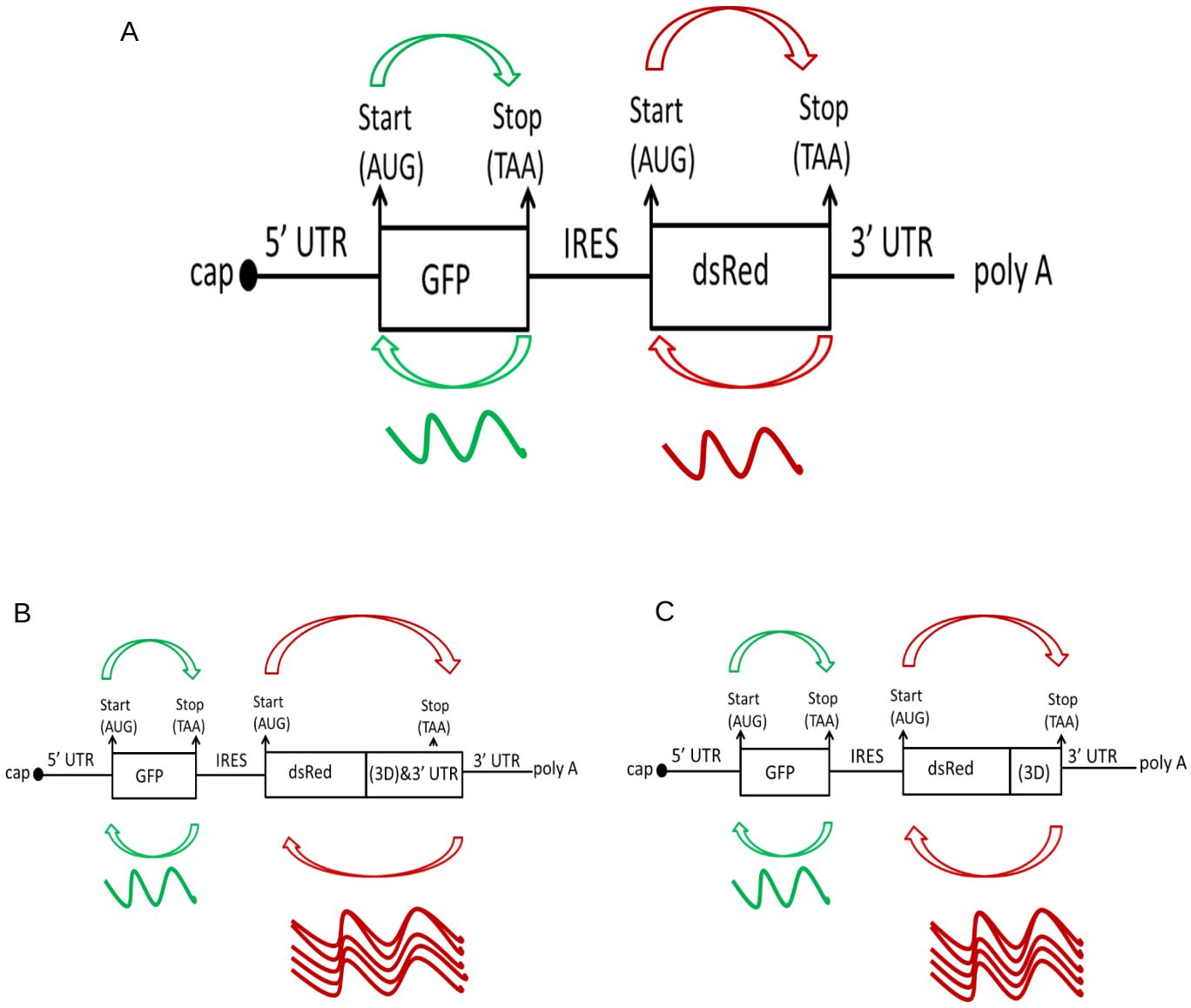


Figure 3.12: Theory behind the constructs made to analyse whether the 3DS structure enhances IRES-driven translation. A) is the original and control construct used to investigate if the addition of (3D)&3' UTR (which is part of the $3D^{Pol}$ containing the 3DS structure to be analyzed and the complete 3' UTR sequence) in B) or (3D) (which is the same part of the $3D^{Pol}$ but does not include the 3' UTR sequence) in C) would bring about the production of more dsRed.

The results showed no consistent significant difference between the dsRed signal from the control (T5dsRed) and the test constructs (T5dsRed3DS) and (T5dsREd3D) after viewing the cells in different fields of each slide (Figure 3.13). This suggests that IRES-driven translation is not greatly enhanced by the 3DS structure, but small differences could be missed.

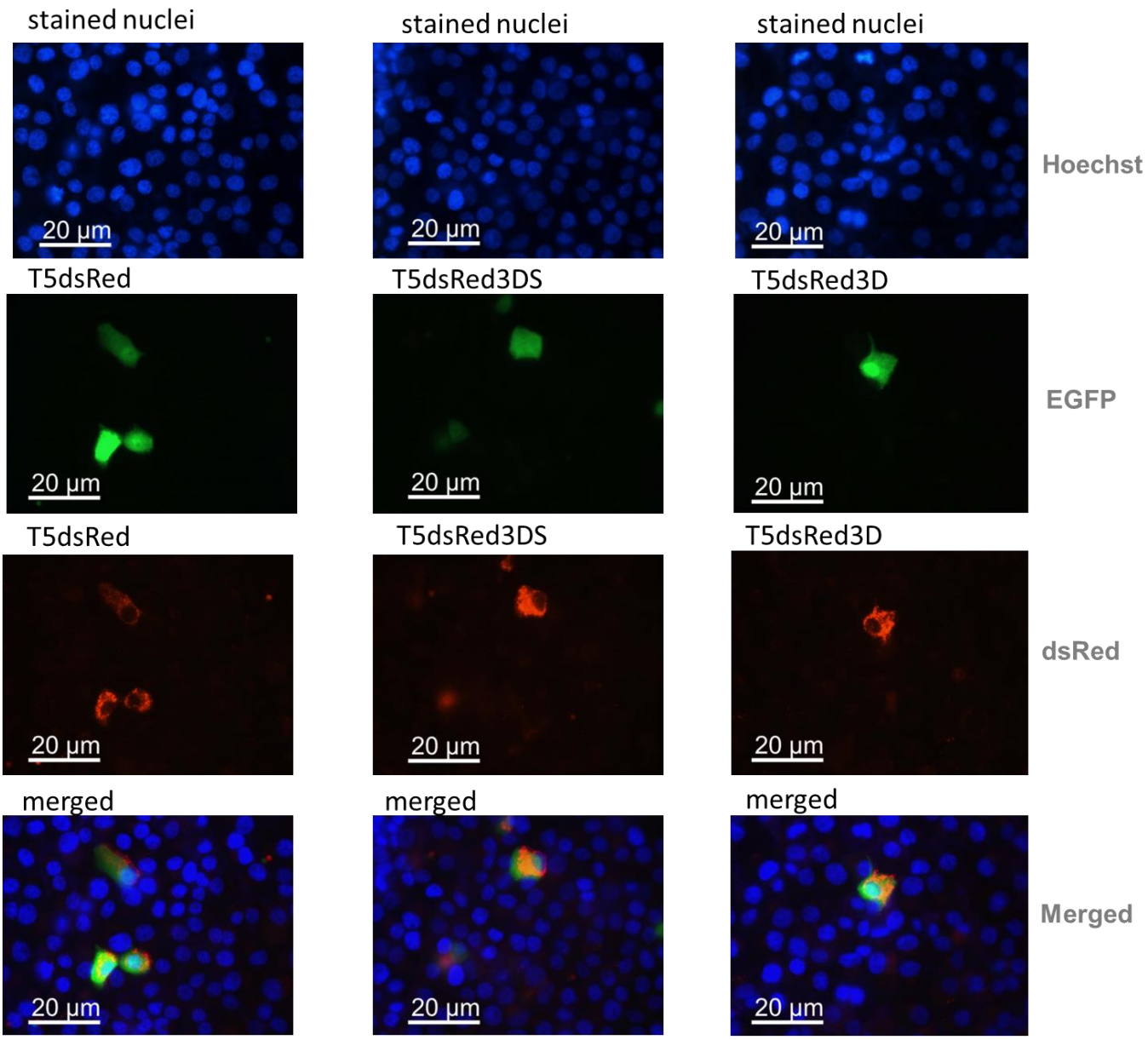


Figure 3.13: GMK cells were transfected with T5dsRed, T5dsRed3DS and T5dsRed3D using Lipofectin. Cells were treated with Hoechst stain for 30 min, 46 ± 2 hr after transfection, and the unfixed cells were examined using a BX 41 microscope at an exposure time of 1 ms, 500 ms and 1 s for the blue, green and red channels respectively. The exposure times were kept constant for each individual channel during microscopy of all the fields. Scale: 20 µm.

3.3.3 Analysis of different HPeV1 regions compared to the 3DS conserved structure

A full description of each DNA used is found in Table 3.2. The regions are all out of frame and all were preceded by a stop codon to ensure that no extra amino acids are added to the vector proteins. All the DNA constructs analysed therefore give the same C-terminus to EGFP, which is the same as that given by the control pEGFPeC. The DNA constructs were sequenced and each sequence was shown to be correct (data not shown).

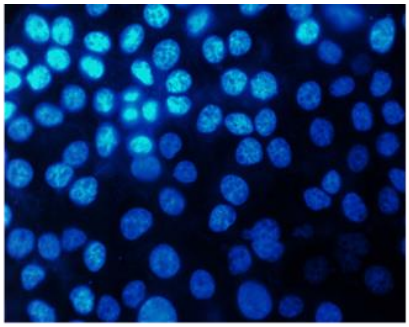
Table 3.2: Description of each construct used in the expanded set to study if the 3DS structure affects RNA stability. The size and origin of PCR fragments ligated to cut pEGFP-C1 vector are shown.

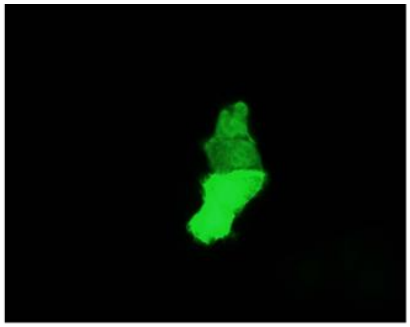
Construct	Primers (F & R)	Description	HPeV1 cDNA position/ size (bp)
pEGFPeC	OL2065 & OL2066	Randomly selected nucleotides to generate the same end to the EGFP protein as in the other constructs	0/ 18
pEGFP3DS	OL2063 & OL2064	Part of the $3D^{Pol}$ sequence (containing the 3DS structure to be analyzed) plus the complete HPeV1 3' UTR	6822 – 7339/ 518
pEGFPN3	OL2063 & OL2154	Part of the $3D^{Pol}$ sequence (containing the 3DS structure to be analyzed) without the HPeV1 3' UTR	6822 – 7249/ 428
pEGFPS	OL2155 & OL2156	Shortened $3D^{Pol}$ region, containing only the 3DS structure	7008 – 7126/ 119
pEGFPVP1	OL2149 & OL2150	Part of the HPeV1 VP1-encoding region equal in size to the pEGFP3DS insert	2351 – 2868/ 518
pEGFP3UTR	OL2151 & OL2064	Complete HPeV1 3' UTR	7250 – 7339/ 90

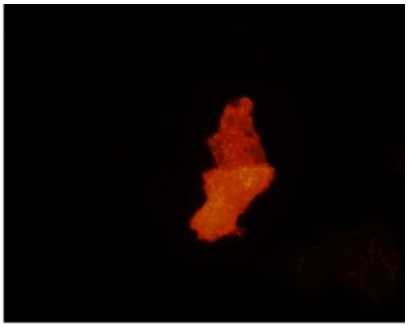
3.3.4 The highly conserved structure in the 3D^{pol} might affect viral RNA stability

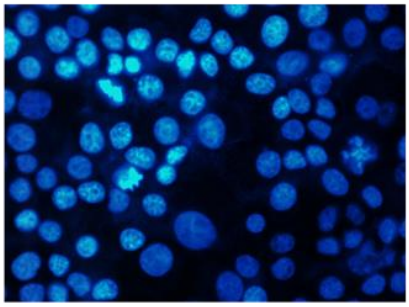
It was decided to increase the range of constructs available. All those used are listed in Table 3.2. The insertion of a large (518 nucleotides) sequence into the 3' UTR of the EGFP construct is the construct pEGFP3DS and it may be itself affecting translation efficiency. To control for this, a section of the VP1-encoding region of the same size was randomly chosen and introduced to pEGFP-C1 to make the construct pEGFPVP1. The 518 nucleotide sequence in pEGFP3DS contains the whole of the HPeV 3' UTR as well as part of the 3D^{pol} region including 3DS. To investigate if the HPeV 3'UTR may be complicating the results, two separate constructs were made including only the HPeV 3'UTR (pEGFP3UTR) and only the 3D^{pol} region present in pEGFP3DS (pEGFPN3). Finally, the minimal 3DS sequence of 119 nucleotides was introduced to give the construct pEGFPS.

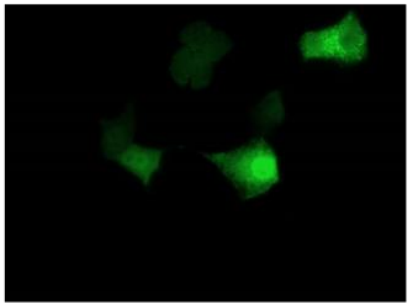
The DNA constructs were co-transfected with pmCherry-C1 into to GMK cells and fluorescence from different channels were observed with a fluorescence microscope (Figure 3.14). The EGFP fluorescence is variable in individual cells in a field which makes interpretation difficult and the results showed similar fluorescence in cells transfected with all the constructs except that the EGFP signal in pEGFP3DS and pEGFPN3 appeared to be lower compared to all other constructs. The results therefore suggest that the 3DS structure is not enhancing RNA stability and giving an increase in fluorescence through increased levels of EGFP.

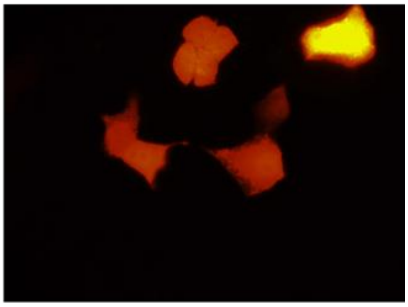

stained nuclei

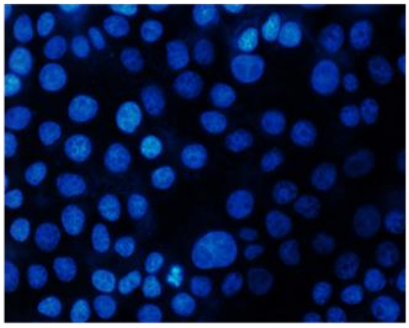

pEGFP-C1

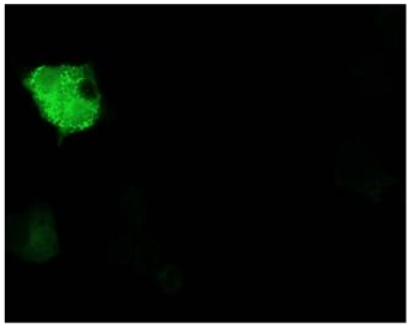

pmCherry-C1

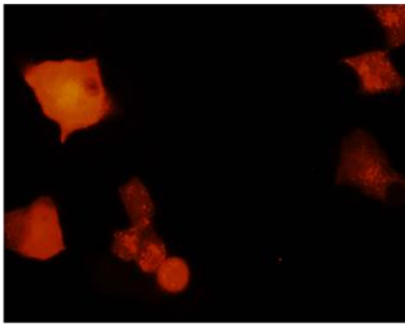

stained nuclei

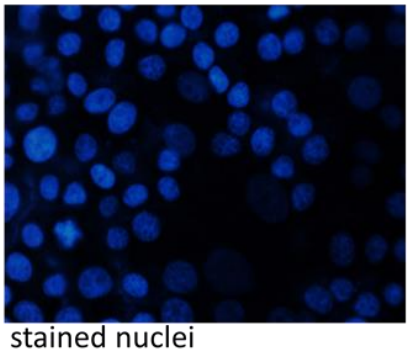

pEGFP3UTR

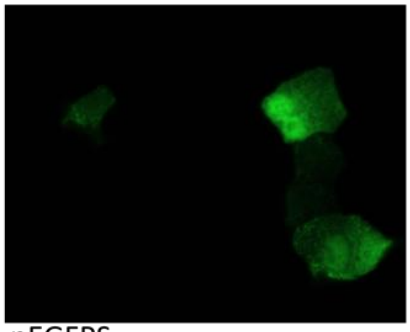

pmCherry-C1

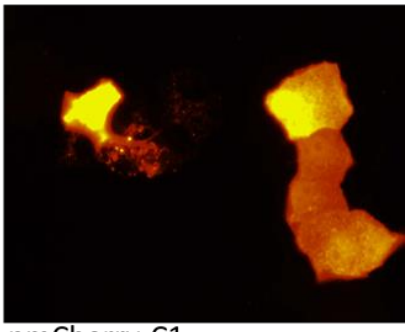

stained nuclei


pEGFPeC


pmCherry-C1


stained nuclei


pEGFPS


pmCherry-C1

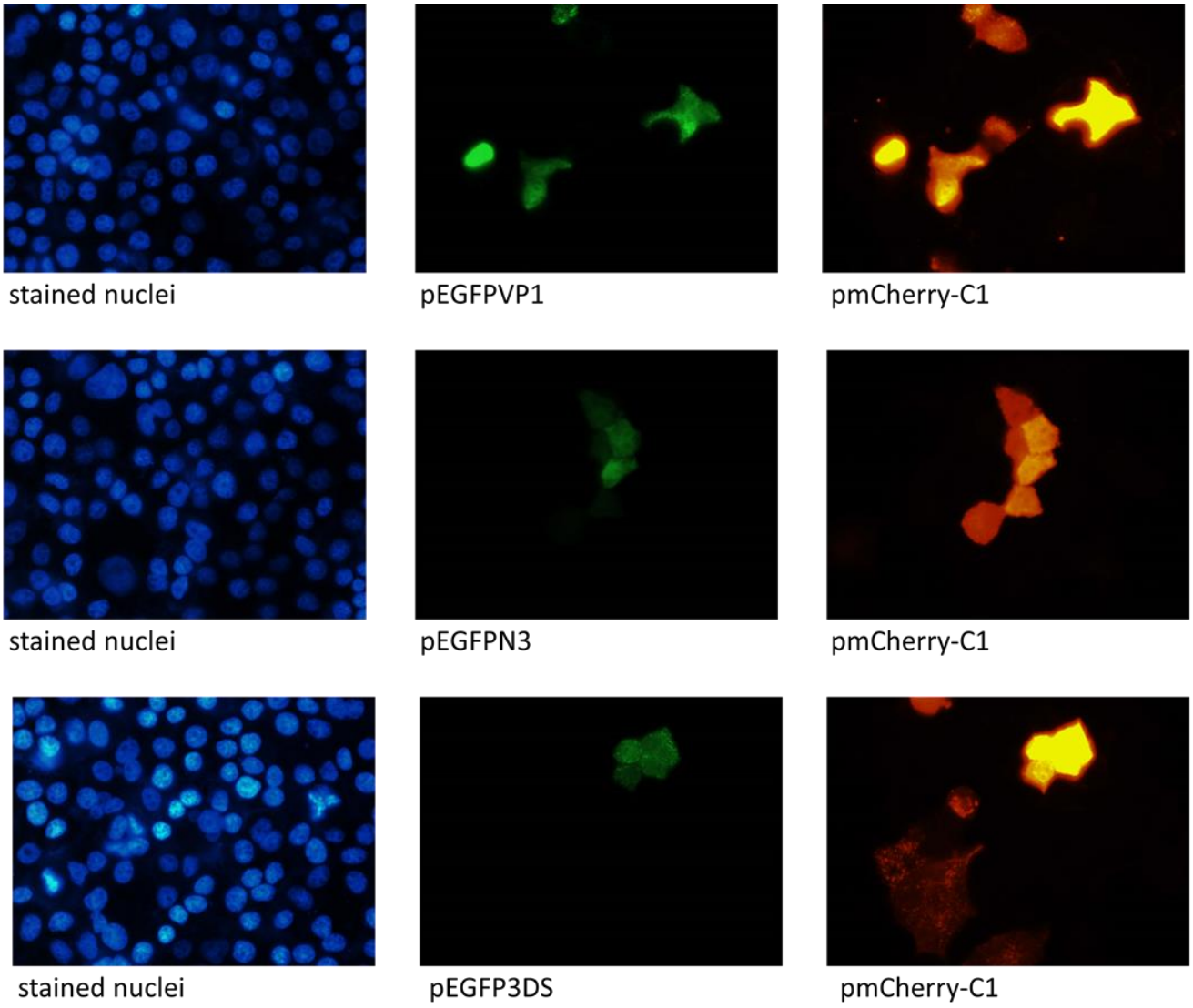


Figure 3.14: Co-transfection of pmCherry-C1 (red) with pEGFP-C1, pEGFP3UTR, pEGFPeC, pEGFPS, pEGFPVP1, pEGFPN3 and pEGFP3DS respectively into GMK cells using Lipofectin. The constructs containing EGFP were viewed from the green channel of the microscope. Cells were observed unfixed treated with Hoechst stain for 30 min, 46 ± 2 hr after transfection. A BX 41 microscope was at an exposure time of 2 ms, 20 ms and 100 ms for the blue, green and red channels respectively. The exposure times were kept constant for each individual channel during microscopy of all the fields. Scale: 20 µm. The selected images show typical results from 10 random fields from each construct.

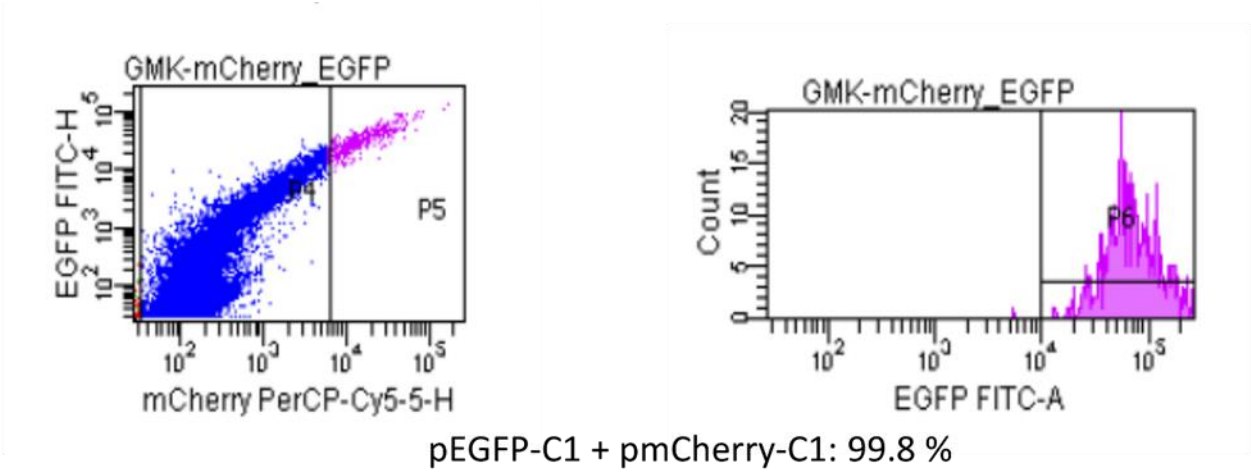
3.4 Investigation of fluorescence using FACS analysis

The variation of fluorescence between transfected cells in a field makes it difficult to see if there is a difference between the constructs. Any difference may be relatively subtle and not obvious by fluorescence microscopy. A potentially more sensitive and less subjective measure of fluorescence was therefore used.

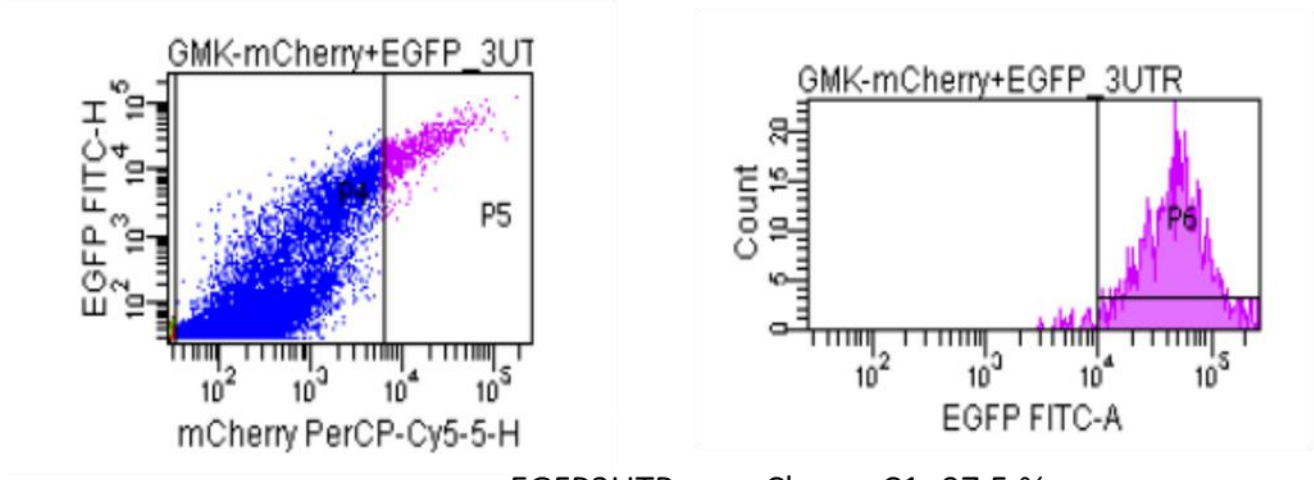
To quantitate the fluorescence, FACS analysis was performed using a FACS Aria flow cytometer. The constructs were co-transfected with pmCherry-C1.The theory behind co-transfecting the constructs in EGFP with pmCherry-C1 is that transfected cells should take up similar amounts of the EGFP and mCherry vector. However, if RNA transcribed from an EGFP construct has increased stability, then higher fluorescence should be produced. Figure 3.15 shows a typical example of how the flow cytometry assay was quantified using the FITC for green on the y-axis (which measured EGFP) and PerCP.Cy5.5 on the x-axis (which measured mCherry). P4 and P5 population show dot plot for double positives of cells expressing both EGFP and mCherry at different levels (10^2, 10^3, 10^4 and 10^5). Gating was thus done at 10^4 to find out the percentage cells expressing very bright fluorescence of EGFP. The P6 population shows the histogram plot of the gating at 10^4.

From Figure 3.15, it is noticed that the minimal 3DS construct, pEGFPS, gives a high fluorescence (96.8%), but not significantly greater than the control pEGFPeC. It is also noticeable that the construct pEGFP3DS gives the lowest fluorescence (51.3%), however, the removal of the 3’ UTR from pEGFP3DS enhances fluorescence (63.6% for the construct pEGFPN3). The construct with the VP1 region insertion (pEGFPVP1) gives a lower fluorescence than pEGFPS but this is greater than pEGFP3DS and pEGFPN3. The

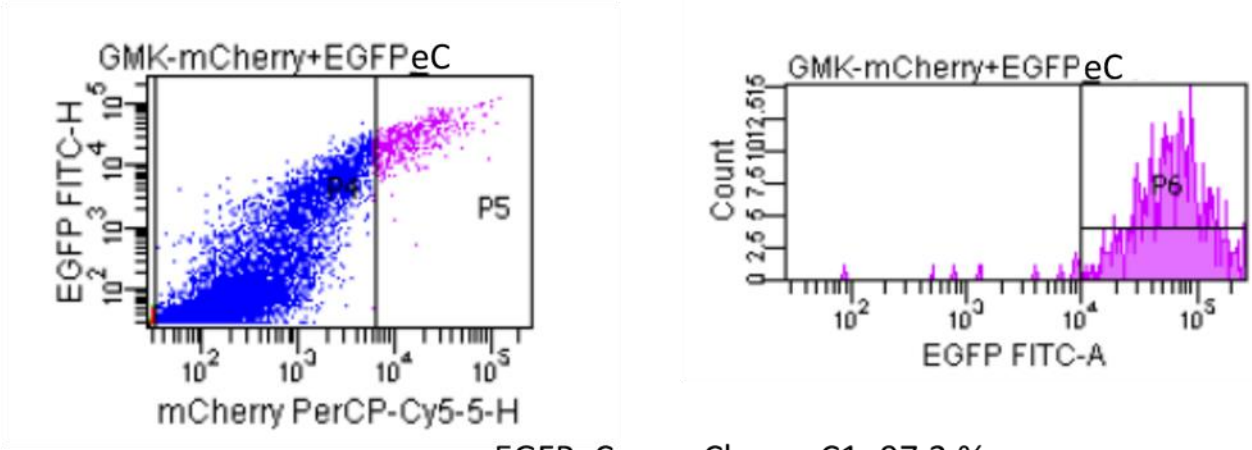
3’ UTR of HPeV1 construct, pEGFP3UTR, gave the highest fluorescence (97.5%) similar to the control. A bar chat (Figure 3.16) represents the result of the mean fluorescence produced by each of the constructs in 3 independent experiments.



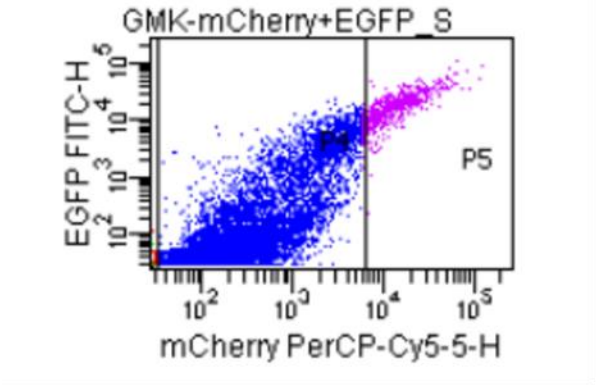
pEGFP-C1 + pmCherry-C1: 99.8 %

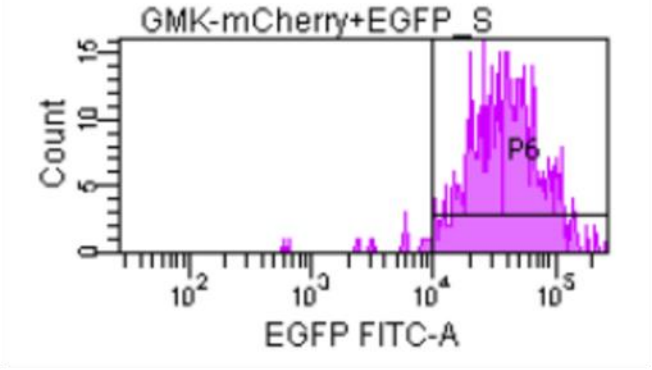


pEGFP3UTR + pmCherry-C1: 97.5 %

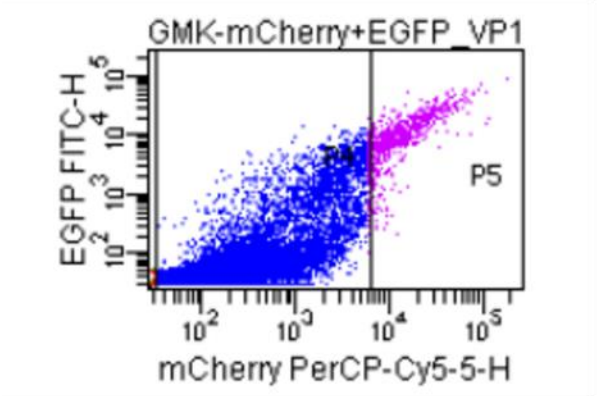


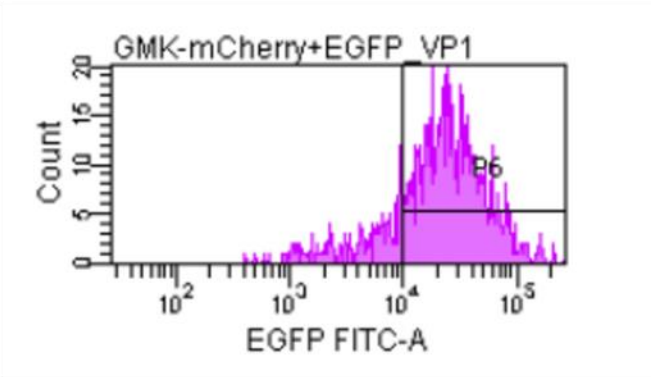
pEGFPeC + pmCherry-C1: 97.2 %



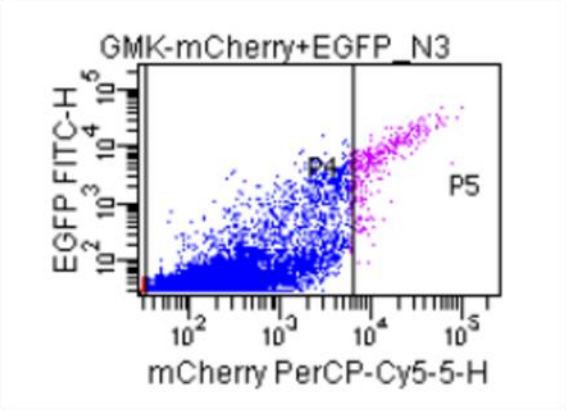


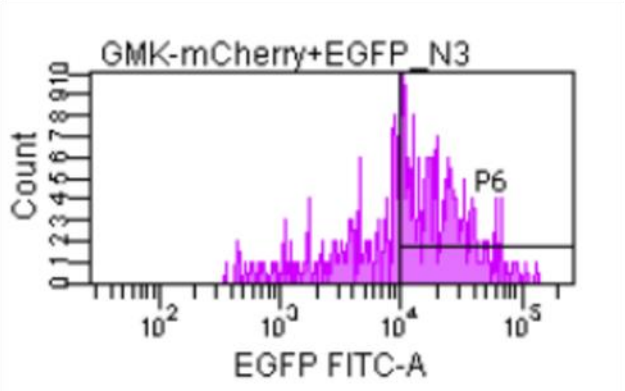
pEGFPS+ pmCherry-C1: 96.8 %





pEGFPVP1+ pmCherry-C1: 81.6 %





pEGFPN3+ pmCherry-C1: 63.6 %

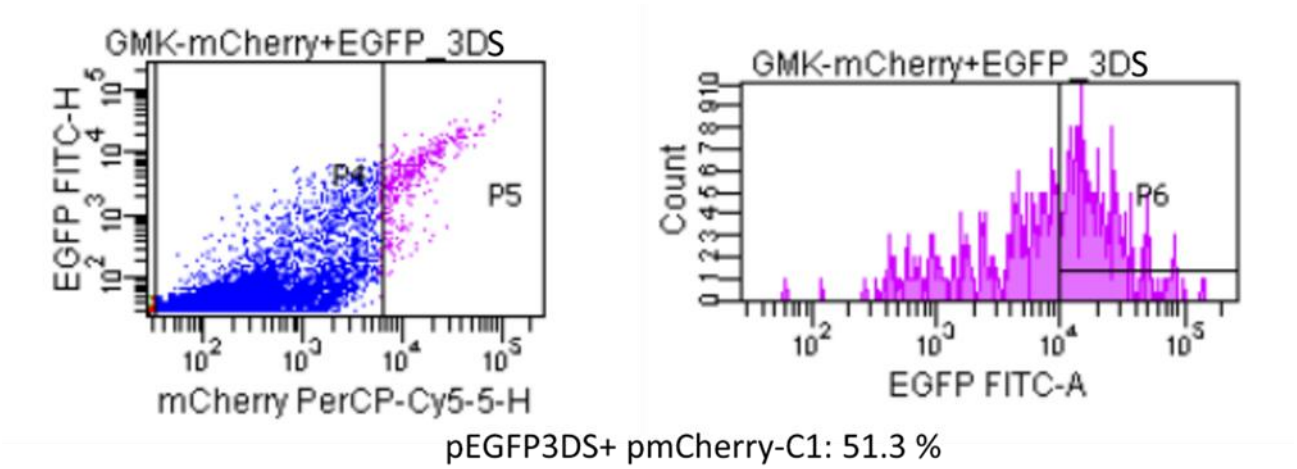


Figure3.15: Flow cytometry analysis of constructs after co-transfection of pmCherry-C1 with pEGFP-C1, pEGFP3UTR, pEGFPeC, pEGFPS, pEGFPVP1, pEGFPN3 and pEGFP3DS respectively into GMK cells using Lipofectin. Cells were analysed using a FACS Aria 46 ± 2 hr after transfection. The blue region (P4) represents living cells transfected with the EGFP construct and MCherry vector while the purple region (P5) shows gating of fluorescence above 10^4 of double positive cells (EGFP construct and MCherry whole vector). The percentages (P6) represent the number of bright EGFP fluorescence events present in the analysed sample after gating at 10^4.

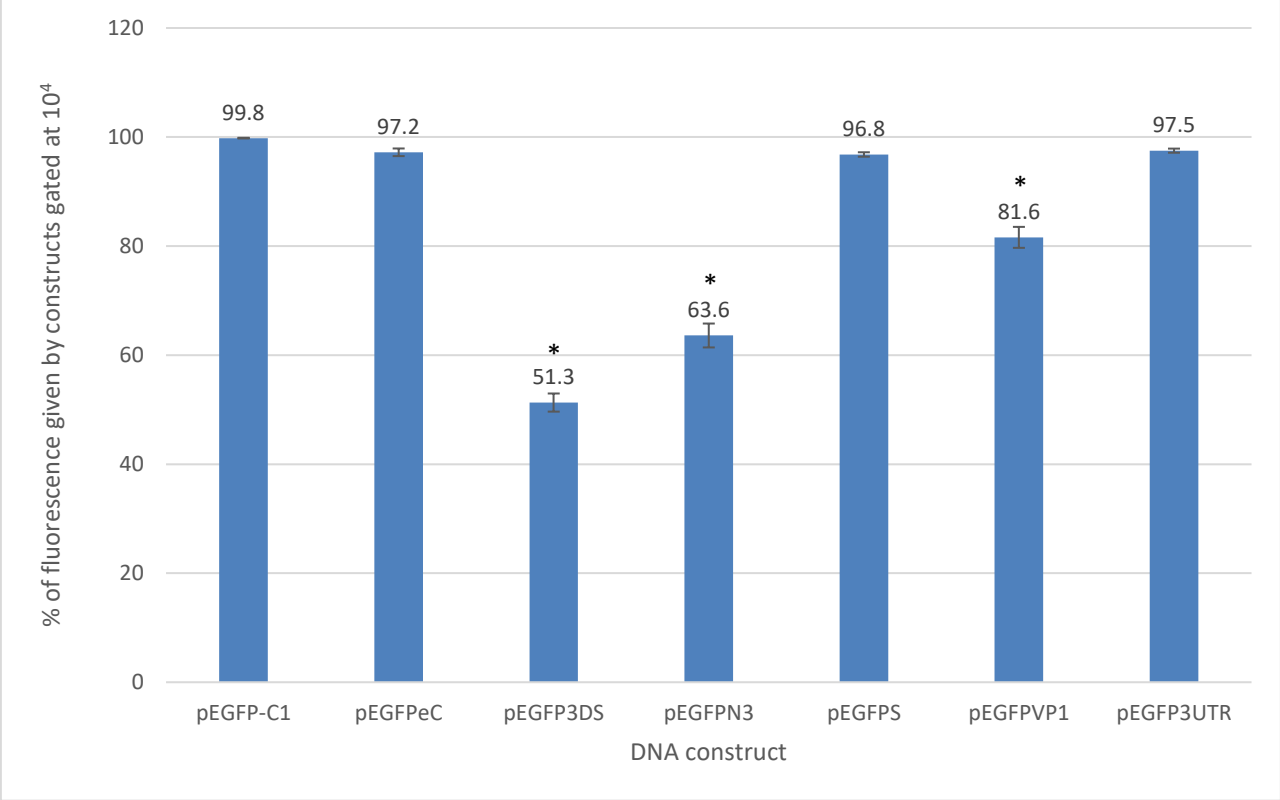


Figure 3.16: Percentage fluorescence given by constructs gated at 10^4 using Fluorescence Activated Cell Sorter (FACS). The constructs; pEGFP-C1, pEGFPeC, pEGFP3DS, pEGFPN3, pEGFPS, pEGFPVP1, and pEGFP3UTR were co-transfected into GMK cells using Lipofectin. The cells were co-transfected with 1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed 46 ± 2 hr after transfection. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates p < 1.8E-03

3.4.1 **Investigation via comparison of DNA construct sizes**

As the longer constructs give lower fluorescence, this was further investigated. Segments of different viral genomes were ligated into pEGFP-C1 vector, also between XhoI and BamHI sites as for the previously made DNA constructs. They also thus give exactly the same C-terminus to EGFP except for pEGFP-C1 itself. Table 3.3 gives a complete list of the constructs used in these experiments.

The constructs were explored to investigate the relation between the fluorescence of EGFP produced by the cells after transfection and the size of the DNA because it is probable that longer lengths of DNA fragments brought about a reduction in fluorescence of the EGFP protein as seen in the results so far (Figure 3.17). In addition, the inserts had different virus origins. A random section of 518 nucleotides was used from both CAV9 and HRV1B. HRV1B, like HPeV1, has a relatively high A+U content, while CAV9 has a lower A+U content. Also, cellular mRNA stability is often influenced by sequences in the 3' UTR. Apart from the HPeV1 3' UTR construct, constructs containing the CAV9 and HRV1B 3' UTR were also made. Finally, a construct containing the Aichivirus stem-loop structure which is conserved in the 3' UTR of several diverse picornaviruses was also produced.
The DNA constructs were co-transfected along with pmCherry-C1 to GMK cell line. Gating was done on mCherry positive cells because cells that take in pmCherry-C1 should also take in pEGFP-C1. The mean fluorescence of EGFP present in the cells for each construct was automatically generated from the FACS. The normalization to pEGFP-C1 + pmCherry-C1 was carried out by dividing the amount of fluorescence obtained from each DNA construct by the amount of fluorescence obtained from pEGFP-C1 + pmCherry-C1.

The mean of the results for three repeats of the experiment was used to plot the graph for constructs derived from different viruses and consisting of different sizes [1000 bp (2 constructs), 518 bp (4 constructs), 428 bp (1 construct), 160 bp (1 construct), 90 bp (1 construct), 56 bp (1 construct), 40 bp (1 construct) and 18 bp (1 construct) (Table 3.3)] and it is shown in Figure 3.17.

It can be seen that there is a large difference in fluorescence obtained from the different constructs. A scatter plot of the fluorescence against size of insert showed that generally, the larger the insert, the lower the fluorescence (Figure 3.18). All the constructs with the 518 nucleotides insertion (pEGFP3DS, pEGFPVP1, pEGFPHRV(3D5) and pEGFPCAV9(3D)) gave similar fluorescence and so the 3DS structure did not significantly affect the results. The shorter insert constructs, gave the highest level of fluorescence and the CAV9 3' UTR and HPeV1 3' UTR constructs were similar. Interestingly, the Aichi structure construct gave a high fluorescence, but this was similar to the control. In contrast, the HRV1B 3' UTR gave greatly enhanced fluorescence.

Table 3.3: List of constructs made to study the effect of size of insert on EGFP fluorescence in the assay. Forward & reverse primers, description of each construct used for analysis and size of PCR fragment ligated to cut pEGFP-C1 vector are shown.

Construct	Oligonucletides (F & R)	Description	DNA position/ size (bp)	Origin
pEGFPeC	OL2065 & OL2066	Randomly selected nucleotides to generate the same end to the EGFP protein as in the other constructs	0/ 18	Nil
pEGFP3DS	OL2063 & OL2064	Part of the $3D^{Pol}$ sequence (containing the 3DS structure to be analyzed) plus the complete HPeV1 3' UTR	6822 – 7339/ 518	HPeV1
pEGFPVP1	OL2149 & OL2150	Part of the HPeV1 VP1-encoding region equal in size to the pEGFP3DS insert	2351 – 2868/ 518	HPeV1
pEGFP3UTR	OL2151 & OL2064	Complete HPeV1 3' UTR	7250 – 7339/ 90	HPeV1
pEGFPVP3/1	OL2183 & OL2150	Parts of the HPeV1 VP3 and VP1-encoding region which gives 1000 bp	1869 – 2868/ 1000	HPeV1
pEGFPN3	OL2063 & OL2154	Part of the $3D^{Pol}$ sequence (containing the 3DS structure to be analyzed) without the HPeV1 3' UTR	6822 – 7249/ 428	HPeV1
pEGFPCAV9(3UTR)	OL2186 & OL2187	Complete CAV9 3' UTR	7347 – 7452/ 106	CAV9
pEGFPCAV9(3D)	OL2184 & OL2185	Part of the $3D^{Pol}$ -encoding region	6638 – 7155/ 518	CAV9
pEGFPHRV(3D5)	OL2188 & OL2189	Part of the $3D^{Pol}$ -encoding region equal in size to the pEGFP3DS insert	6072 – 6589/ 518	HRV
pEGFPHRV(3D1)	OL2190 & OL2191	Part of the $3D^{Pol}$ region	5715 – 6714/ 1000	HRV
pEGFPHRV(3UTR)	OL2192 & OL2193	Complete 3' UTR	7094 – 7133/ 40	HRV
pEGFPAichi(3UTR)	OL2200; OL2201 & OL2202; OL2203	Part of the 3' UTR of Aichi virus A which forms a predicted conserved structure (Figure 3.4)	8102 – 8156/ 56	Aichivirus

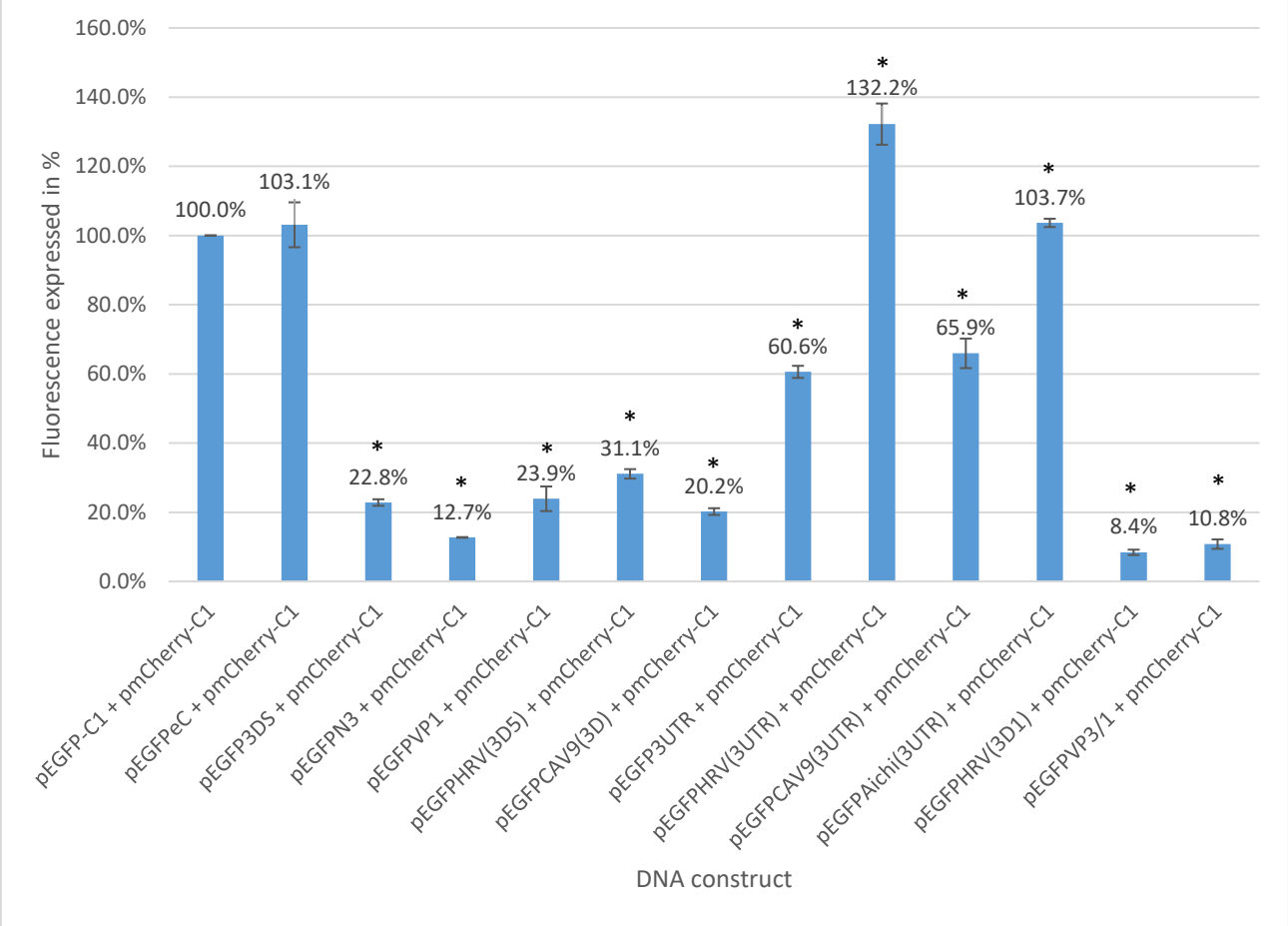


Figure 3.17: Mean fluorescence of EGFP constructs gated on mCherry fluorescence. GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 48 hr after transfection. Error bars show the Standard error (SE) from 3 independent experiments. * indicates $p < 3.8E-02$.

* indicates significant difference of all DNA constructs to the control, pEGFP-C1.

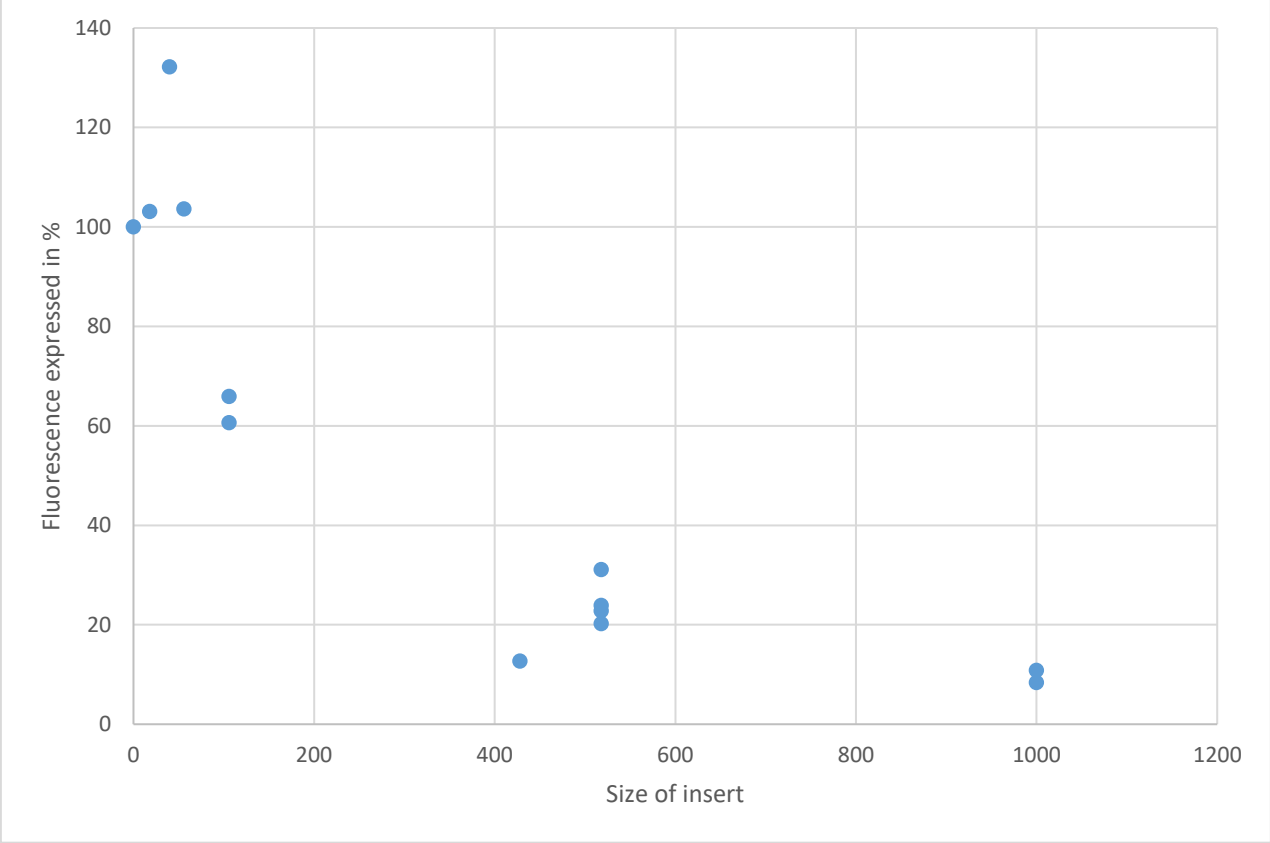


Figure 3.18: Scatter plot of the fluorescence expressed by each construct against size of insert.

3.4.2 **Investigation of the 'CG' dinucleotide in pEGFP3DS**

There is a conservation of CG dinucleotide at position 6935/6 and 6989/90 in the HPeV genome (Williams *et al.*, 2009). This was investigated using three other new constructs, pEGFP(CG1)3DS, pEGFP(CG2)3DS and pEGFP(CG1+2)3DS, made from pEGFP3D by introducing a silent mutation of 'A' to 'C' to determine if fluorescence in pEGFP3DS is affected by conserved 'CG' dinucleotide. Mutations were introduced using overlap PCR from the primers OL2196 & OL2197 for 1st mutation, OL2198 & OL2199 for the 2nd mutation and pEGFP(CG1+2)3DS was produced from another overlap PCR using pEGFP(CG1)3DS as template. Table 3.4 Show a summary of how the overlap PCR was carried out. The results are shown in Figure 3.19. Neither single mutant nor the double mutant gave different fluorescence compared to the original pEGFP3DS. This suggests that the CG dinucleotides do not influence RNA stability/translation.

Table 3.4: Summary of how overlap PCR was carried out to produce pEGFP(CG1)3DS, pEGFP(CG2)3DS and pEGFP(CG1+2)3DS.

Initial product	Template	Oligonucleotides	Final DNA construct
Fragment 1	HPeV1 cDNA	OL2063 + OL2197	
Fragment 2	HPeV1 cDNA	OL2096 + OL2064	
	Fragment 1 + fragment 2	OL2063 + OL2064	pEGFP(CG1)3DS
Fragment 1	HPeV1 cDNA	OL2063 + OL2199	
Fragment 2	HPeV1 cDNA	OL2198 + OL2064	
	Fragment 1 + fragment 2	OL2063 + OL2064	pEGFP(CG2)3DS
Fragment 1	pEGFP(CG1)3DS	OL2063 + OL2199	
Fragment 2	pEGFP(CG1)3DS	OL2198 + OL2064	
	Fragment 1 + fragment 2	OL2063 + OL2064	pEGFP(CG1+2)3DS

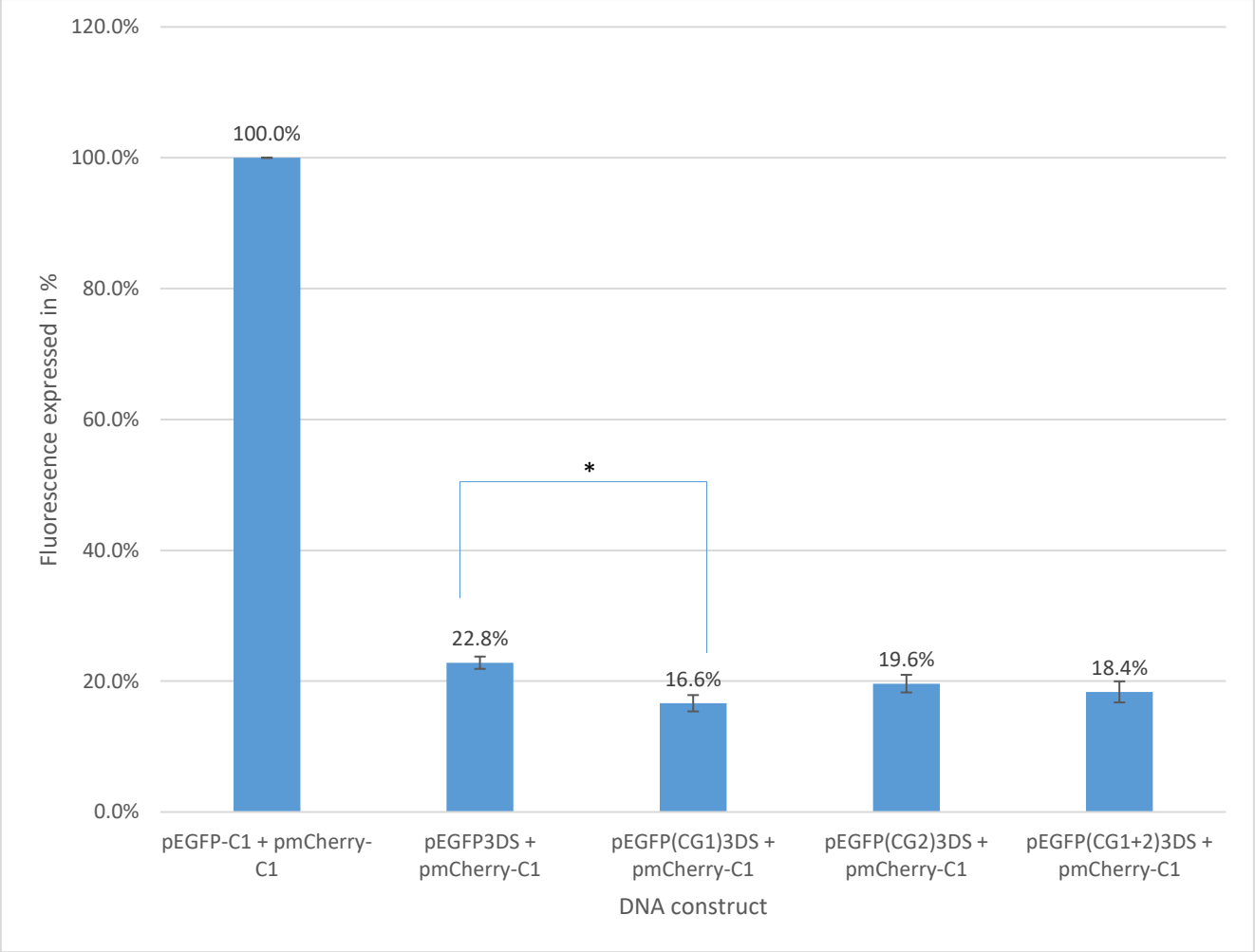


Figure 3.19: Mean fluorescence of the structure in the 3D region (pEGFP3DS) in comparison to the CG dinucleotide mutations; pEGFP(CG1)3DS, pEGFP(CG2)3DS and pEGFP(CG1+2)3DS gated on mCherry fluorescence. The original DNA construct without any mutation is highlighted in blue. GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) 48 hr after transfection. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates p = 1.7E-02.

* indicates significant difference when comparing pEGFP3DS and pEGFP(CG1)3DS.

3.4.3 Mutant plasmid construction

The 3DS structure was mutated by introducing 21 silent mutations into the pEGFP3DS clone. The designed sequence was synthesized commercially by Epoch Life Science, INC. and plasmid received was cut with XhoI and BamHI prior to ligation into pEGFP-C1 cut with the same enzymes to form pEGFP(21mut)3DS. Clones containing a shorter fragment including only the viral RNA structure in the $3D^{pol}$ region were generated by PCR amplification of the native and 21 mutation sequences.

3.4.3.1 The highly conserved structure in the $3D^{pol}$ appears to affect viral RNA stability

Experiments were designed to include the same number of base pairs of the wild type and mutants. 21 mutations were made on the secondary structure in the 3D encoding region of HPeV1 made up of three stem-loops and a tertiary structure kissing interaction and ligated into cut pEGFP-C1 to produce pEGFP(21mut)S (119 bp) and pEGFP(21mut)3DS (518 bp). Mfold program was used to determine the pattern of the mutations carried out on the structure (Figure 3.20). Table 3.5 shows the position in 3DS structure of HPeV1 in which the 21 silent mutations were introduced. The proposed theory for the RNA stability is that for more stable RNAs, each molecule should be translated more often and so stronger fluorescence should be seen when the DNA is transfected into the cells. The fluorescence of pEGFPS, pEGFP3DS, pEGFP(21mut)S and pEGFP(21mut)3DS) were then compared (Figure 3.21). All the constructs gave reduced fluorescence compared to the control pEGFP-C1. However, between the pEGFP3DS and the mutant construct, there was a significant difference. The 21 mutations caused a reduction in fluorescence. The

same result was seen in the minimal structure construct. The results suggest that the 3DS structure does have an effect on RNA stability.

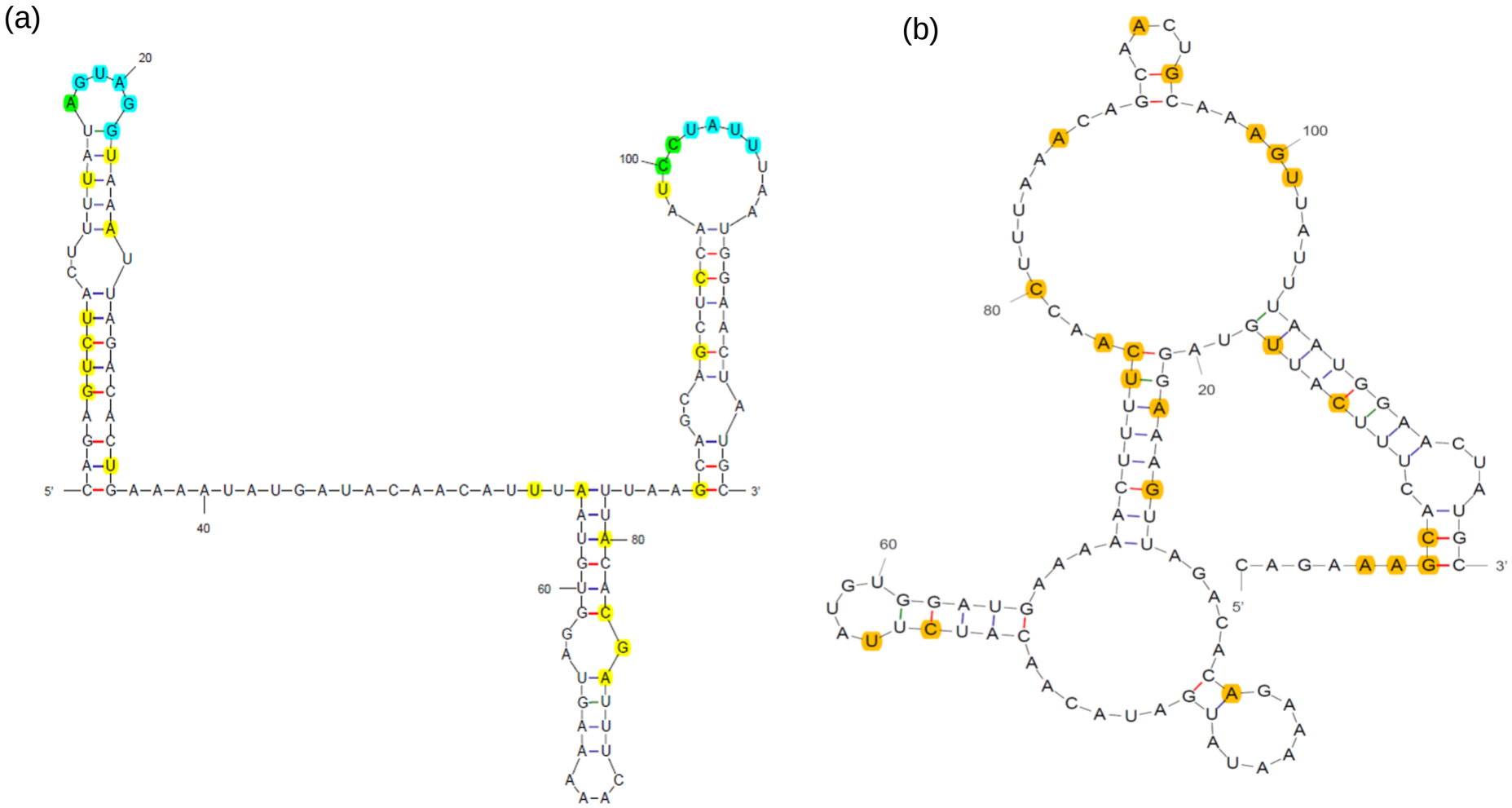


Figure 3.20: (a) Schematic representation of the predicted RNA secondary structure in the 3D encoding region of HPeV1. The tertiary structure kissing interaction between two of the loops is highlighted in blue while the points of mutations within the loops are highlighted in green. The yellow highlights shows the target points of mutations in the original structure. dG = -20.40 [initially -20.40]. (b) 21 mutations introduced into the structure thereby disrupting the three stem-loops highlighted in orange. dG = -10.19 [initially -18.60]The schematic figures were generated using the default parameters of the RNA folding form for Mfold.

Table 3.5: Summary of the 21 silent mutations introduced into the 3DS structure

S/n	Position	Original bp	Mutated bp
1	7012	G	A
2	7013	U	A
3	7014	C	G
4	7015	U	C
5	7021	U	C
6	7024	A	U
7	7030	U	A
8	7033	A	G
9	7042	U	A
10	7061	U	C
11	7063	A	U
12	7082	A	U
13	7083	G	C
14	7084	C	A
15	7087	A	C
16	7093	G	A
17	7099	G	A
18	7102	C	G
19	7106	U	A
20	7107	C	G
21	7108	C	U

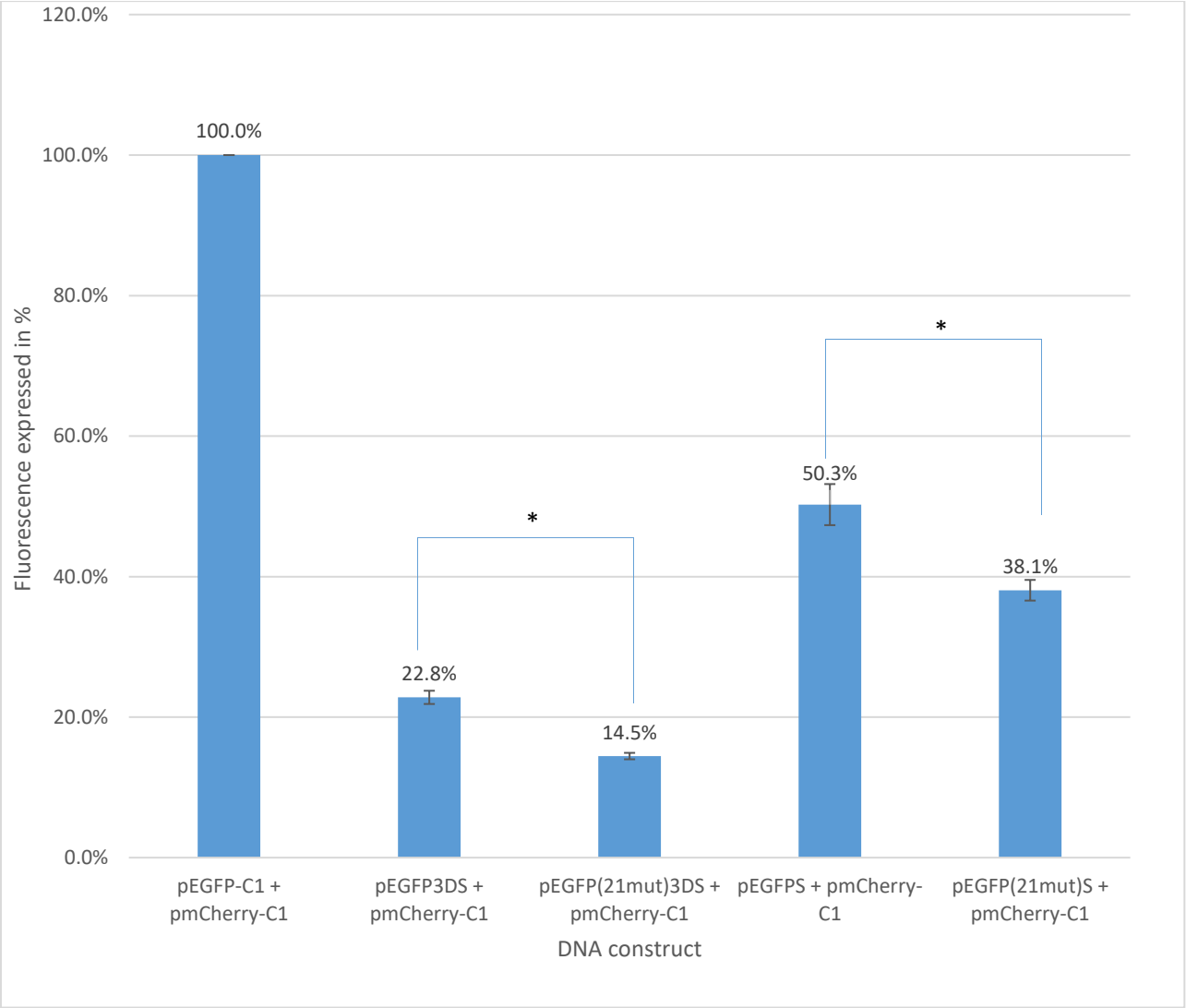


Figure 3.21: Mean fluorescence of the structure in the 3D region (pEGFP3DS & pEGFPS) in comparison to the structure containing 21 mutations (pEGFP(21mut)3DS and pEGFP(21mut)S respectively) gated on mCherry fluorescence. The original DNA constructs without mutations are highlighted in blue. GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) 48 hr after transfection. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates $p < 2.0E-02$.

* indicates significant difference when comparing the structure and the mutant.

3.5 Discussion

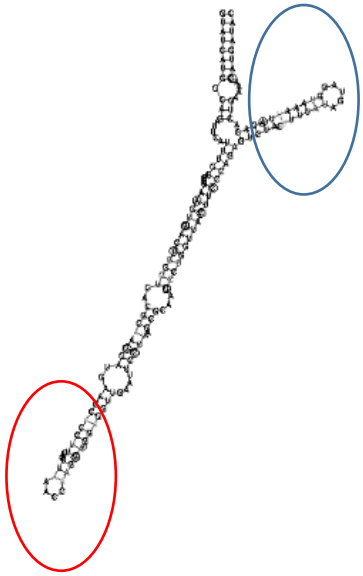
The genomes of positive strand RNA viruses are capable of forming secondary structures that are important in virus replication. These structures can also participate in intra-molecular (RNA/RNA) and inter-molecular (RNA/host cell or virus proteins) interactions to bring about replication. One example is the cis-acting replication element (cre), found at any part of the genome depending on the kind of virus e.g., 2C coding region in enteroviruses, VP1, VP2 or 2A in different human rhinovirus species and VP0 in HPeV (McKnight & Lemon, 1998; Goodfellow *et al.*, 2000; Witwer *et al.*, 2001; Al-sunaidi *et al.*, 2007; Liu *et al.*, 2009a). There are a number of other predicted structures in the 5' UTR and 3' UTR of positive strand RNA viral genomes that have role to play in transcription, replication and encapsidation (Sasaki & Taniguchi, 2003b; Liu *et al.*, 2009a). The location of the cre, which is in the coding region in most picornaviruses, is evidence that important RNA structures can exist in the coding region and other structures are starting to be found. A conserved structure in the $3C^{Pro}$-encoding region of PV and other members of *Enterovirus C*, excluding CAV1, CAV19 and CAV22, consists of four stem-loops and is proposed to play an important role of preventing the RNA genome from being cleaved by a cellular enzyme, RNase L, which was reported to cleave RNA in other groups of *Enterovirus* without the structure (Han *et al.*, 2007; Townsend *et al.*, 2008). A structure in the $3D^{pol}$-encoding region of *Enterovirus C* members has been reported to be involved in replication. It does not seem to be present in other enterovirus species (Song *et al.*, 2012; Burrill *et al.*, 2013).

Williams *et al*., (2009) compared sequences of several HPeVs in the $3D^{pol}$-encoding region and they were all highly conserved in a region of 119 nucleotides. A structure could be predicted, which was conserved in all the sequences. In this thesis, all the available

HPeV sequences, 254, were analysed (Figure 3.2). It is made up of three stem-loops with a kissing interaction between the loops of the first and third stem-loops (Figure 3.1). The alignment of all the available sequences shows that the region containing the structure is highly conserved (Figure 3.2). The few changes are mainly either in regions predicted to be unpaired or where the base pair is maintained. There are a few examples of covariant changes, and several more where the base pair is maintained through G-C/U or U-G/A interactions. The non-covariant changes, leading to the loss of a base pair are mainly at positions 35 in the alignment (base of the first stem-loop). This part of the stem-loop may therefore not exist. The position 8/29 base pair, where there are several examples of covariance, is disrupted in a few cases, but the position of the base pair means that the overall stem-loop structure is not weakened greatly. The same is true of a few changes in the second stem-loop. The conservation of the sequence in this region, together with the pattern of nucleotide differences is strong evidence that there are pressures to maintain the structure because it is functionally significant.

Using thermodynamic predictions and conservation of predicted base pairs in aligned sequences, Witwer et al., (2001), identified a number of conserved structures in several picornavirus genera. Although several of the cre predictions have been shown to be inaccurate, it is interesting that a large structure was predicted in the 3D-encoding region of HPeVs (Figure 3.22).

A



B

```
Query        1        ACGC-AAA-GATGAACCTCCTTCACTTAAACCAAGAATGGAGGTTGAATT-C-TT-AAA  54
L02971       6934     ....-...-.....................................-.-..-...  6987
EU374210     493      ....-...-.....................................G..G..-.-..-...  546
AM933161     493      ....-..G-.....................................G..-.-..-...  546
EU022171     6667     ....-...-.....................................G..-.-..G...  6721
EU024629     6667     ....-...-.....................................G..-.-..G...  6721
GU946950     400      ....-...-.....................................G..T.-.G-...  454
AM235749     6915     ....-..G-.....C...............................-.-C.-...  6968
EU077513     131       ...-...-.....C...........G..................-T-..-...  183
JN106520     369      ....-...-.....C.....C..G.....................-.-C.-...  422
AM933164     493      ....-..G-.....A....C.........................-.-C.-...  546
KJ152442     6817      ..T-..G-.....C.....C.........................-.-..-...  6869
JN106521     369      ....-...-.....C.....C..G.........C...........-.-C.-...  422
HQ696574     6484      ..G-..G-..C.....A.....................G..-.-..-...  6536
HQ696571     6484      ..G-...-..C.....A..G..................G..-.-..-...  6536
HQ696570     6484      ..G-..G-..C.....A.....................G..-.-..-...  6536
GU946961     401       ..G-..G-..C.....A.....................G..-.-..-...  453
GU946959     401       ..G-...-..C.....A..G..................G..-.-..-...  453
GU946958     400       ..G-..G-..C.....A.....................G..-.-..-...  452
AM933163     493      ....-..G-.....A..AA....A.....................-.C..-...  547
AM933160     493      ...G-..G-.....C.....C..G.....................-.-C.-...  546
KT626009     6652      ...-...-.....A...T.A.....C...........G.....-.-..       6701
JN106501     370       ..G-...-.....A..A.....A.....T..................T.-..-...  423
JN106494     369      ...A-..G-.....A..A.....A.....T.................-.C..-...  423
JN106478     368       ...G...G.....A..A.....A.....T.................-.C..-...  423
GU946955     399       ...G...G.....A..A.....A.....T.................-.C..-...  454
GU946949     399       ...G...G.....A..A.....G.....T.................-.C..-...  454
GU946948     399       ...G...G.....A..A.....A.....T.................-.C..-...  454
GU946947     399       ...G...G.....A..A.....A.....T.................-.C..-...  454
```

Figure 3.22: Analysis of the structure predicted to be present in the HPeV 3D-encoding region (Witwer *et al*., 2001) which overlaps with the Williams et al. (2009) structure. A: The predicted Witwer *et al*. structure. The stem-loop also predicted by Williams et al. is circled in blue. The core of the structure, analysed below, is circled in red. B: Alignment of the HPeV1 Harris sequence and its closest relatives found by a Blast search in the region of the core structure. The nucleotides predicted to base pair are highlighted in the core structure. Differences which disrupt the structure are highlighted in red and ones which are not predicted to base pair are not marked. There are no differences which lead to conservation of the structure.

The regions containing the Witwer *et al*. and Williams *et al*. 3DS structures overlap and the first stem-loop of the 3DS structure is also predicted by Witwer *et al*. However, aligning the sequences making up the top of the very large stem-loop structure gives no support for the Witwer *et al*. structure as many of the differences in this area disrupt the predicted base pairs. The Witwer *et al*. structure was based on only 3 HPeV sequences and so there was little data to give an accurate prediction, but it is interesting that one of the 3DS domains was predicted which suggests that it is energetically favourable.

Aligning HPeV sequences with the sequences of other members of the *Parechovirus* genus (Ljugan virus, Sebokele virus and Ferret Parechovirus 1), belonging to species B, C and D respectively) showed no evidence of a corresponding structure, except in Sebokele virus. Here, 3 stem-loops plus a kissing interaction were also predicted (Figure 3.3). This provides further evidence that the structure is important for the viruses. The lack of a structure in the other parechoviruses argues against this, but this is not unusual. Neither of the structures present in polioviruses and other members of *Enterovirus C* are seen in other enterovirus species (Song *et al.*, 2012; Burrill *et al.*, 2013).

The HPeV 3DS structure is located in the RNA sequence encoding the C-terminal part of 3D. The *Enterovirus C* structures (Song *et al.*, 2012; Burrill *et al.*, 2013) are also located in the 3D-encoding region and to find if there is any relationship between 3DS and the *Enterovirus C* structures, an amino acid alignment was performed and the corresponding nucleotide sequences were added (Figure 3.23). It can be seen that 3DS and the *Enterovirus C* structures are in the same general area of the genome but there is only a short overlap of 3DS with one of the regions containing the *Enterovirus C* structures. This region of overlap does not include the well-conserved stem-loop which is the key component of this *Enterovirus C* structure and so there seems to be no similarity between

the HPeV 3DS and the Enterovirus C structures. It is interesting, though, that they are located in a similar region and, in addition, a comparison between FMDV and PV1 revealed FMDV RNA structures that overlap with the PV1 (Enterovirus C) structures (Martin Ryan, unpublished). A number of other *Picornaviridae* genera were shown to have potential RNA structures in this region (Hershan, 2012). The reason why RNA structures within the coding region tend to be concentrated in the 3D C-terminus-encoding area is unclear, but in some cases it may be related to RNA replication, as negative strand synthesis is initiated at the 3' terminus of the genome, which is located close to these possible structures.

```
AACTTGATTATCAGGACACTCTTACTGAAAACCTACAAGGGCATAGATTTAGACCACCTAAAA
 N  L  I  I  R  T  L  L  L  K  T  Y  K  G  I  D  L  D  H  L  K
 L  M  M  C  I  Y  T  T  N  L  I  S  P  G  I  D  C  L  P  -  -
TTGATGATGTGTATTTACACTACAAATTTAATCAGTCCTGGAATTGATTGTCTACCA------

ATGATTGCCTATGGTGATGATGTAATTGCTTCCTACCCCCATGAAGTTGACGCTAGTCTCCTA
 M  I  A  Y  G  D  D  V  I  A  S  Y  P  H  E  V  D  A  S  L  L
 -  I  V  Y  G  D  D  V  I  L  S  L  D  K  E  I  E  P  E  K  L
---ATCGTATATGGCGATGATGTTATTCTTTCATTGGACAAAGAAATTGAACCAGAGAAACTG

GCCCAATCAGGAAAAGAC---TATGGACTAACTATGACTCCAGCTGACAAATCAGCTACATTT
 A  Q  S  G  K  D  -  Y  G  L  T  M  T  P  A  D  K  S  A  T  F
 Q  S  I  M  A  D  S  F  G  A  E  V  T  G  S  R  K  D  E  P  P
CAAAGTATCATGGCAGATTCATTTGGAGCCGAAGTGACTGGGTCACGCAAAGATGAACCTCCT

GAAACAGTCACATGGGAGAATGTAACATTCTTGAAGAGATTCTTCAGGGCAGACGAGAAATAC
 E  T  V  T  W  E  N  V  T  F  L  K  R  F  F  R  A  D  E  K  Y
 S  L  K  P  R  M  E  V  E  F  L  K  R  K  P  G  Y  F  P  E  S
TCACTTAAACCAAGAATGGAGGTTGAATTCTTAAAGCGGAAACCTGGTTATTTCCCAGAGTCT

CCATTTCTTATTCATCCAGTAATGCCAATGAAGGAAATTCATGAATCAATTAGATGGACTAAA
 P  F  L  I  H  P  V  M  P  M  K  E  I  H  E  S  I  R  W  T  K
 T  F  I  V  G  -  K  L  D  T  E  N  M  I  Q  H  L  M  W  M  K
ACTTTTATAGTAGGT---AAATTAGACACTGAAAATATGATACAACATTTAATGTGGATGAAA

GATCCTAGGAACACTCAGGATCACGTTCGCTCTCTGTGCCTTTTAGCTTGGCACAATGGCGAA
 D  P  R  N  T  Q  D  H  V  R  S  L  C  L  L  A  W  H  N  G  E
 N  F  S  T  F  K  Q  Q  L  Q  S  Y  L  M  E  L  C  L  H  G  K
AACTTTAGCACATTTAAGCAGCAGCTCCAATCCTATTTAATGGAACTATGCCTCCATGGAAAA

GAAGAATATAACAAATTCCTAGCTAAAATCAGGAGTGTGCCAATTGGAAGAGCTTTATTGCTC
 E  E  Y  N  K  F  L  A  K  I  R  S  V  P  I  G  R  A  L  L  L
 D  T  Y  Q  H  Y  I  K  I  L  E  P  Y  L  Q  E  W  N  I  T  V
GACACTTATCAACACTACATTAAGATCTTGGAACCATATCTACAGGAATGGAATATCACTGTG

CCAGAGTACTCAACATTGTACCGCCGTTGGCTTGACTCATTT------
 P  E  Y  S  T  L  Y  R  R  W  L  D  S  F  -  -
 D  D  Y  D  V  V  I  T  K  L  M  P  M  V  F  D
GATGATTATGATGTGGTTATAACTAAGTTGATGCCCATGGTGTTTGAT
```

Figure 3.23: Alignment of the amino acid sequences of the 3D protein C-terminus from PV1 Mahoney (red font) and HPeV1 Harris (black font). The corresponding nucleotide sequences are also shown and the location of RNA structures are indicated by highlights. The HPeV 3DS structure is highlighted in grey. The two functionally redundant structures found in PV and other members of *Enterovirus C* are highlighted in green (Song *et al.*, 2012). The first of these is contained within a larger structure predicted by (Burrill *et al.*, 2013) and the additional nucleotides making that structure are highlighted in yellow. The figure was based on a comparison between PV1 and FMDV (Martin Ryan, unpublished) and was produced by aligning the amino acid sequences using ClustalW and manually adding the corresponding nucleotide sequences.

The 3’ UTR of Aichivirus contains an RNA stem-loop structure which is also found in a few other picornaviruses (Boros *et al.*, 2012). We extended this observation by taking the Aichivirus structure sequence and comparing with the database. Repeating the process with each hit, led to the identification of the structure in several picornaviruses (Figure 3.4). Interestingly, we also found the structure in two members of a different virus family, *Caliciviridae*. The structure therefore has some similarity to the s2m motif found in *Astroviridae*, *Coronaviridae* and in some picornaviruses. The Aichivirus structure is a long stem-loop with two large bulges which have highly conserved sequences among the different viruses. The function of the s2m and Aichivirus structure is not known.

To study the 3DS structure (Williams *et al.*, 2009) and structures such as the Aichivirus structure, we used an EGFP-based assay. Similar EGFP assays have been used to study post-transcriptional processes such as RNA stability and RNA editing (Benjamin *et al.*, 2006; Laloo *et al.*, 2009; Severi & Conticello, 2015). Differences in RNA levels, due to increased or decreased stability should affect the observed EGFP fluorescence.

The RNase L binding structure in poliovirus prevents degradation of RNA (Han et al., 2007). In addition, it was found a number of years ago that HPeV1 (then called echovirus 22) RNA is stable *in vitro* and resistant to pancreatic RNase (Seal & Jamison, 1984). Affecting HPeV RNA stability may therefore be part of the role of the 3DS structure. A number of DNA constructs were designed where sequences were added downstream of a stop codon at the end of the EGFP encoding region in pEGFP-C1 or mCherry encoding region in pmCherry-C1. In all cases apart from pEGFP-C1 and pmCherry-C1 themselves, the constructs included a stop codon at the same position to give the same differences in protein distribution or stability. A random short sequence was added to cut pEGFP-C1 (pEGFPeC) and cut pmCherry-C1 (pMCheC) to control for this difference from the

proteins expressed from the original vectors. PCR reactions were carried out using *Pfu* polymerase to amplify the region to be studied, because it has proofreading ability while *Taq* polymerase does not have this ability (Lawyer *et al.*, 1993a; Cline *et al.*, 1996). All constructs were analysed by sequencing to ensure no additional changes were introduced. A summary of the constructs are shown in Figure 3.23.

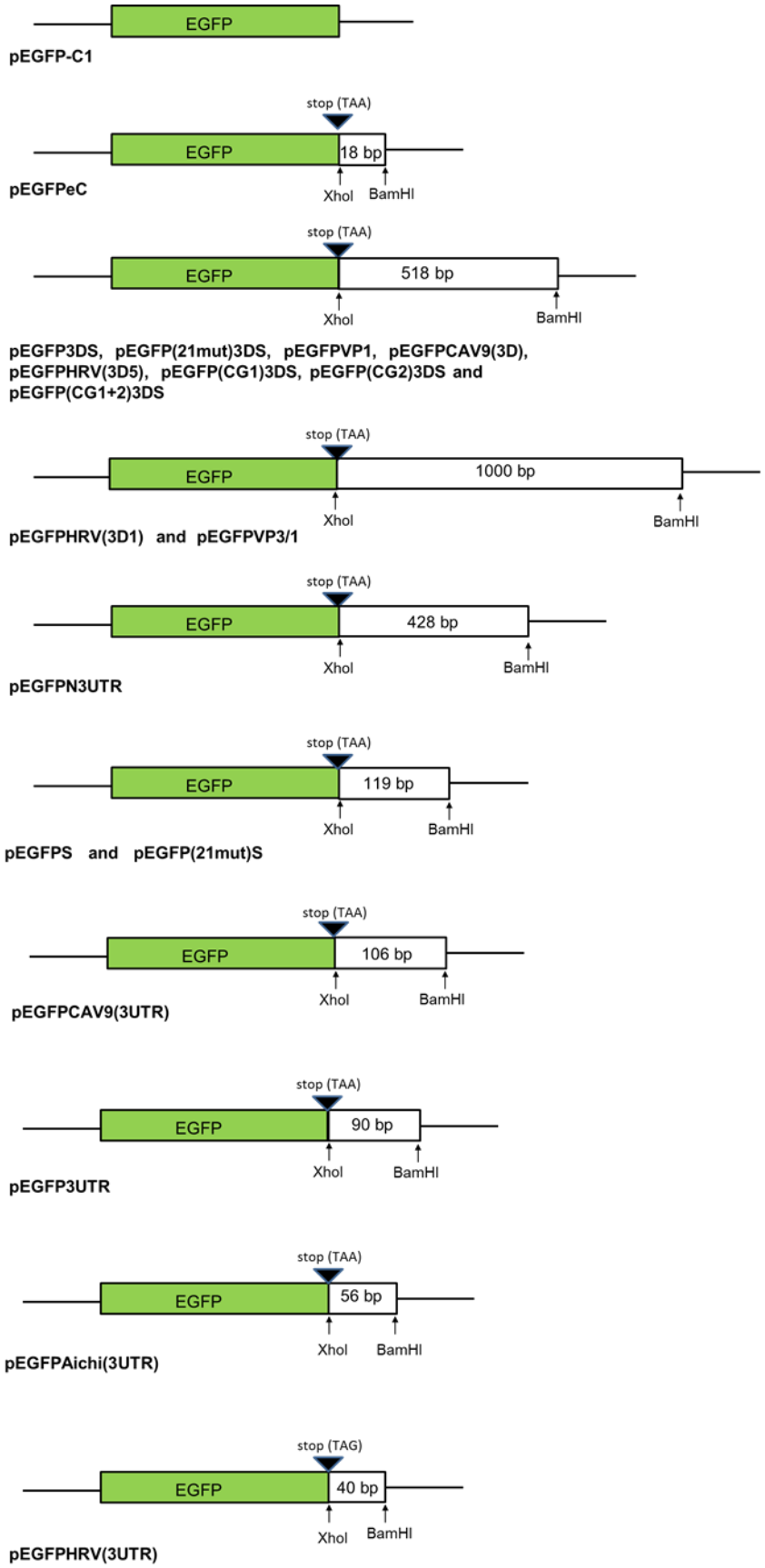


Figure 3.24: Schematic representation of pEGFP-C1 with all other constructs analyzed indicating the length of the inserted sequence. All apart from pEGFP-C1 give an EGFP protein with the same C-terminal sequence.

3.5.1 **Investigation of the role of the 3DS structure via microscopy**

Several different cell lines were used to study the effect of transfection of the DNAs in the cells. The MCF-7 cell line was transfected better than some of the others, but optimal results were obtained for the GMK cell line. These yielded the highest transfection efficiency compared to the other cell lines used (data not shown). At the initial stages of the experiment where microscopy was used, pEGFP3DS and pEGFPeC were studied. These were distributed all over the cell with some concentration of EGFP at the nucleus of cells transfected with pEGFPeC although transfection with pEGFP-C1 produced brighter fluorescence all over the cells. This may be due to differences in the C-terminus of the EGFP.

For a clearer understanding of the results, pmCherry-C1 was co-transfected with pEGFP3DS and pEGFPeC to GMK cells. The EGFP fluorescence from pEGFPeC and pEGFP-C1 was always relatively higher, in comparison with the mCherry fluorescence from pmCherry-C1, than pEGFP3DS. This suggests that RNA transcribed from pEGFPeC and pEGFP-C1 is more stable than pEGFP3DS (Figures 3.6 and 3.7).

The EGFP constructs, pEGFP3DS and pEGFPeC, were reproduced in pmCherry-C1 to produce pMCh3DS and pMCheC respectively. This was done to see if the same results would be obtained. Co-transfection of ‘pEGFP3DS with pMCheC’ and ‘pEGFPeC with PMCh3DS’ to GMK cells did not give useful results because the mCherry signal was weak and microscopy revealed abnormal localisation. Another problem was that cells seemed to be bleached easily when the exposure time in the red channel was increased to aid visibility, hence it was difficult to see details in the cells. Co-transfection results of pEGFP-C1 with pMCh3DS and pMCheC showed similar disadvantages, whereas transfection with pEGFP-C1 always gave clear results.

The experiment was aimed at determining the stability effect by examining how much fluorescence would be produced, i.e. if the RNA were more stable, each molecule should be translated more often and so stronger EGFP fluorescence should be seen. The results, however, showed that there was no clear cut difference between pEGFP3DS, pMCh3DS and their control experiments.

The stability of HPeV1 Harris RNA due to the highly conserved region of the $3D^{Pol}$ could not be studied using the above explained experimental design as both the control and main experiment co-incidentally show highly similar results.

3.5.2 **Investigation of the 3DS structure using IRES driven translation**

An experiment was conducted to study the possible effect of the 3D structure on translation by producing a bicistronic construct with dsRed as the reporter of IRES-driven translation. Three bicistronic constructs, T5dsRed, T5dsRed3DS and T5dsRed3D, were made up of GFP, already present in the Topo vector used as the first cistron and the 2^{nd} cistron is controlled by the IRES in the 5' UTR of HPeV1. If the 3DS structure enhances IRES driven translation, then the dsRed signal should be increased. The results from this experiment showed that T5dsRed, T5dsRed3DS and T5dsRed3D were transfected very inefficiently into GMK cells. Results revealed properties typical to EGFP (distribution all over the cell) and dsRed (distributed only in the cytoplasm, as the protein is tetrameric and is too large to go into the nucleus) nevertheless; no major observed difference was seen in their fluorescence. No conclusion could be drawn from studies with the bicistron experimental designs and so new experiments were designed to study the structure in more details.

3.5.3 Further investigation of the role of the 3DS structure via microscopy and FACS

A number of other constructs were studied by co-transfecting with pmCherry-C1 into GMK cells (Figure 3.8). This was carried out to provide a comparison and control for differences in the DNA transfection and therefore the level of EGFP fluorescence present in the cells. Significant differences in fluorescence could not be observed microscopically.

Due to the above explained limitation, cells transfected with the same DNA constructs listed above were analysed using a FACS to quantitatively examine the difference in fluorescence of the constructs. Typical results are shown in Figure 3.15. Gating was done at 10^4 to analyse for efficiently transfected and stable cells because from 10^4 and above, a high level of fluorescence is expressed. Double positive cells (expressing both EGFP and mCherry fluorescence) were gated on the P5 population and then only the EGFP fluorescence expressed by the cells was recorded in the P6 population in percentage (Figure 3.15). It was observed that transfection with the control, pmCherry-C1 and pEGFP-C1, had 99.8% of EGFP fluorescence. The constructs with shorter inserts (pEGFP3UTR [90 bp], pEGFPeC [18 bp] and pEGFPS [119 bp]) gave similar values to the control. pEGFP3DS and pEGFPVP1, which was used to control for the size of the insert in pEGFP3DS (518 bp) gave lower levels of EGFP fluorescence. It is also noticeable from the output in Figure 3.15 that a number of cells have lower EGFP fluorescence. This was more noticeable for pEGFP3DS (51.3% EGFP fluorescence) and pEGFPN3 (63.6% EGFP fluorescence) than pEGFPVP1 (81.6% EGFP fluorescence).
This finding led to an inference that there may be a region in the section of the $3D^{pol}$-ecoding sequence present in pEGFP3DS which brings about the reduction of fluorescence in the cells. One possible candidate is that there are 2 strongly conserved

CG dinucleotides at positions 6935/6 and 6997/8 (Williams *et al.*, 2009). CG dinucleotides are highly under-represented in HPeV genomes and this conservation is therefore unusual. These positions were re-examined using the much larger HPeV dataset now available (299 sequences in this region). It was found that 79% of the sequences had CG at position 6935/6 and 95% had CG at position 6978/9 suggesting that these sequences play some role in HPeV replication. These dinucleotides are present in the strongly affected constructs pEGFP3D and pEGFPN3, but not in any of the other constructs. To investigate if they are responsible for the reduced EGFP signal, they were mutated. pEGFP(CG1)3DS, pEGFP(CG2)3DS and pEGFP(CG1+2)3DS are constructs derived from silently mutating the 1st, 2nd and both the 1st and 2nd 'CG' dinucleotides respectively to 'AG'. The results (Figure 3.19) show that none of the mutations was greatly different from the original pEGFP3DS construct suggesting that these CG dinucleotides do not affect stability/translation. However, after mutating the 1st CG dinucleotide, there was a reduction in fluorescence of about 27.17% compared to the wild type, pEGFP3DS, therefore suggesting that the 1st CG dinucleotide is useful to the virus possibly more than the 2nd CG dinucleotide which was only 14.02% lesser in fluorescence than the original DNA, pEGFP3DS.

3.5.4 **The relationship between insert size and EGFP fluorescence produced**

As the constructs with the smallest inserts seemed to give the largest fluorescence, the reduced fluorescence given by pEGFP3DS may be related to size of insert. To analyse this further, several new constructs were made, giving a total of 19 (Figure 3.23). These included inserts taken from CAV9 and HRV1B as these have slightly different nucleotide compositions from each other, with HRV1B being similar to HPeVs (Stanway, 1990; Al-

sunaidi *et al.*, 2007). This was intended to control for the effect of nucleotide composition on the results obtained.

Cells co-transfected with pmCherry-C1 and pEGFP-C1 was used as the control and set as the basis to calculating the expression of EGFP in all other constructs. The mean fluorescence of each construct was generated by the FACS after gating for only mCherry positive cells and then the result obtained was normalised to pEGFP-C1 + pmCherry-C1 control expressed in percentage. Normalization was carried out by dividing the fluorescence result obtained from each construct by the result obtained from pEGFP-C1 + pmCherry-C1. Transfection was carried out in 48 hr and after three repeats of the experiment, it was noticed that there was consistent loss of EGFP fluorescence in most of the constructs compared to pEGFP-C1 (always expressed as 100%). Also, it was mostly noticed that the longer length of DNA insert give lesser fluorescence in the cells when compared to shorter ones. In fact, the DNA constructs, pEGFPHRV(3D1) and pEGFPVP3/1, which were 1000 bp in size had a loss of fluorescence of 91.59 % and 89.22% respectively when compared with the fluorescence expressed by the control, pEGFP-C1. All the constructs with an insert of 518 nucleotides (from HPeV 3D or VP1, CAV9 and HRV1B) gave similar fluorescence suggesting that nucleotide composition is not a factor.

3.5.5 **Mutation of the 3DS structure**

We decided to mutate the 3DS structure in the construct pEGFP3DS and also in the minimal 3DS constructs, pEGFPS. 21 silent mutations were introduced to disrupt the structure as much as possible (Figure 3.20). No CG dinucleotides were produced by the mutations as it is now known that these can affect virus replication, although mutation of the pair of conserved CG dinucleotides did not suggestively affect the results shown in

Figure 3. 19 (Atkinson *et al.*, 2014; Tulloch *et al.*, 2014a). It has been shown by introducing multiple silent mutations into the genome of an echovirus that increasing the number of UA dinucleotides can also reduce picornaviruses replication (Atkinson *et al.*, 2014; Tulloch *et al.*, 2014a), but in our work the mutated 3DS region had 8 UA dinucleotides, while the wild type region had 12. In addition, in the echovirus work significant effects on virus replication were only observed when large regions of the genome were mutated and many CG and UA codons were introduced. It is therefore, unlikely that the few silent mutations introduced into 3DS would have an effect on genome stability due to changes up in CG/UA dinucleotides, rather than due to the disruption of the RNA structure.

There was an observed reduction in fluorescence from 50.3% to 38.1% when the structure, pEGFPS, was completely broken by introducing 21 silent mutations to it (pEGFP(21mut)S). The pEGFP(21mut)3DS also showed a decrease in fluorescence compared to the original DNA (pEGFP3DS). This suggests that the 3DS structure may be playing some role in RNA stability/translation.

3.5.6 **Other RNA structures**

The 3’ UTR of HPeV1, CAV9, HRV and part of the 3’ UTR of Aichivirus (the promiscuous element) was also studied. The regions studied were made up of different DNA insert sizes, 90, 106, 40 and 56 bp respectively. The most interesting result was obtained from pEGFPHRV(3UTR) with 32.21 % more fluorescence than the control, pEGFP-C1. The other inserts had no significant effect on fluorescence.

Surprisingly, there was a 44.26 % loss in fluorescence in pEGFPN3 when the 3’ UTR of HPeV1 was removed from pEGFP3DS (518 bp) to give pEGFPN3 (428 bp). This suggests that the HPeV 3’ UTR could have a region which makes RNA from pEGFP3DS

more stable than pEGFPN3, although the construct pEGFP3UTR, which contains only the HPeV 3' UTR did not enhance the fluorescence compared to the pEGFP-C1.

In summary, the size of the DNA relatively affects the fluorescence but some DNA having the same size as others are more stable than the others.

Rhinoviruses have a short 3' UTR made up of a single stem-loop (Palmenberg *et al.*, 2009). The HRV1B 3' UTR is 40 nucleotides long and is A/U rich, giving a weak secondary structure (Hughes *et al.*, 1988) and this is typical of all rhinoviruses. The 3' UTR of HRV14 can be largely deleted to give a viable virus with poor growth properties (Todd & Semler, 1996). This suggests that it is not essential to the virus but plays some role. Our results show that one possible role is in stabilising the RNA or affecting translation in some way. It would be interesting to study this further by making mutants and by analysing the effect of other rhinovirus 3' UTRs in our system.

Several viruses are known to have structures which affect RNA stability. These include a diverse set of DNA and RNA viruses which contain an element called the ENE (Tycowski *et al.*, 2012) and Rous sarcoma virus also has a stabilising element (Withers & Beemon, 2011).

Measurement of the half-lives of 21,248 yeast mRNA sequences identified many sequences that stabilise or destabilise the mRNAs (Geisberg *et al.*, 2014). Genome-wide analysis of mammalian mRNAs also identified a number of elements which stabilise the RNA (Goodarzi *et al.*, 2012). Given the possibility of being recognised as foreign by the infected cell, it is likely that RNA viruses such as HPeVs or rhinoviruses could have elements which stabilise the viral genome.

Although the effect of mutation on the 3DS structure was modest, marginal effects on factors such as RNA stability could play a large role in HPeV evolution. It would be interesting to carry out both infection and transfection of the DNA constructs on cells to observe if the structure works better in infected cells or in other stress conditions to simulate virus infection. The possible role of the three stem-loop structure located in the $3D^{pol}$-encoding region of HPeV1 in withstanding stress within cells is studied in Chapter 4.

Chapter 4

Investigation of the effect of RNA secondary structures on RNA stability in response to cell stress

4.1 Introduction

Secondary structures are important elements in RNA viruses as they are involved in various activities such as the initiation of translation and transcription, prevention of cleavage to viral mRNA and also for replication of the viral genome (Han *et al.*, 2007; Liu *et al.*, 2009b; Song *et al.*, 2012). However, the predicted structure located in the 3D encoding region of HPeV1 has not been previously studied, thus, there is little understanding on how this RNA structure works. The work described in chapter 3 identified some effect of the structure in the EGFP fluorescence assay. The effect was small, but it would be interesting to find if it was more obvious in virus infected cells, or in cells under different stress conditions which may occur in virus-infected cells.

ER stress activates the unfolded protein response (UPR) which helps to restore the protein homeostasis. This ER stress response could cause or aggravate diseases such as neurodegenerative diseases, inflammation, cancer and metabolic diseases. Failure of the ER function due to ER stress could lead to apoptosis and ultimately cell death (Sano & Reed, 2013). Viral replication could also induce ER stress within cells and also cell death (He *et al.*, 2006). ER stress could also promote oxidative stress due to the accumulation of endogenous peroxides (Yokota *et al.*, 2003).

Oxidative stress is another type of cell stress and it is caused by the accumulation of reactive oxygen species (ROS) which are by products of aerobic metabolism within cells (Schieber & Chandel, 2014). This stress could come from within the cells or from the external environment and it may bring about alteration of RNA transcription (causing the transcription of more genes or even the downregulation of other genes) (Mager *et al.*,

2000). ROS production may induce damage to lipids, nucleic acids and proteins leading to cellular degradation (Yokota *et al.*, 2003).

Several viruses seem to induce oxidative stress (Schwarz, 1996; Beck *et al.*, 2000). Osmotic stress is another type of stress that cells encounter. This could be hyperosmotic stress; which occurs due to the efflux of water from the cells causing the cells to shrink and become flaccid, or hypo-osmotic stress. This occurs when there is an influx of water into the cells, causing swelling and ultimately turgidity or bursting of the cells (Mager *et al.*, 2000). There is some overlap in how the cell responds to oxidative and osmotic stress (Burg *et al.*, 2007).

The infection of cells by some members of the *Picornaviridae* family causes the cells to modify their biochemical properties to defend itself from the viral infection, although, the virus must be able to bypass this defensive mechanism (Han *et al.*, 2007; Morrison & Racaniello, 2009). Some picornaviruses are known to shutoff cap-dependent translation processes by infected cells thereby allowing the virus to usurp the available resources in the cell for its replication within its host (Svitkin *et al.*, 2005). Virus infection also manipulates RNA stress granules and in some cases uses heat shock proteins (White *et al.*, 2012; Wang *et al.*, 2013).

In this chapter, the effect of the 3DS structure was under conditions of ER, oxidative, osmotic stress was analysed as well as in virus-infected cells (Krupina *et al.*, 2010).

4.2 Analysis of optimum Actinomycin-D concentration to prevent transcription

One difficulty with the experiment in Chapter 3 is that the EGFP fluorescence is high making analysis of the results difficult. In addition, in theory, RNA is being produced continuously in transfected cells and this may mask any effect seen on the stability of the RNA. Actinomycin-D inhibits the cell transcription machinery (Romanova *et al.*, 2009; Krupina *et al.*, 2010). Treating cells with Actinomycin-D after transfection should prevent further transcription and the RNA should then decay with a half-life depending on the fluorescence/absence of stabilizing/destabilizing elements. This drug has been used widely in studies on RNA stability (Saunus *et al.*, 2007). 8 concentrations of Actinomycin-D were analysed to find the maximum concentration that would not be harmful to the cells, because Actinomycin-D could induce apoptosis within cells (Romanova *et al.*, 2009).

The concentrations were introduced to confluent cells in a 24 well plate also containing two controls; an empty well (blank) and a well containing only cells which would be used for normalizing and calculating the relative cell death. The experiment was left in an incubator for 24 hr and then observed by eye and with a light microscope (OLYMPUS CK2) (Table 4.1). A crystal violet growth assay was used to study the amount of cell damage due to the Actinomycin-D after 24 hr of treatment by measuring the amount of cells left in the well. Figure 4.1 shows the relative cell death due to the different concentrations of Actinomycin-D analysed, therefore, a concentration of 0.01 µg/ml was used for the experiment as this is the highest concentration giving no major effect on the cells. A sub-set of constructs was used (Table 4.2).

Table 4.1: 8 concentrations of Actinomycin D analysed after 24 hr of treatment to GMK cells by observation by eye and light microscope

Concentration (µg/ml)	**24 hr treatment and observation by eye**	**24 hr treatment and observation using the light microscope**
1000	Many detached and floating cells	Completely rounded with lots of dead cells and spaces
100	Few detached cells and spaces between clusters of cells	Cells looked very stressed, shrunk down but with lesser spaces
10	Intact monolayer	Lesser shrinkage and fewer spaces
1	Intact monolayer	Lesser stress and much fewer spaces
0.1	Intact monolayer	Cells appeared healthy
0.01	Intact monolayer	Cells appeared healthy
0.001	Intact monolayer	Cells appeared healthy
0.0001	Intact monolayer	Cells appeared healthy

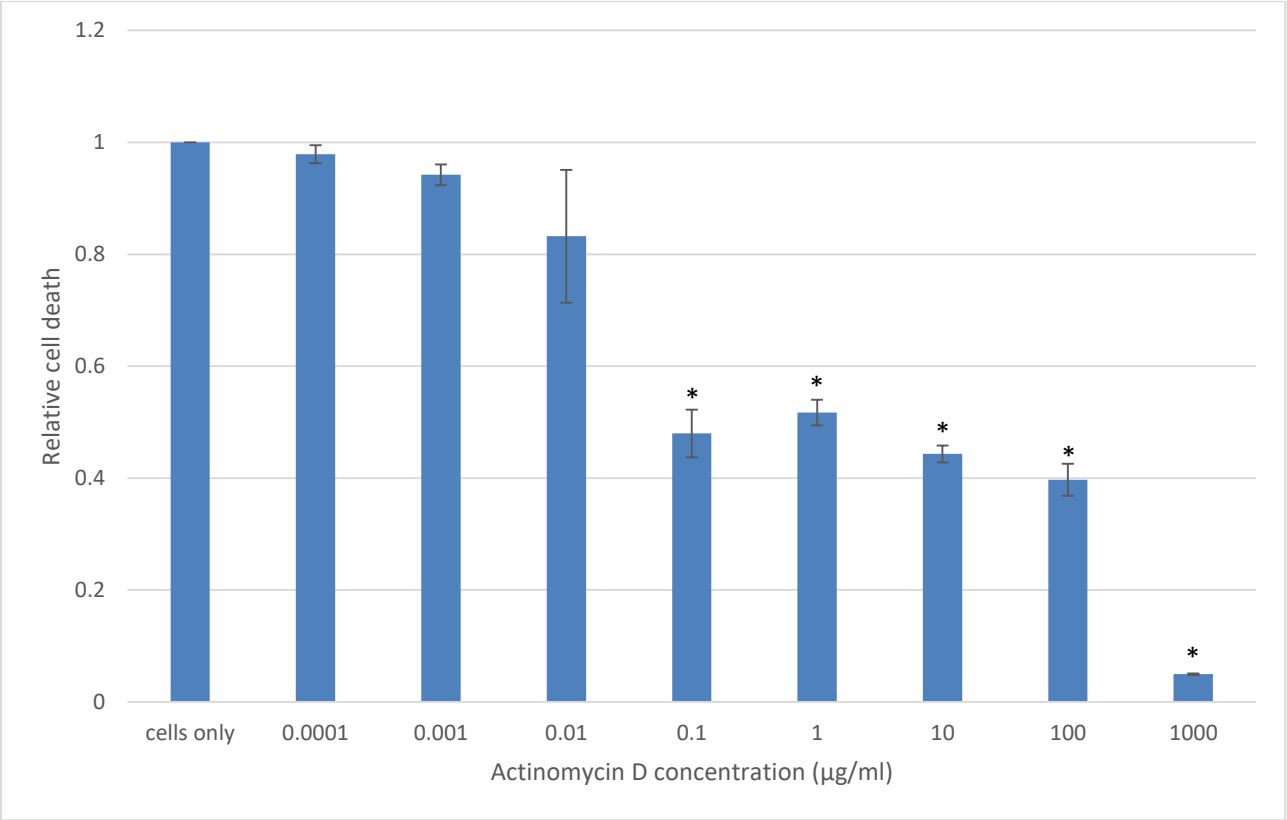


Figure 4.1: 8 concentrations of Actinomycin D were analysed using a crystal violet assay after 24 hr treatment to GMK cells, by normalizing the cell death at each concentration to cells only (normalized to 1.0). The error bars show the Standard Error (SE) from 2 independent experiments. * indicates $p <$ 3.0E-05.

* indicates significant difference of Actinomycin-D concentration to the control, cells only.

Table 4.2: Name, insert size ligated to cut pEGFP-C1 and origin of each construct used for analysis.

Construct	DNA size (bp)	Origin
pEGFPeC	18	Nil
pEGFP3DS	518	HPeV1
pEGFP(21mut)3DS	518	HPeV1
pEGFPS	119	HPeV1
pEGFP(21mut)S	119	HPeV1
pEGFPHRV(3UTR)	40	HRV
pEGFPAichi(3UTR)	56	Aichivirus

4.2.1 Regulation of the expression of transfected constructs by Actinomycin-D.

Figures 4.2 and 4.3 show the comparison between ordinary transfection (where the DNA was left to be transfected for 24 hr and then the growth media was changed to allow for EGFP/mCherry expression for another 24 hr) and the treatment with the optimized concentration of Actinomycin-D (where the DNA was left to be transfected for 24 hr before being treated with growth media containing 0.01 µg/ml of Actinomycin D for 24 hrs after transfection). Transfection was carried out using Lipofectin as described in Section 2.2.2.1 and only the GMK cell line was used.

The data was normalized in two ways. Firstly by normalizing the results from each construct with the pEGFP-C1 + pmCherry-C1 results for the same treatment. This should show the difference between the different constructs. Secondly, by normalizing the results to the pEGFP-C1 + pmCherry-C1 from the untreated control. This should show the overall effect of the treatment.

4.2.2 Normalization to individual (pEGFP-C1 + pmCherry-C1) for each experiment

Normalization was carried out by dividing the amount of fluorescence obtained in each DNA construct by the amount of fluorescence obtained in (pEGFP-C1 + pmCherry-C1) for each individual experiment (in treated and untreated cells) and then the average was calculated and was used to plot the graph. Hence, all the results for no Actinomycin-D was normalized to the control; (pEGFP-C1 + pmCherry-C1) with no Actinomycin-D while all the results for Actinomycin-D, was normalized to the control; (pEGFP-C1 + pmCherry-C1) which was treated with 0.01 µg/ml of Actinomycin-D.

Figure 4.2 shows the results. It can be seen that there is little difference between the treated and the untreated values, which implies that the effect of the treatment was equal regardless of the construct.

4.2.3 Normalization to only untreated (pEGFP-C1 + pmCherry-C1)

The direct comparison between normal transfected cells and cells treated with Actinomycin-D was studied by normalizing all results from the experiment to only the fluorescence expressed by the untreated control; (pEGFP-C1 + pmCherry-C1). The results are shown in Figure 4.3. It would be expected that there would be a reduction in fluorescence as Actinomycin-D should prevent further transcription (Krupina *et al.*, 2010). However, for some of the samples, the opposite is the case, although, the results are not statistically significant. Comparing with the constructs, the reduced fluorescence of pEGFP3DS compared to pEGFP-C1 and pEGFPeC was seen but any effect of the mutation was not as clear.

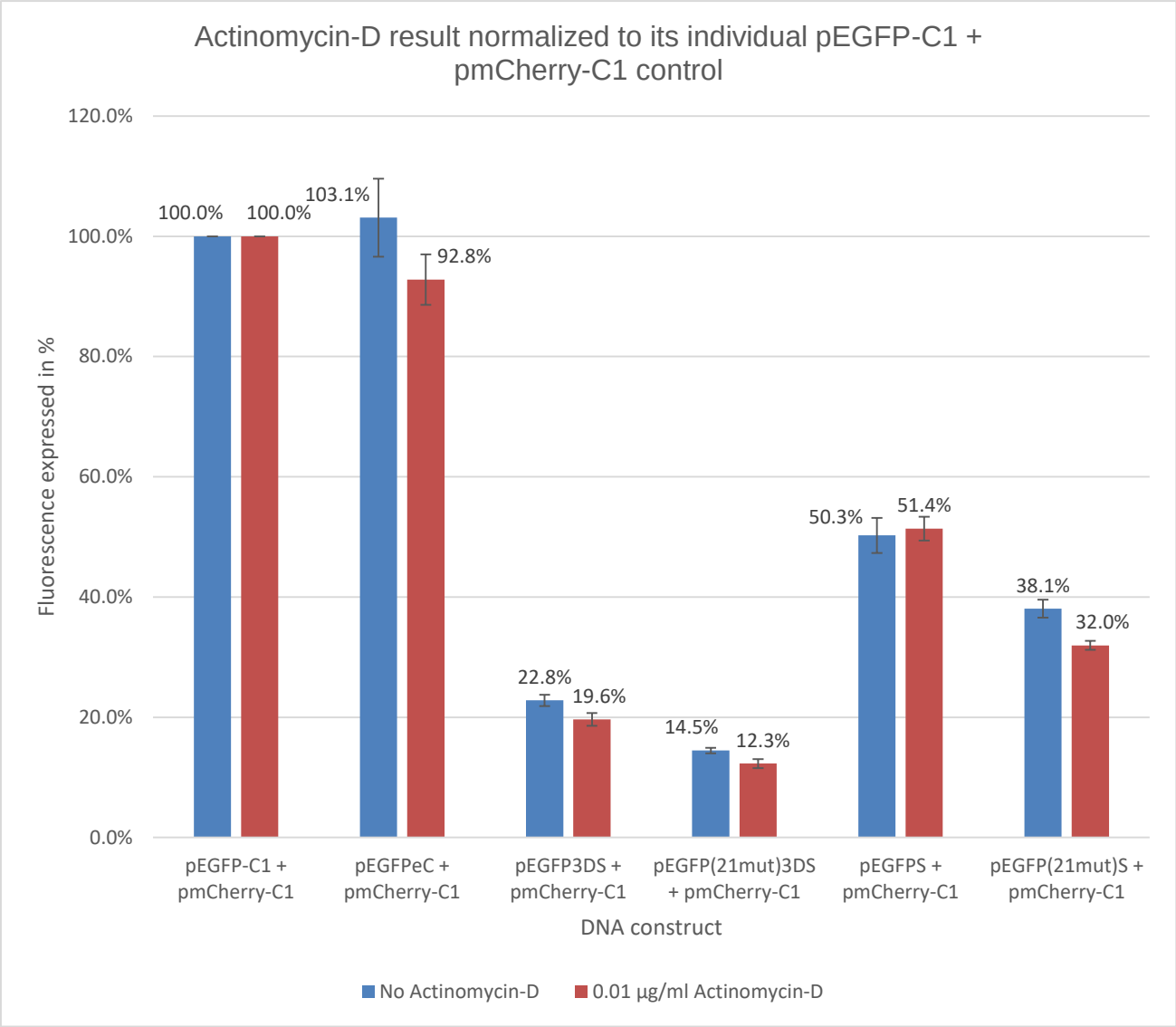


Figure 4.2: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. The result was normalized to (pEGFP-C1 + pmCherry-C1) of each independent experiment i.e. cells without Actinomycin-D and cells treated with 0.01 µM Actinomycin-D for 24 hr. GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 24 hr of transfection and 24 hr expression of the DNA. The blue bars show normally transfected cells while the red bars show cells treated with 0.01 µM concentration of Actinomycin-D. Error bars show the Standard Error (SE) from 3 independent experiments.

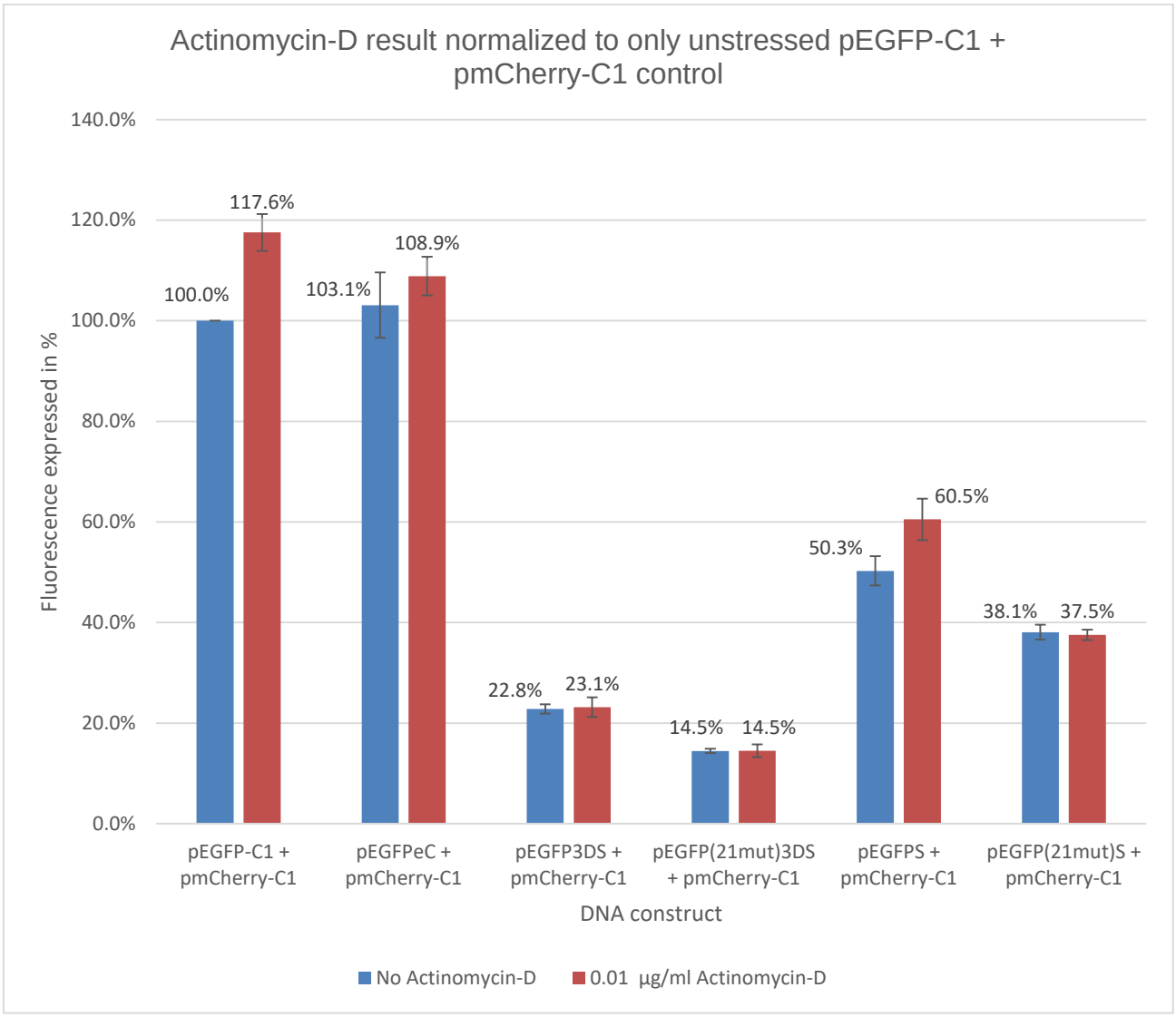


Figure 4.3: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. All results were normalized to only (pEGFP-C1 + pmCherry-C1) without treatment with Actinomycin-D. Cells were analysed after 24 hr of transfection and 24 hr expression of the DNA. The blue bars show normally transfected cells while the red bars show cells treated with 0.01 µM concentration of Actinomycin-D. Error bars show the Standard Error (SE) from 3 independent experiments.

The result suggests that Actinomycin-D treatment is not useful in subsequent analysis to stop transcription of the DNA constructs in cells, because all of the constructs (except pEGFP(21mut)S) if anything had an increase in fluorescence in cells treated with Actinomycin-D when compared to untreated transfected cells.

4.3 Investigation of stress factors

To quantitate the amount of fluorescence produced by each DNA construct after being exposed to different cell stress conditions, the constructs were also co-transfected with pmCherry-C1 and then, the fluorescence of EGFP and mCherry were compared in each case.

The effect of secondary structures on RNA stability in cells responding to osmotic, oxidative and ER stress

Only GMK cells were used in this study because of the adequate transfection efficiency. Each pEGFP DNA construct was co-transfected with pmCherry-C1 and then the cells were introduced to different stresses of a low and high concentration of each (i.e. 140 mM and 400 mM NaCl, 50 µM and 500 µM H_2O_2 and 0.3 µM and 1 µM Thapsigargin) for 6 hr after 6 hr of transfection. 14 hr later, the fluorescence data was obtained. The low cell stress concentration were used for the experiment because the cells were relatively healthy when compared with unstressed cells, while the high stress concentrations were used because the cells were relatively unhealthy when compared with the unstressed cells and so any stress effects should be greater. Unstressed cells and stressed cells (with all stress parameters) were not fixed before analysis.

4.3.1 Analysis of NaCl concentrations for osmotic stress

10 concentrations of NaCl were tested to pick out two different stress conditions i.e. low concentration and high concentration that would be used to carry out the experiment. Cells were split into 11 wells of a 24 well plate and confluent cells were exposed to the 10 different stress concentrations. Two controls were used, an empty well (blank) and untreated cells for normalizing and calculating the relative cell death. The plate was left in an incubator for 6 hr and then analysed using Wst 1 as described in Section 2.2.11 to show the amount of cell damage due to the NaCl by measuring the amount of live cells left in each well.

Figure 4.4 shows the relative cell death and thus, concentration of 140 mM and 400 mM were used as relatively low and high concentrations of NaCl respectively for the experiment.

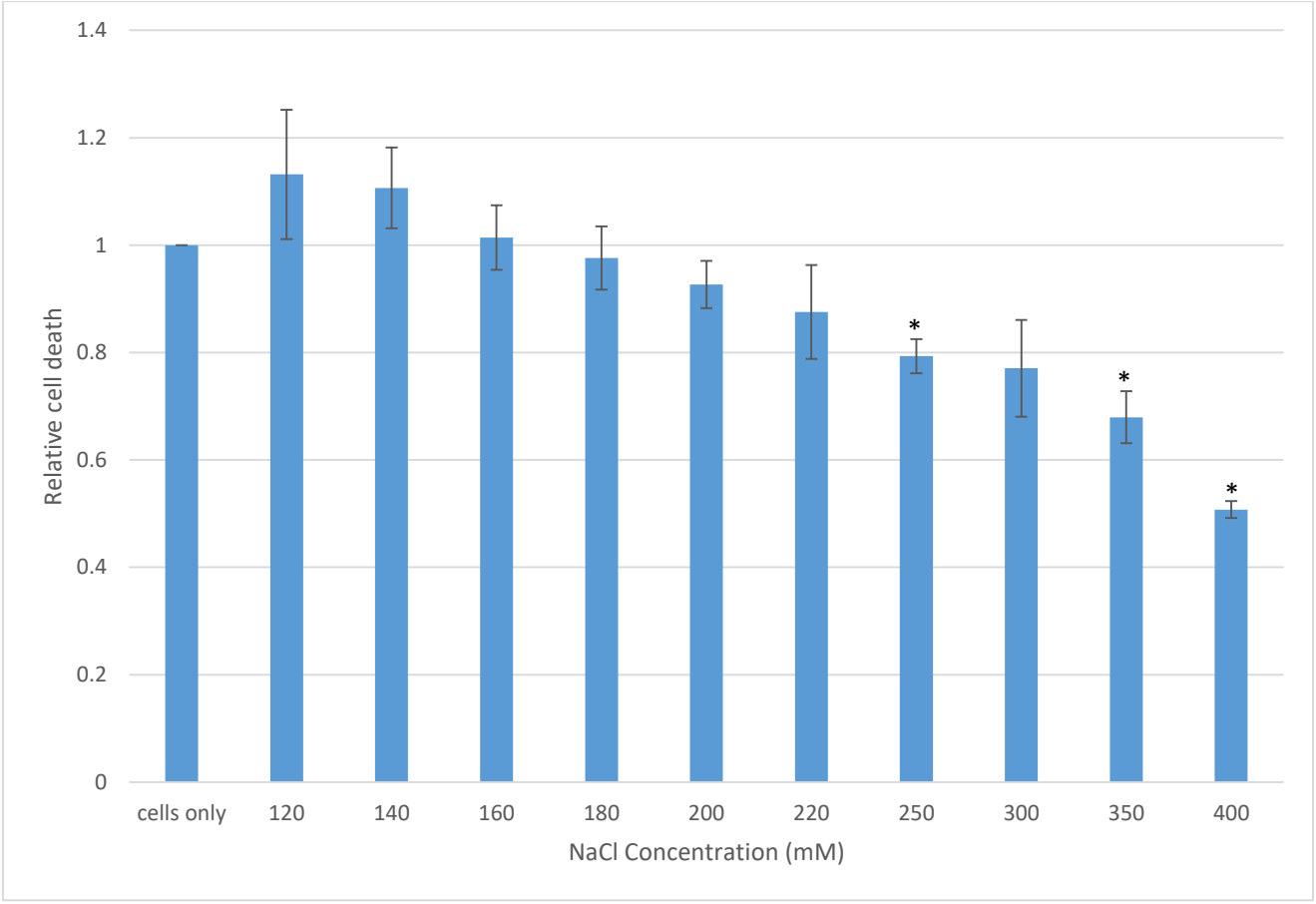


Figure 4.4: GMK cells were treated with different concentrations of NaCl for 6 hr. Wst 1 assay was used to measure the relative cell death compared to untreated (cells only) cells. The error bars show the Standard Error (SE) from 3 independent experiments. * indicates $p < 2.7E-03$.

* indicates significant difference of NaCl concentration to the control, cells only.

4.3.1.1 Normalization to individual (pEGFP-C1 + pmCherry-C1) for each experiment

The two ways of normalising the data described for Actinomycin-D, were used. The data using the appropriate treatment control is shown in Figures 4.5 and 4.7. Normalization was carried out by dividing the amount of fluorescence obtained in each DNA construct by the amount of fluorescence obtained in the (pEGFP-C1 + pmCherry-C1) for each experiment (i.e. in unstressed and stressed cells). The average was used to plot the graphs. The two independent results have been plotted side by side each other to give a direct picture of the expression of the fluorescence observed.

4.3.1.2 Normalization to only unstressed (pEGFP-C1 + pmCherry-C1)

The direct comparison between unstressed and stressed cells with NaCl was studied to see the effect of the stress on the cells. This was done by normalizing all results from the experiment to only the fluorescence expressed by the unstressed control; (pEGFP-C1 + pmCherry-C1). The results shown in Figures 4.6 and 4.8 thus show the 11 result values of constructs which included those both in stressed and unstressed cells, normalized to only (pEGFP-C1 + pmCherry-C1) to give a direct picture of the expression of the fluorescence observed in the 140 mM NaCl stressed cells and also the 400 mM NaCl stressed cells respectively. Surprisingly, some of the constructs, pEGFP-C1, pEGFP(21mut)S and pEGFPHRV(3UTR) were shown to express higher amount of fluorescence in 140 mM NaCl while pEGFP-C1, pEGFPeC, pEGFP(21mut)S and pEGFPHRV(3UTR) expressed higher amount of fluorescence in 400 mM NaCl (Figures 4.6 and 4.8). In the unstressed cells, there is a significant difference between pEGFPS and pEGFP(21mut)S. Also, in both 140 and 400 mM concentration of NaCl, there was a significant difference between stressed pEGFPS and stressed pEGFP(21mut)S but no significant difference between unstressed PEGFPS and stressed pEGFP(21mut)S.

In general, the result of the stress with both the low and high concentration of NaCl show that there is no significant difference between the fluorescence expressed by any particular construct in both unstressed and stressed cells.

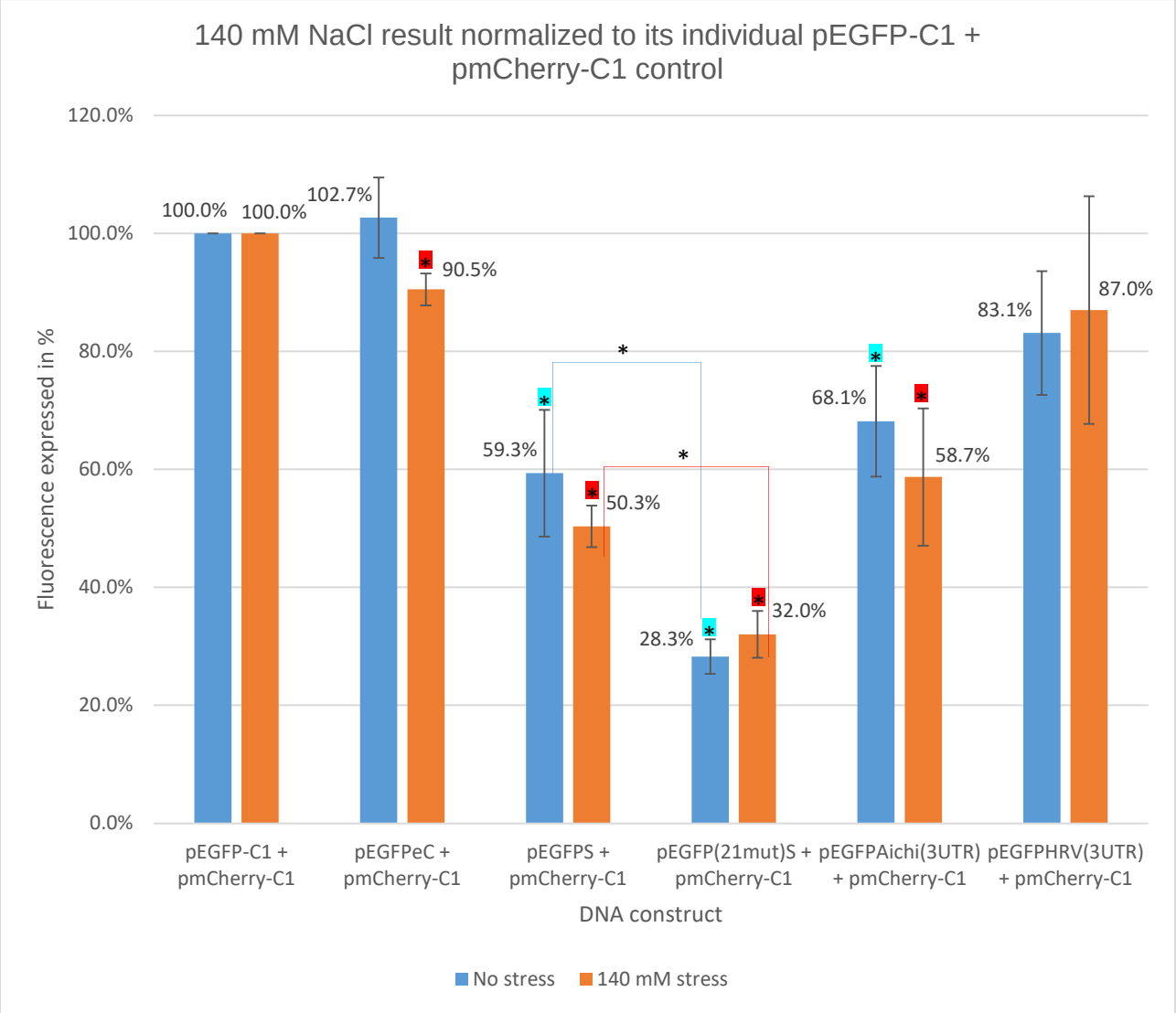


Figure 4.5: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. The result was normalized to (pEGFP-C1 + pmCherry-C1) of each independent experiment i.e. unstressed cells and stressed cells. GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr stress with NaCl. The blue bars show unstressed cells while the red bars show cells treated with the 140 mM concentration of NaCl. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates $p < 2.5E-02$, * indicates $p < 2.7E-02$ and * indicates $p < 4.9E-02$.

* indicates significant difference under 140 mM stress condition

* indicates significant difference under no stress

* indicates significant difference when comparing pEGFPS and pEGFP(21mut)S.

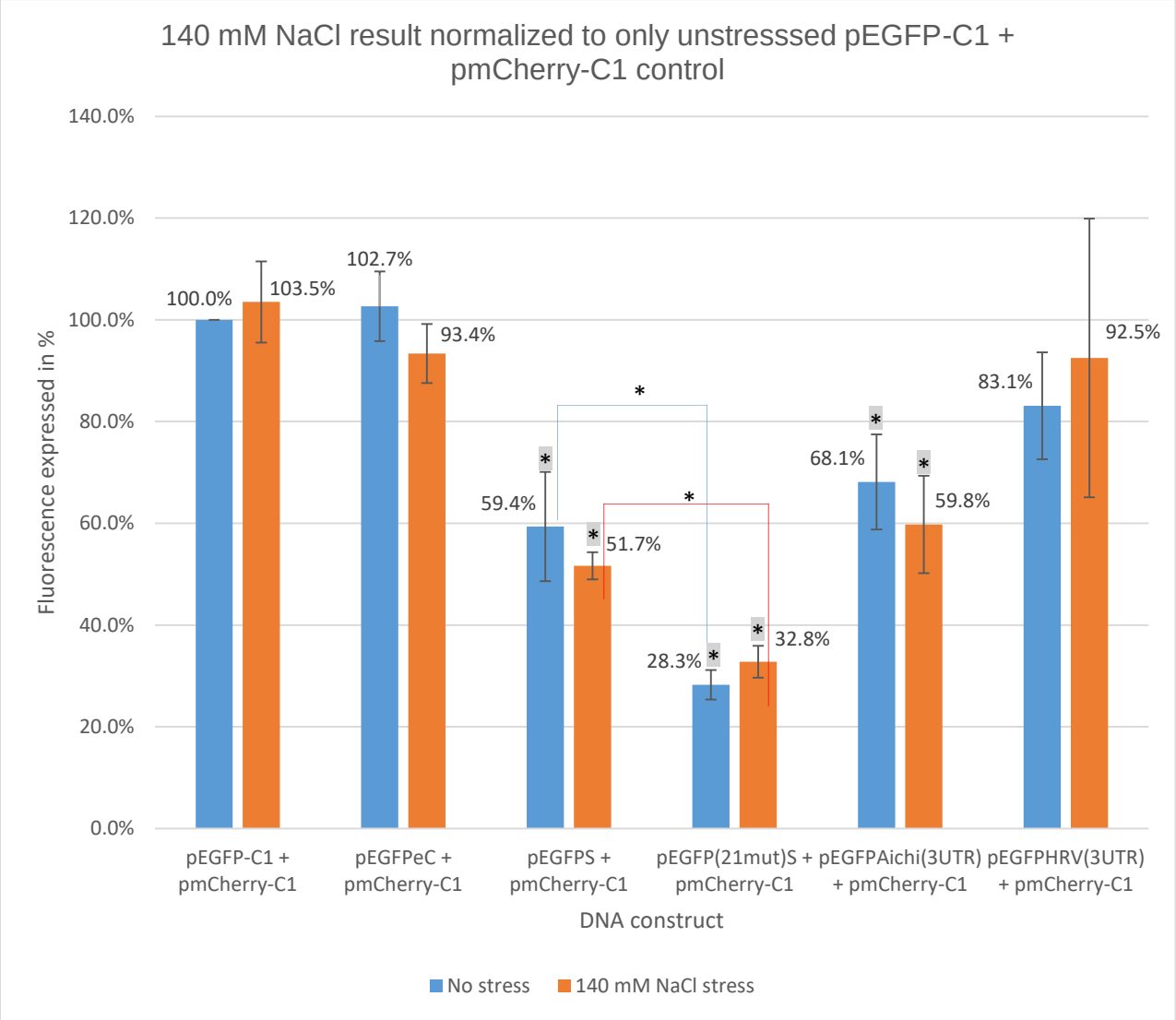


Figure 4.6: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. All results were normalized to only unstressed (pEGFP-C1 + pmCherry-C1). GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr stress with NaCl. The blue bars show unstressed cells while the red bars show cells treated with 140 mM concentration of NaCl. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates $p < 4.9E-02$ and * indicates $p < 2.7E-02$.

* indicates significant difference of all DNA constructs (stressed and unstressed) to the unstressed control, pEGFP-C1.

* indicates significant difference when comparing pEGFPS and pEGFP(21mut)S.

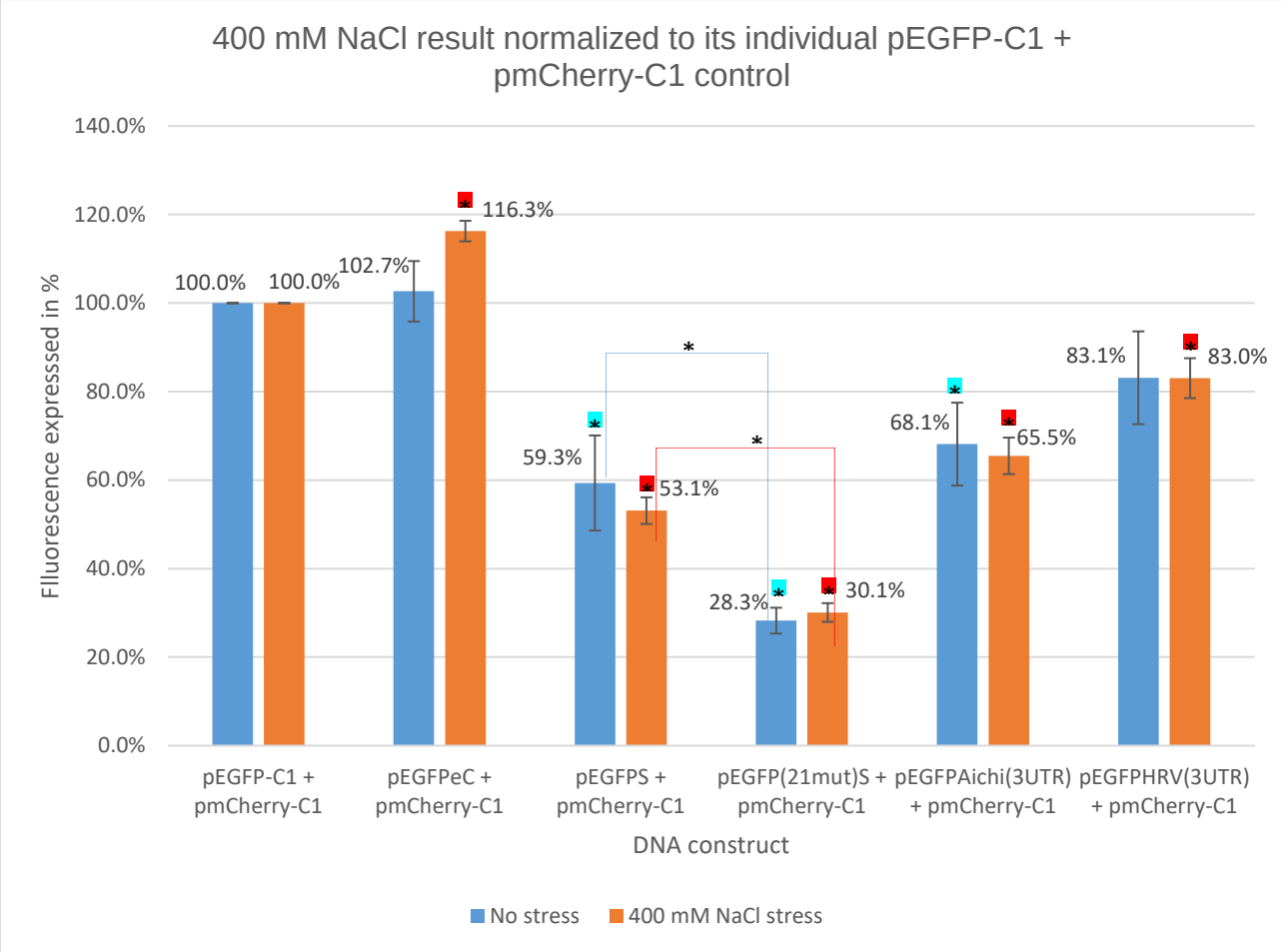


Figure 4.7: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. The result was normalized to (pEGFP-C1 + pmCherry-C1) of each independent experiment i.e. unstressed cells and stressed cells. GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr stress with NaCl. The blue bars show unstressed cells while the red bars show cells treated with 400 mM concentration of NaCl. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates $p < 1.9E-02$, * indicates $p < 2.7E-02$ and * indicates $p < 4.9E-02$.

* indicates significant difference under 400 mM stress condition

* indicates significant difference under no stress

* indicates significant difference when comparing pEGFPS and pEGFP(21mut)S.

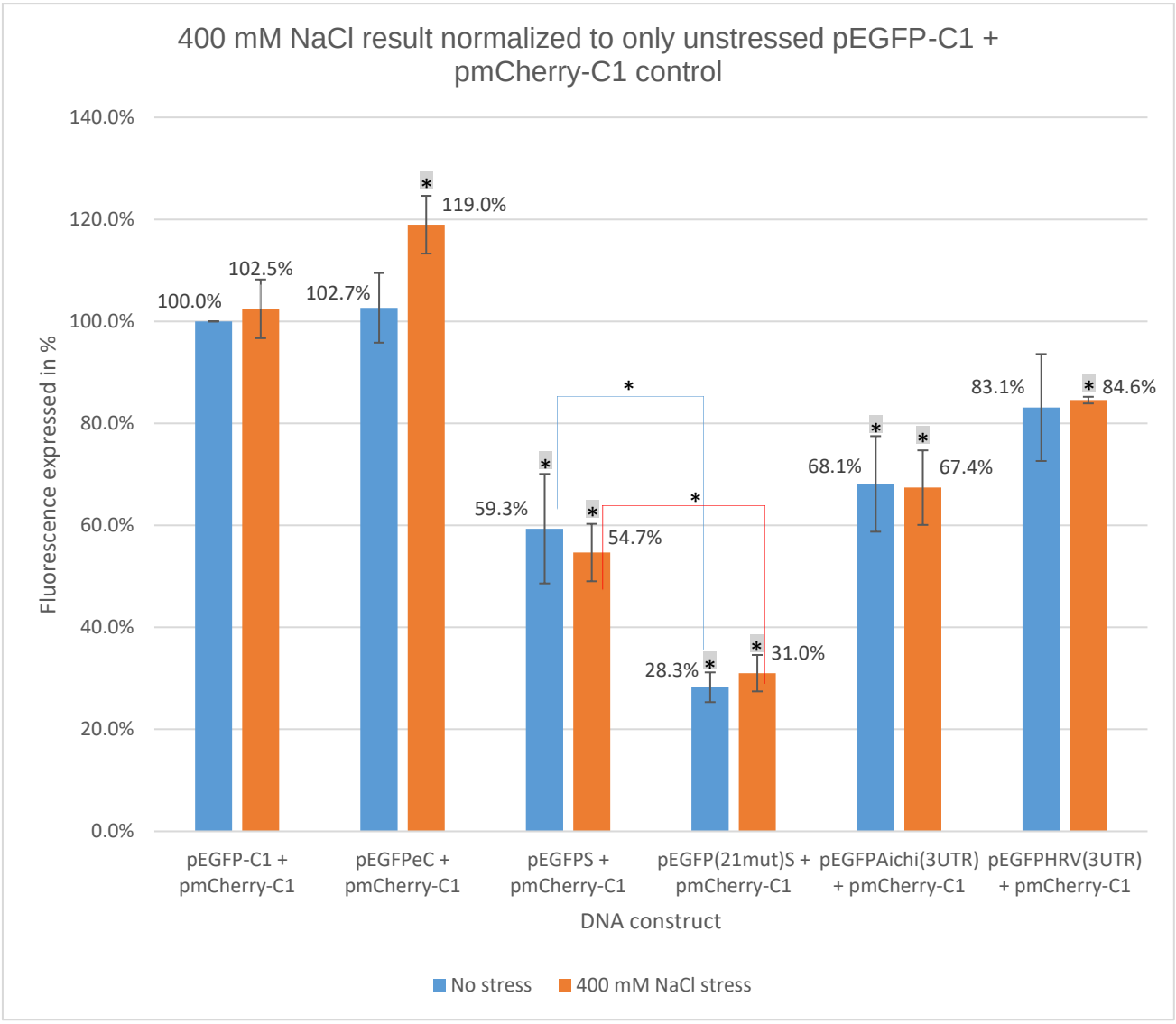


Figure 4.8: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. The result was normalized to (pEGFPeC + pmCherry-C1) of each independent experiment i.e unstressed cells and stressed cells. GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr stress with NaCl. The blue bars show unstressed cells while the red bars show cells treated with 400 mM concentration of NaCl. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates $p < 4.9E-02$ and * indicates $p < 2.9E-02$.

* indicates significant difference of all DNA constructs (stressed and unstressed) to the unstressed control, pEGFP-C1

* indicates significant difference when comparing pEGFPS and pEGFP(21mut)S.

4.3.2 **Analysis of H_2O_2 concentrations for oxidative stress**

Oxidative stress was tested using hydrogen peroxide (H_2O_2) (Halliwell *et al.*, 2000). Different concentrations were analysed to pick out two different stress conditions i.e. one of lower concentration and the other of a higher concentration that would be used to carry out the experiment. 9 different stress concentrations were introduced to confluent cells in a 24 well plate also containing two controls; an empty well and cells only for normalizing and calculating the relative cell death. This was left in an incubator for 6 hr and then analysed using Wst 1 as described in Section 2.2.11 to show the amount of cell damage due to the H_2O_2 by measuring the amount of cells left in each well. Figure 4.9 shows the relative cell death. Concentration of 50 µM and 500 µM were chosen as low and high concentrations respectively for the experiment.

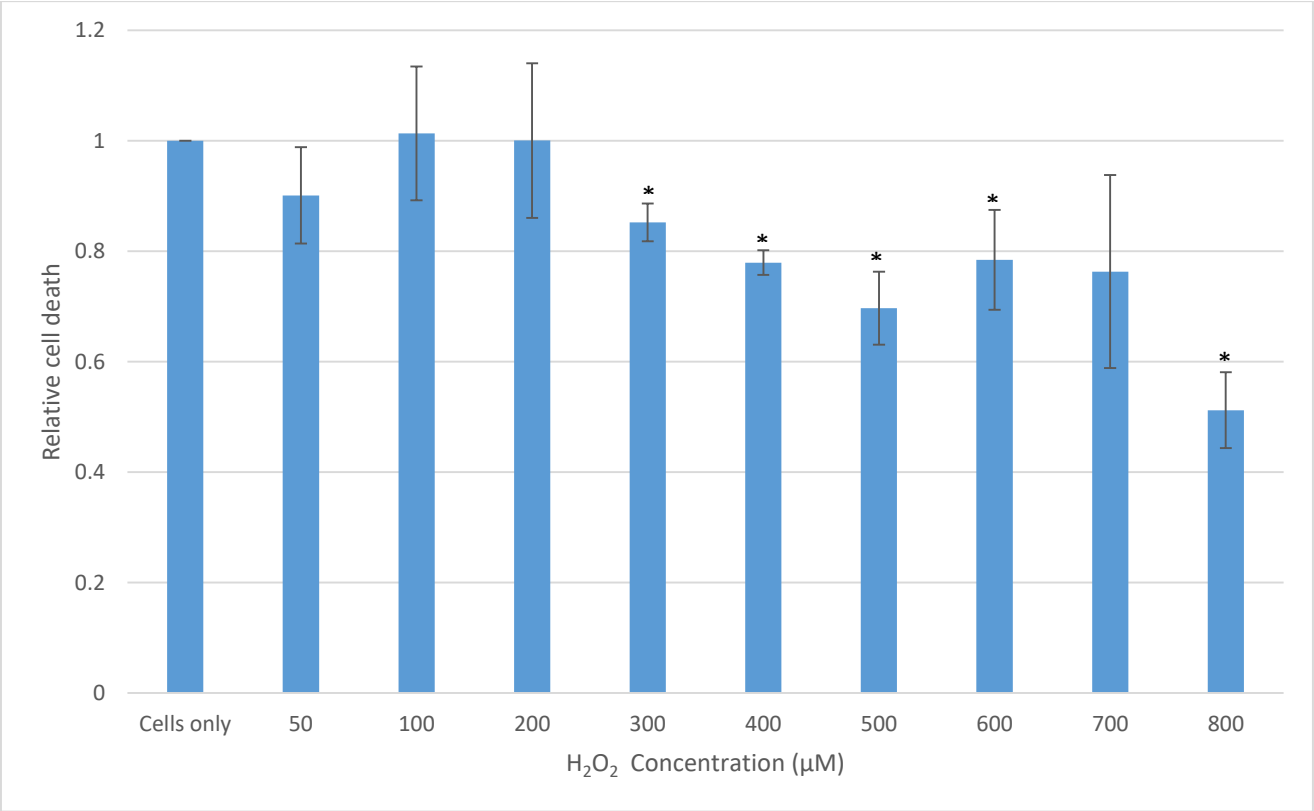


Figure 4.9: GMK cells were treated with different concentrations of H_2O_2 for 6 hr. Wst 1 assay was used to measure the relative cell death compared to untreated (cells only) cells. The error bars show the Standard Error (SE) from 2 independent experiments. * indicates $p < 1.0E-02$.

* indicates significant difference of H_2O_2 concentration to the control, cells only.

4.3.2.1 Normalization to (pEGFP-C1 + pmCherry-C1) for each experiment

The relationship between unstressed cells and cells stressed with relatively high (500 µM) and low (50 µM) concentrations of H_2O_2 were studied as individually different experiments and normalized to the fluorescence expressed by the controls; unstressed and stressed (pEGFP-C1 + pmCherry-C1). The results are shown in Figures 4.10 and 4.12.

In the unstressed cells, there is a clear difference between pEGFPS and pEGFP(21mut)S. At the 50 µM concentration of H_2O_2, pEGFPS has a significantly higher fluorescence than pEGFP(21mut)S, suggesting that the structure has an effect under these conditions (Figure 4.10). However, the same trend is not seen at 500 µM H_2O_2 concentration (Figure 4.12).

4.3.2.2 Normalization to only unstressed (pEGFP-C1 + pmCherry-C1)

The direct comparison between unstressed cells and cells stressed with relatively high (500 µM) and low (50 µM) concentrations of H_2O_2 was studied by normalizing all results from the experiment to only the fluorescence expressed by the unstressed control (pEGFP-C1 + pmCherry-C1). This was done to see how much of the cells would be affected by the stress when compared with the unstressed control. The results are represented in Figures 4.11 and 4.13 and they thus show the 11 values of constructs which included those both in stressed and unstressed cells, normalized to only (pEGFP-C1 + pmCherry-C1) to give a picture of the expression of the fluorescence observed. The 50 µM H_2O_2 stress has little effect on the construct overall, but fluorescence is greatly reduced substantially at 500 µM H_2O_2 (Figures 4.11 and 4.13).

The result of the fluorescence expressed by constructs in 50 µM and 500 µM concentration of H_2O_2 show that there is no clear difference between the fluorescence

expressed by any particular construct in both unstressed and stressed cells (Figures 4.10-4.13).

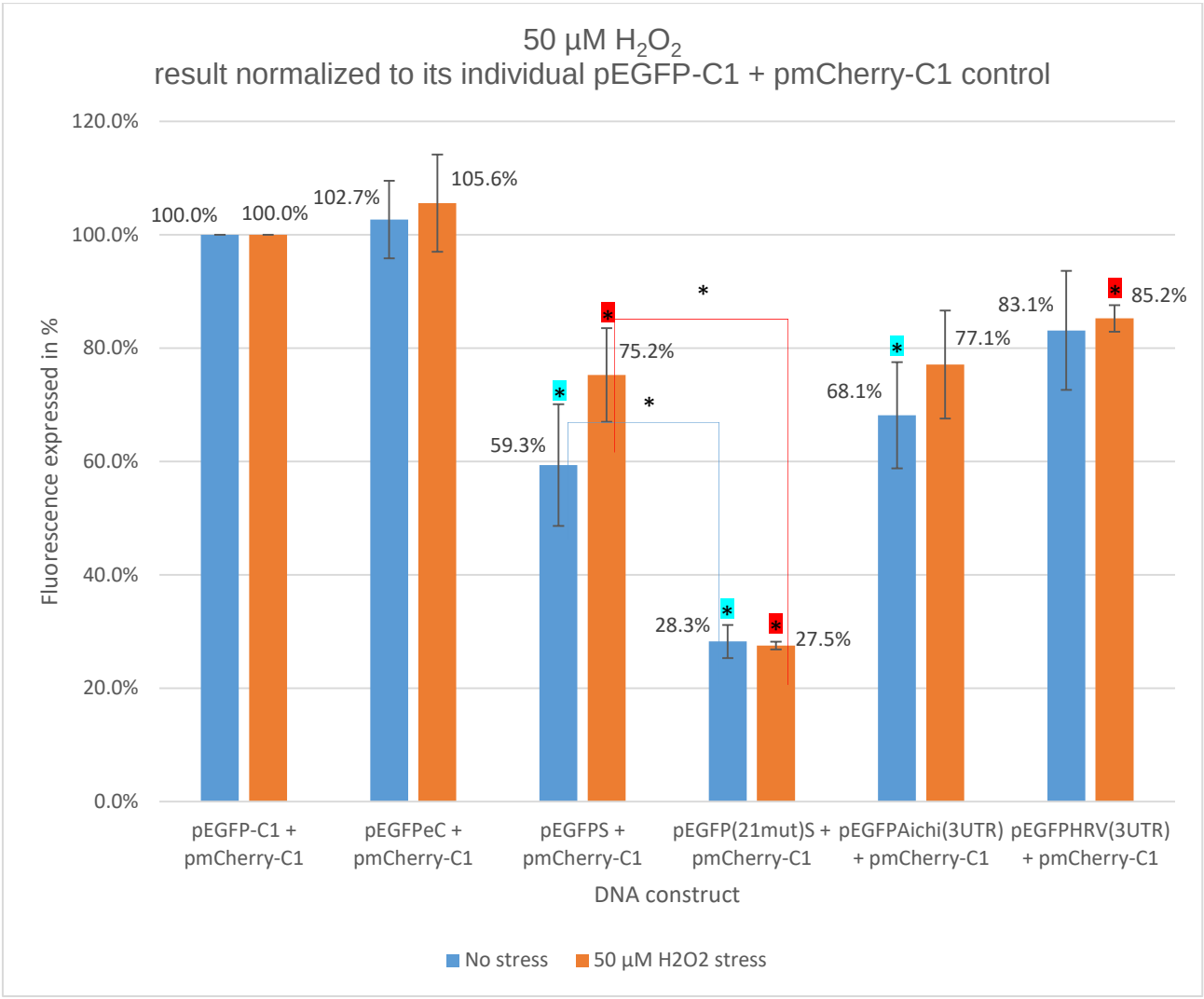


Figure 4.10: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. The result was normalized to (pEGFP-C1 + pmCherry-C1) of each independent experiment i.e. unstressed cells and stressed cells. GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr stress with H_2O_2. The blue bars show unstressed cells while the red bars show cells treated with 50 µM concentration of H_2O_2. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates p < 4.0E-02, * indicates p < 2.7E-02 and * indicates p < 4.9E-02.

* indicates significant difference under 50 µM stress condition

* indicates significant difference under no stress

* indicates significant difference when comparing pEGFPS and pEGFP(21mut)S.

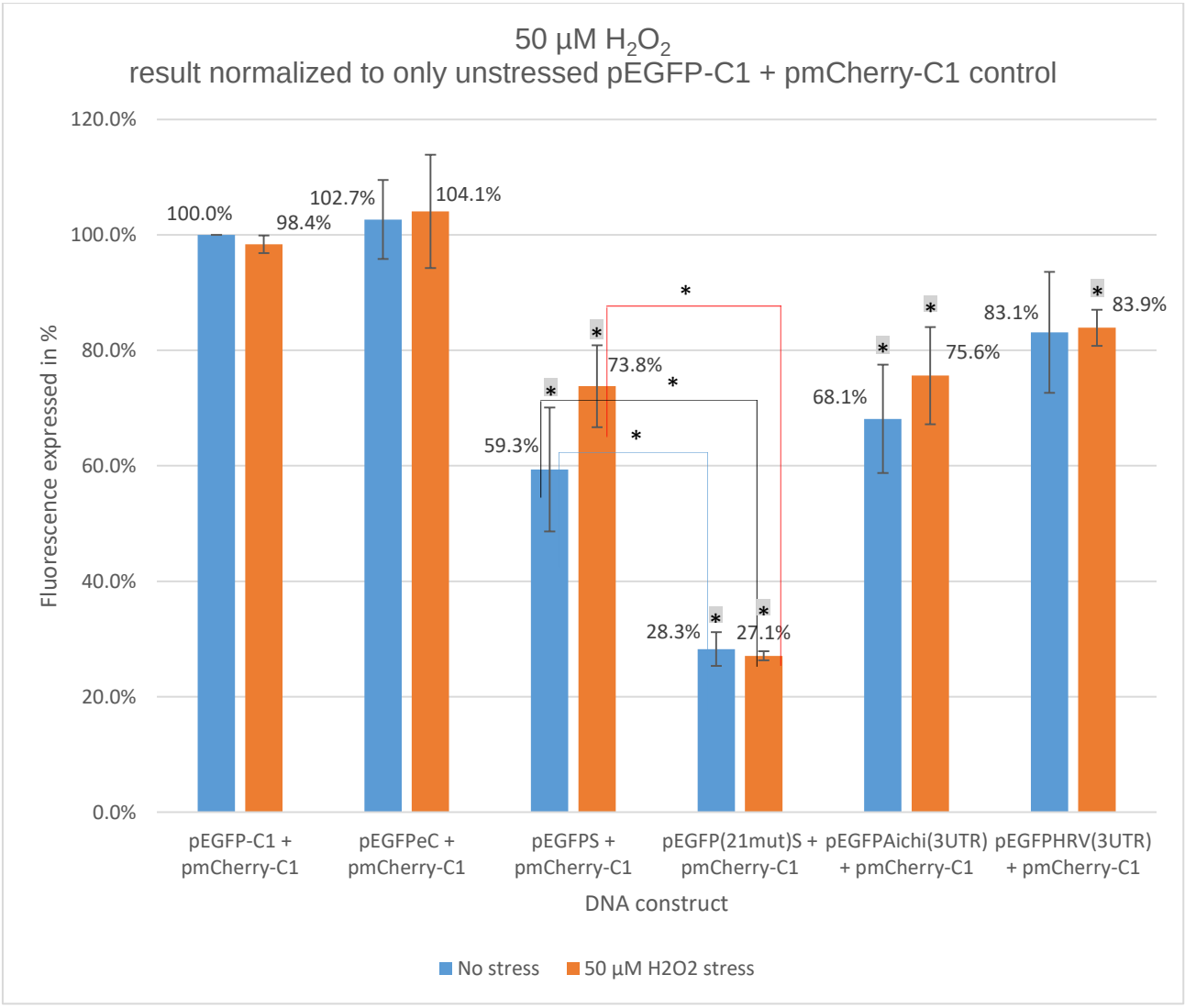


Figure 4.11: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. All results were normalized to unstressed (pEGFPeC + pmCherry-C1). GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr stress with H_2O_2. The blue bars show unstressed cells while the red bars show cells treated with 50 µM concentration of H_2O_2. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates $p < 4.9E-02$ and * indicates $p < 4.4E-02$.

* indicates significant difference of all DNA constructs (stressed and unstressed) to the unstressed control, pEGFP-C1

* indicates significant difference when comparing pEGFPS and pEGFP(21mut)S.

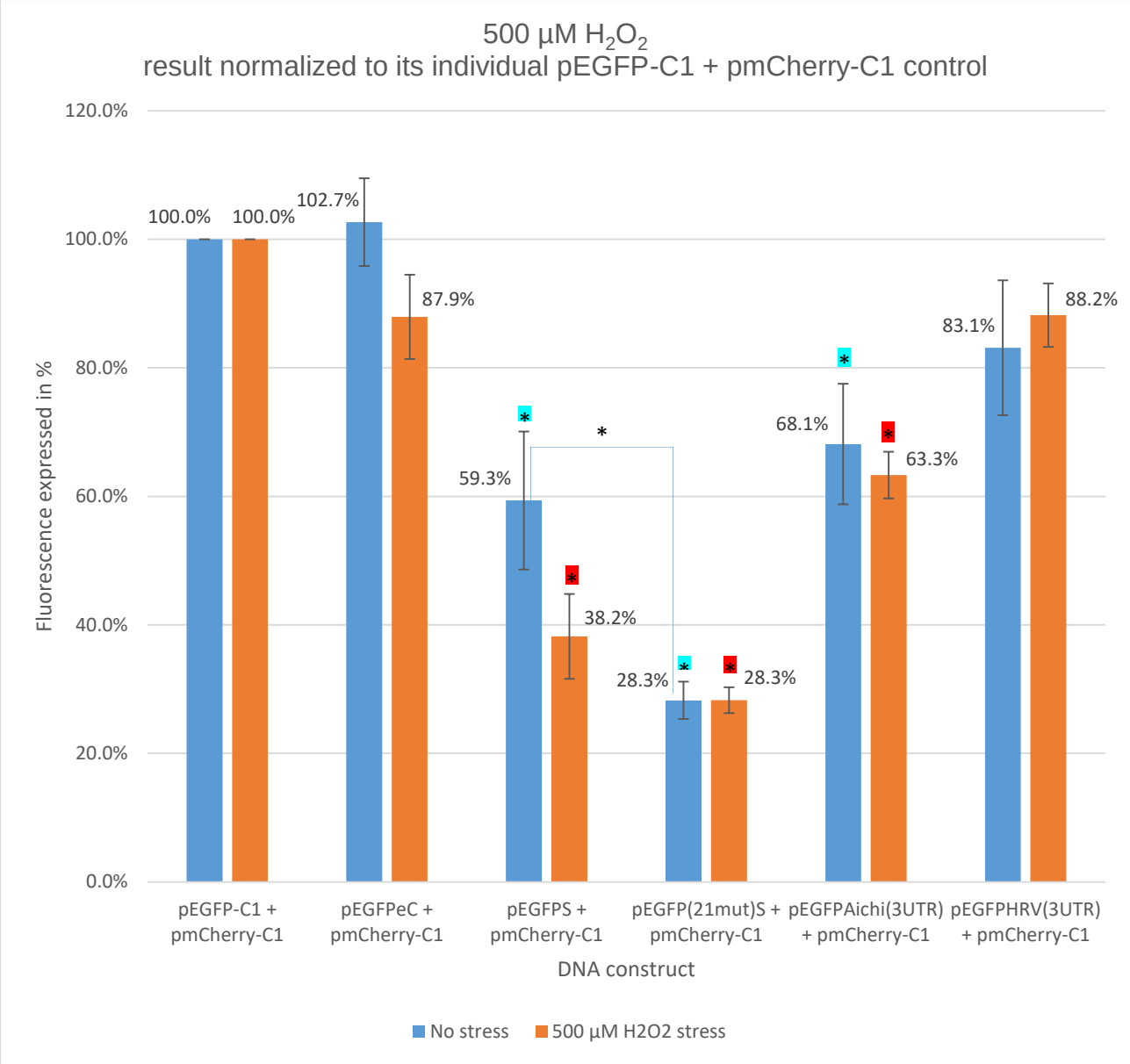


Figure 4.12: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. The result was normalized to (pEGFP-C1 + pmCherry-C1) of each independent experiment i.e. unstressed cells and stressed cells. GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr stress with H_2O_2. The blue bars show unstressed cells while the red bars show cells treated with 500 µM concentration of H_2O_2. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates p < 7.3E-04, * indicates p < 2.7E-02 and * indicates p = 4.9E-02.

* indicates significant difference under 500 µM stress condition

* indicates significant difference under no stress

* indicates significant difference when comparing pEGFPS and pEGFP(21mut)S.

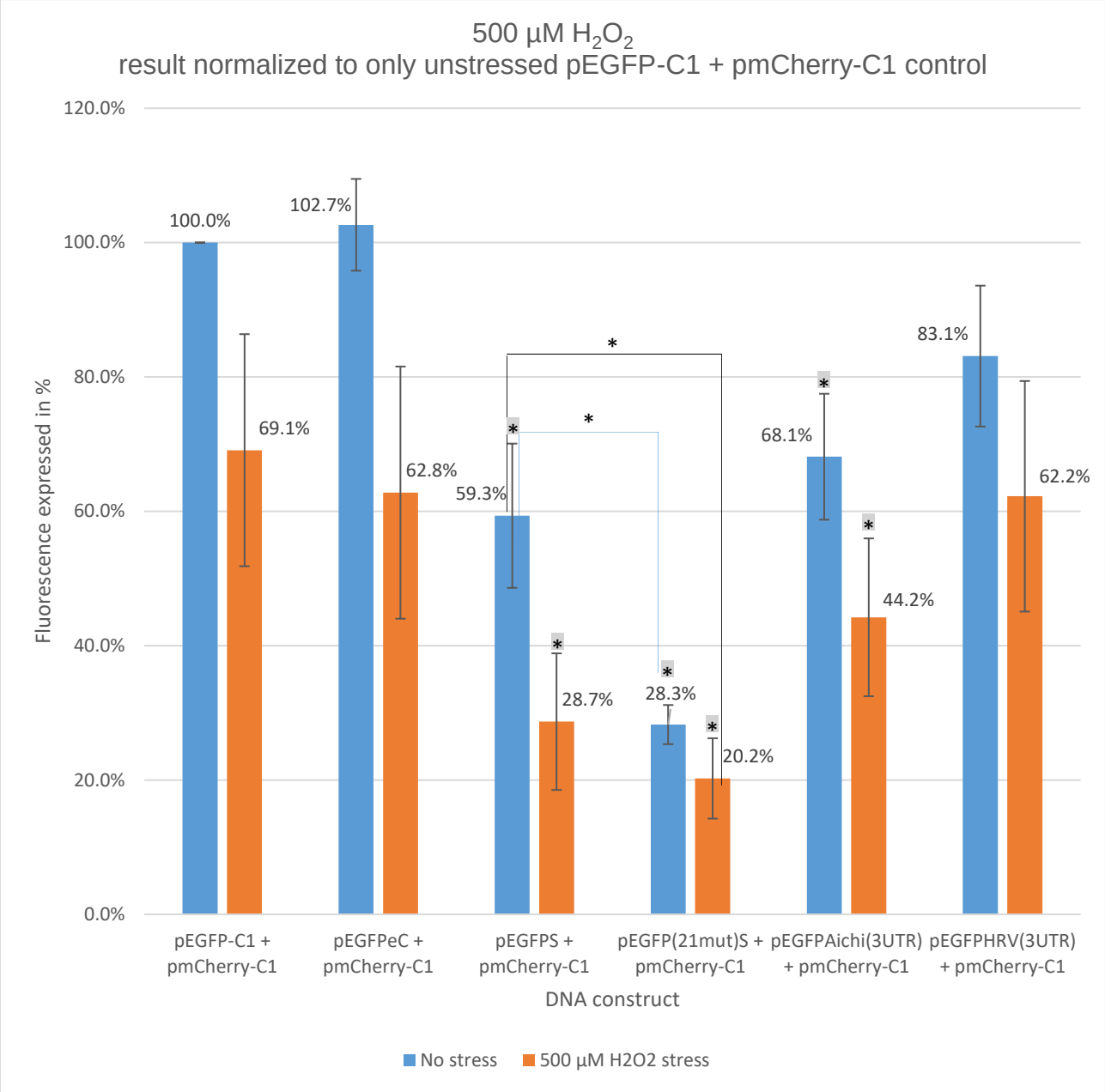


Figure 4.13: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. The result was normalized to (pEGFPeC + pmCherry-C1) of each independent experiment i.e unstressed cells and stressed cells. GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr stress with H_2O_2. The blue bars show unstressed cells while the red bars show cells treated with 500 µM concentration of H_2O_2. Error bars show the Standard error (SE) from 3 independent experiments. * indicates $p < 4.9E-02$ and * indicates $p < 2.7E-02$.

* indicates significant difference of all DNA constructs (stressed and unstressed) to the unstressed control, pEGFP-C1

* indicates significant difference when comparing pEGFPS and pEGFP(21mut)S.

4.3.3 **Analysis of Thapsigargin concentrations for ER stress**

The structure in the 3D polymerase encoding region was tested under endoplasmic reticulum (ER) stress using Thapsigargin, an inducer of the unfolded protein response (UPR) (He *et al.*, 2006) and an inhibitor of Ca^{2+}-ATPase in the ER, which causes an increase in the concentration of calcium within the cells leading to ER stress (Ruan *et al.*, 2007). 5 concentrations of Thapsigargin were analysed to pick out two different stress conditions i.e. low concentration and high concentration that would be used to carry out the experiment. Cells were split into 4 wells of a 24 well plate and confluent cells were exposed to the 5 different stress concentrations. Two controls, an empty well and untreated cells were used for normalizing and calculating the relative cell death. The plate was left in an incubator for 6 hr and then also analysed using Wst 1 assay to show the amount of cell damage due to the Thapsigargin by measuring the amount of live cells left in each well. Figure 4.14 shows the relative cell death. Concentrations of 0.3 µM and 1 µM were used as low and high concentrations respectively for the experiment.

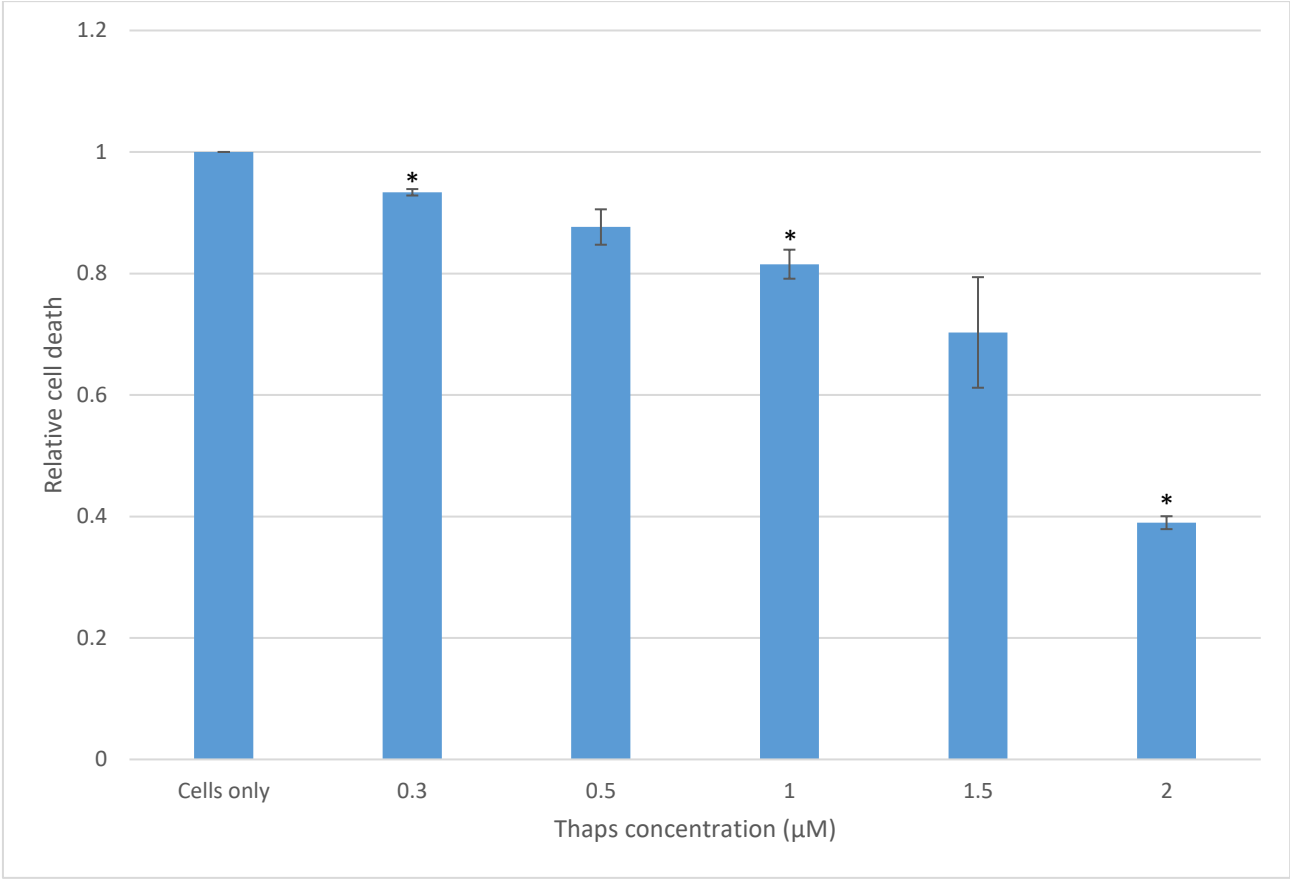


Figure4.14: GMK cells were treated with different concentrations of Thapsigargin for 6 hr. Wst 1 assay was used to measure the relative cell death compared to untreated (cells only) cells. The error bars show the Standard Error (SE) from 2 independent experiments. * indicates $p < 2.0E-02$

* indicates significant difference of Thapsigargin concentration to the control, cells only.

4.3.3.1 Normalization to (pEGFP-C1 + pmCherry-C1) for each experiment

Normalization was carried out by dividing the amount of fluorescence obtained in each DNA construct by the amount of fluorescence obtained in (pEGFP-C1 + pmCherry-C1) for each individual experiment (in unstressed and stressed cells) and then the average was calculated and used to plot the graph. Hence, all the results for no stress was normalized to the control; (pEGFP-C1 + pmCherry-C1) with 'no stress' while all the results for 'stress with Thapsigargin' was normalized to the control; (pEGFP-C1 + pmCherry-C1) which was stressed with Thapsigargin. The two independent results have been plotted side by side each other to give a direct picture of the expression of the fluorescence observed.

4.3.3.2 Normalization to only untreated (pEGFP-C1 + pmCherry-C1)

The direct comparison between stress and unstressed cells treated with relatively high (1 μM) and low Thapsigargin (0.3 μM) was studied by normalizing all results from the experiment to only the fluorescence expressed by the unstressed control; (pEGFP-C1 + pmCherry-C1). The results shown in Figures 4.16 and 4.18 thus show the 11 result values of constructs which included those both in stressed and unstressed cells, normalized to only (pEGFP-C1 + pmCherry-C1) to give a direct picture of the expression of the fluorescence observed.

At the low Thapsigargin concentration, all the constructs gave increased fluorescence compared to the pEGFP-C1 + pmCherry-C1 control (Figure 4.15). This seemed to be due to a larger effect of the drug on the control than on the other constructs (Figure 4.16). At the higher concentration, there is a substantial reduction in fluorescence.

The result of the fluorescence expressed by constructs in 0.3 μM and 1 μM concentration of Thapsigargin show that there is no significant difference between the fluorescence

expressed by any particular construct in both unstressed and stressed cells. However, In 0.3 μM Thapsigargin, the constructs, pEGFPeC, pEGFPS, pEGFP(21mut)S and pEGFPAichi(3UTR) were shown to express higher amount of fluorescence while in 1 μM concentration of Thapsigargin, all constructs expressed lower amount of fluorescence when compared with the unstressed cells (Figures 4.16 and 4.18).

In unstressed cells, there is a significant difference between pEGFPS and pEGFP(21mut)S. In 0.3 μM concentration of Thapsigargin, there was no significant difference between stressed pEGFPS and stressed pEGFP(21mut)S and also between unstressed PEGFPS and stressed pEGFP(21mut)S. In 1 μM concentration of H_2O_2, there was no significant difference between stressed pEGFPS and stressed pEGFP(21mut)S however, there was a significant difference between unstressed PEGFPS and stressed pEGFP(21mut)S.

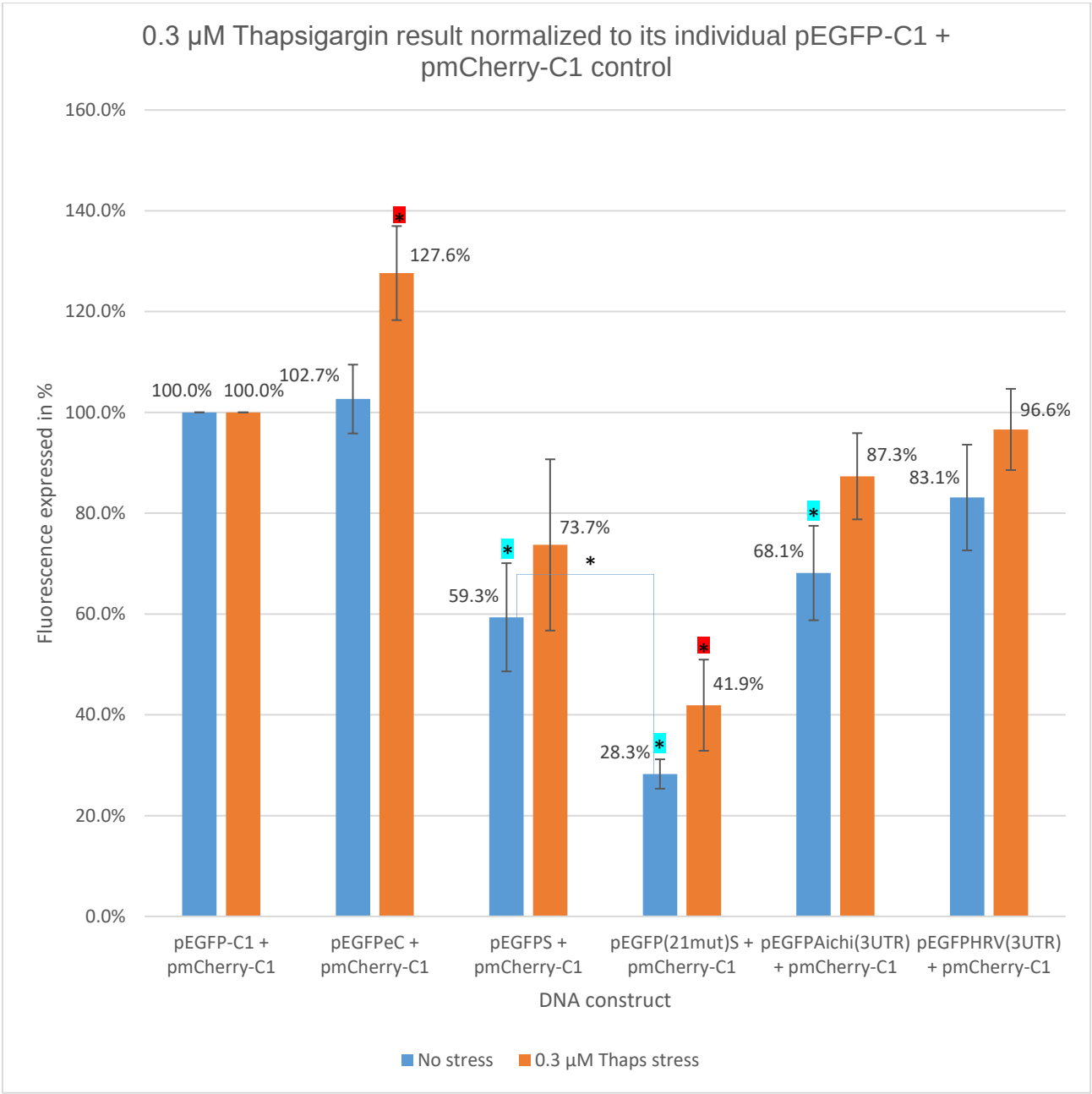


Figure 4.15: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. The result was normalized to (pEGFP-C1 + pmCherry-C1) of each independent experiment i.e. unstressed cells and stressed cells. GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr stress with Thapsigargin. The blue bars show unstressed cells while the red bars show cells treated with 0.3 µM concentration of Thapsigargin. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates $p < 4.2E-02$, * indicates $p < 2.7E-02$ and * indicates $p = 4.9E-02$.

* indicates significant difference under 0.3 µM stress condition

* indicates significant difference under no stress

* indicates significant difference when comparing pEGFPS and pEGFP(21mut)S.

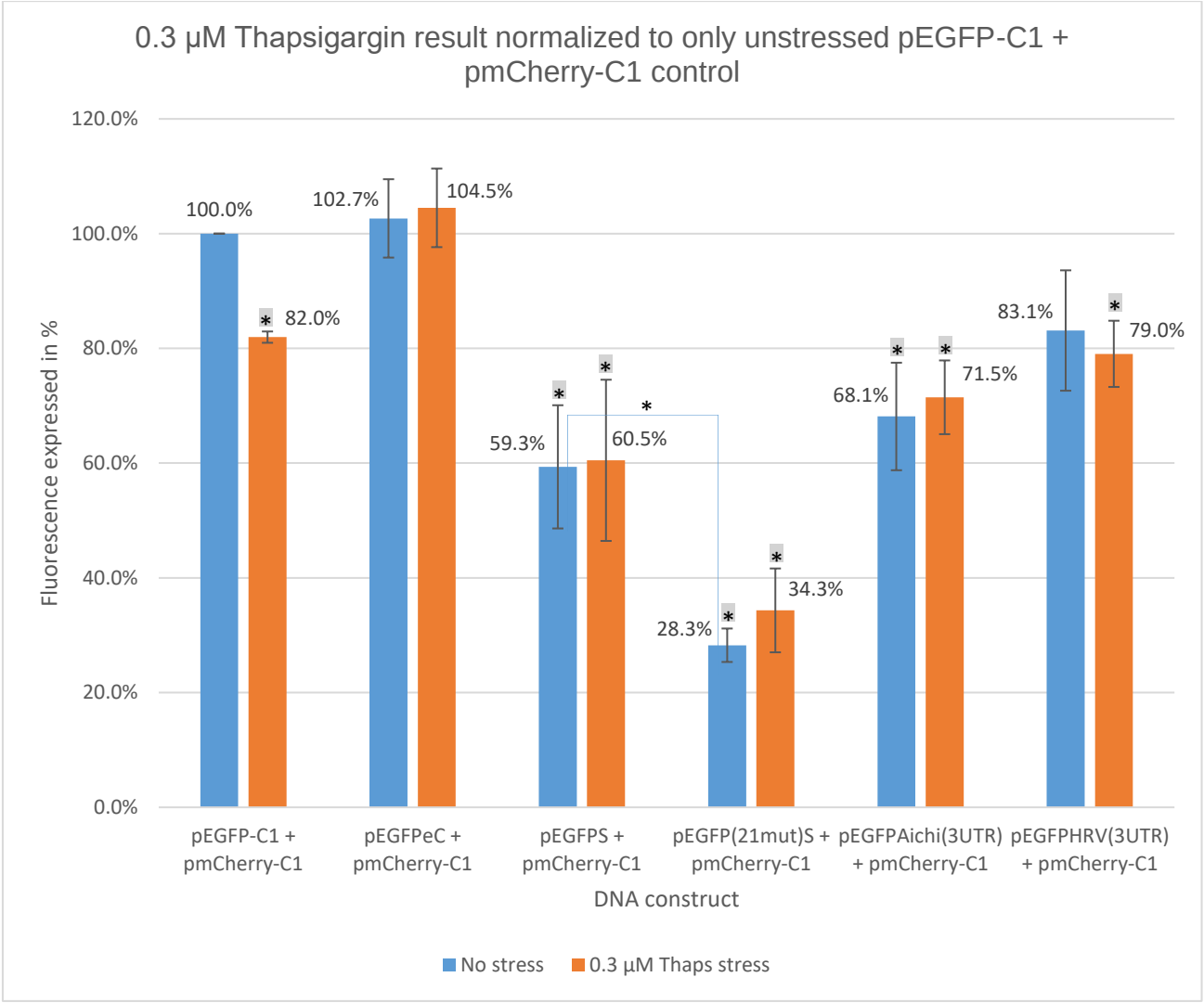


Figure 4.16: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. The result was normalized to (pEGFPeC + pmCherry-C1) of each independent experiment i.e unstressed cells and stressed cells. GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr stress with Thapsigargin. The blue bars show unstressed cells while the red bars show cells treated with 0.3 µM concentration of Thapsigiargin. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates p = 4.9E-02 and * indicates p < 4.8E-02.

* indicates significant difference of all DNA constructs (stressed and unstressed) to the unstressed control, pEGFP-C1

* indicates significant difference when comparing pEGFPS and pEGFP(21mut)S.

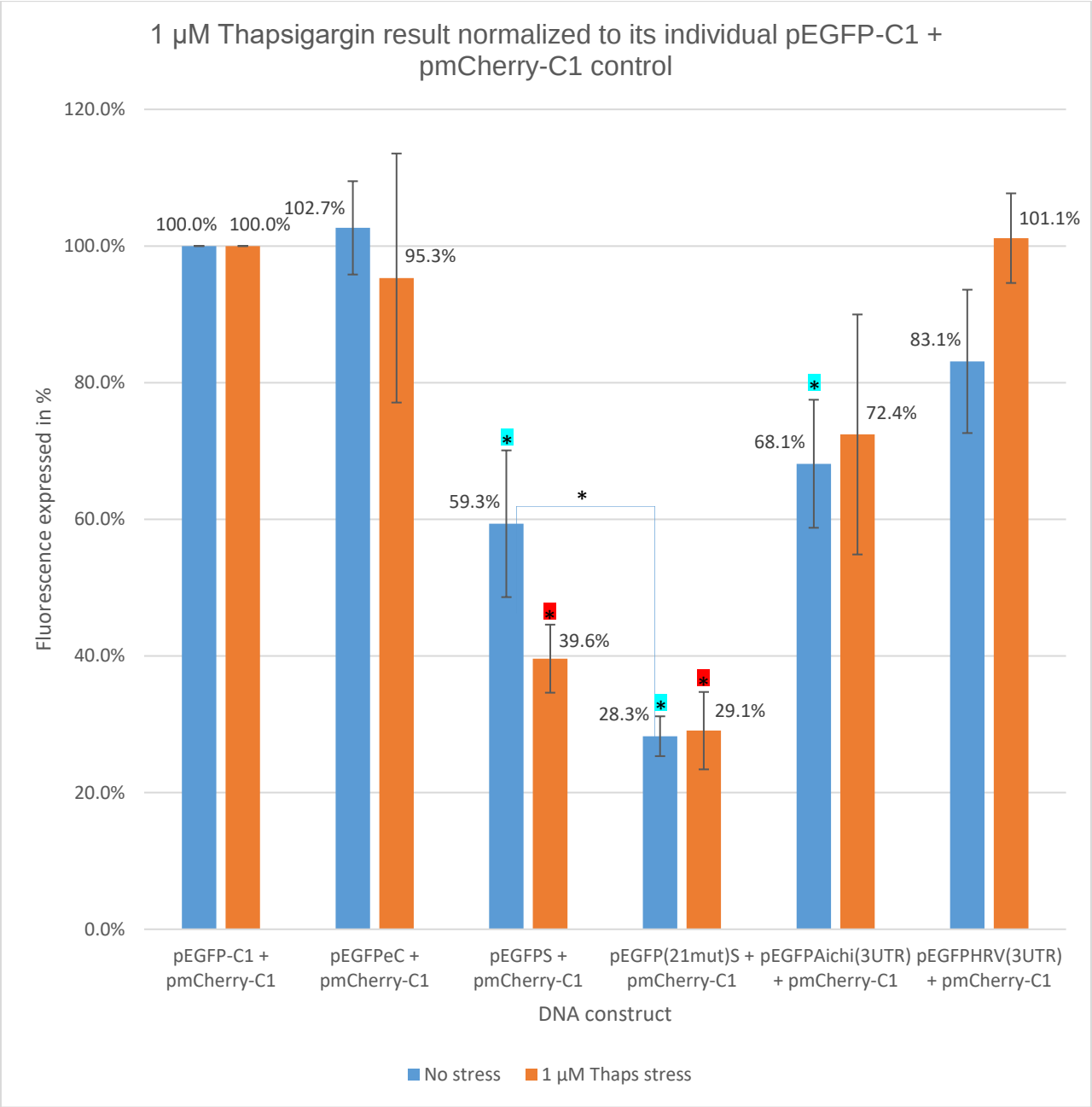


Figure 4.17: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. The result was normalized to (pEGFP-C1 + pmCherry-C1) of each independent experiment i.e. unstressed cells and stressed cells. GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr stress with Thapsigargin. The blue bars show unstressed cells while the red bars show cells treated with 1 µM concentration of Thapsigiargin. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates $p < 2.7E-04$, * indicates $p < 2.7E-02$ and * indicates $p = 4.9E-02$.

* indicates significant difference under 1.0 µM stress condition

* indicates significant difference under no stress

* indicates significant difference when comparing pEGFPS and pEGFP(21mut)S.

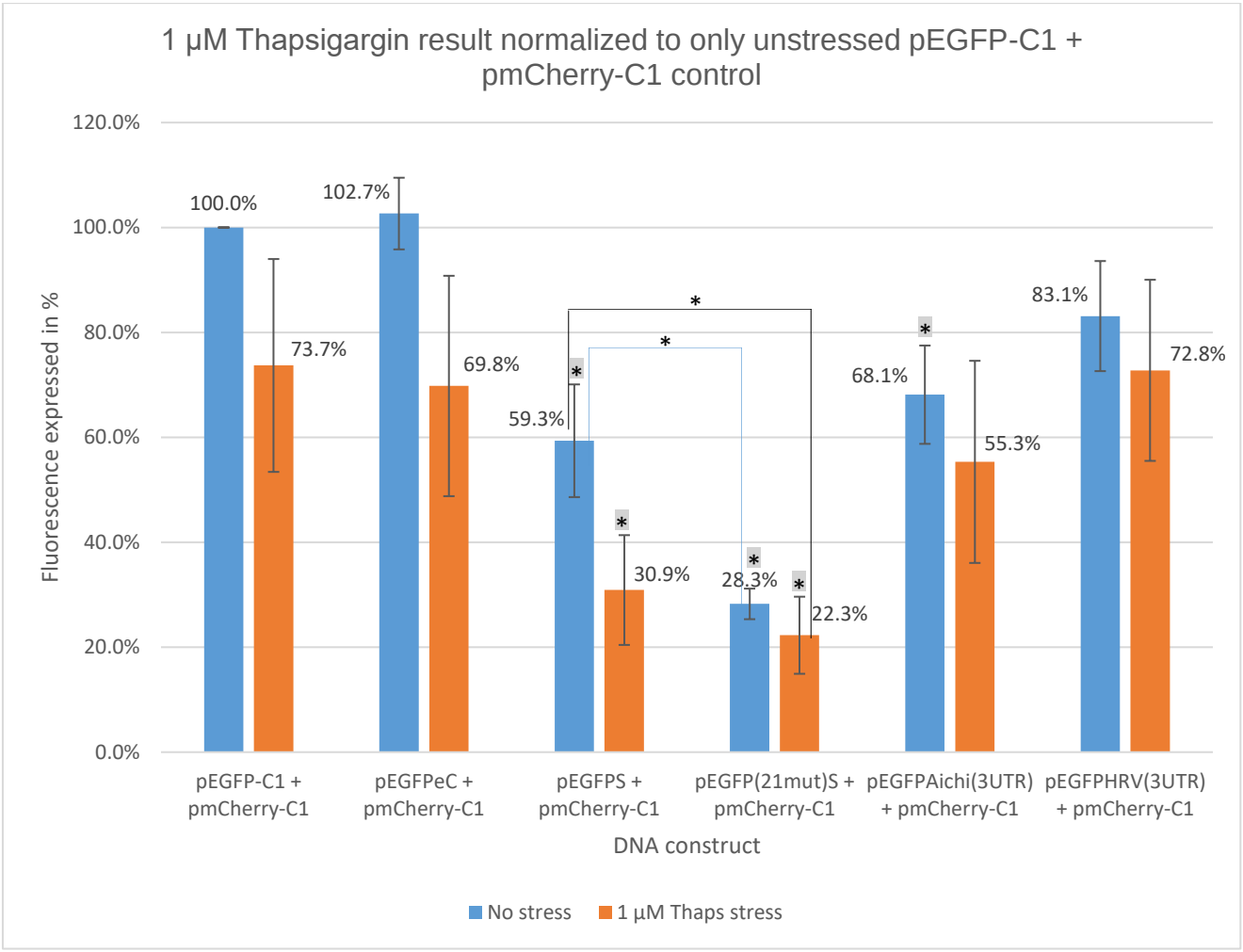


Figure 4.18: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. Result was normalized to (pEGFPeC + pmCherry-C1) of each independent experiment i.e unstressed cells and stressed cells. GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr stress with Thapsigargin. The blue bars show unstressed cells while the red bars show cells treated with 1 µM concentration of Thapsigargin. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates p < 4.9E-02 and * indicates p < 2.7E-02.

* indicates significant difference of all DNA constructs (stressed and unstressed) to the unstressed control, pEGFP-C1

* indicates significant difference when comparing pEGFPS and pEGFP(21mut)S.

4.4 The effect of secondary structures on RNA stability in cells responding to stress due to infection

HPeV1 was used to infect transfected HT29 cells while CAV9 was used to infect transfected GMK cells in this study.

Each DNA construct was co-transfected with pmCherry-C1 for 6 hr then 14 hr after media change, the cells were then infected with the virus for 6 hr. All cells (non-infected and infected) were fixed with 4% paraformaldehyde in ice for 30 min before analysis to deactivate the virus.

4.4.1 Quantification of HPeV1 used for Infection of cells transfected with DNA constructs

Plaque assay was used to determine the *pfu*/ml of HPeV1 (2.49×10^4 pfu/ml) which was used to calculate the MOI value to be 2.5 that was used to infect cells already co-transfected with the DNA constructs and pmCherry-C1 using Lipofectin as described in Section 2.2.2.1. The aim of this experiment was to investigate how the constructs behave in virus-infected cells.

4.4.2 Quantification of CAV9 used for Infection of cells transfected with DNAs

Plaque assay was used to determine the *pfu*/ml of CAV9 (1.26×10^5 pfu/ml) which was used to calculate the MOI value (12.6) that was used to infect cells already co-transfected with the DNA constructs and pmCherry-C1 was also carried out using Lipofectin.

4.4.2.1 Normalization to (pEGFP-C1 + pmCherry-C1) for each experiment

The relationship between non-infected and infected cells when normalized to the fluorescence expressed by the control (pEGFP-C1 + pmCherry-C1) are shown in Figures 4.19 and 4.21. Normalization was carried out by dividing the amount of fluorescence obtained from each DNA construct by the amount of fluorescence obtained from (pEGFP-C1 + pmCherry-C1) for each experiment (in non-infected and infected cells) and then taking the average which was used to plot the graphs.

4.4.2.2 Normalization to only non-infected (pEGFP-C1 + pmCherry-C1)

The direct comparison between non-infected and infected HT29 cells was studied by normalizing all results from the experiment only to the fluorescence expressed by the non-infected control (pEGFP-C1 + pmCherry-C1). This was done to measure the amount of fluorescence in the infected cells when compared to the non-infected control to see the effect of infection overall. The results shown in Figures 4.20 and 4.22 thus show the 11 result values of constructs which included those both in stressed and unstressed cells, normalized to only (pEGFP-C1 + pmCherry-C1) to give a direct picture of the expression of the fluorescence observed.

HT29 cells were used, as HPeV1 grows better in these cells than in GMK cells. In uninfected cells, the overall pattern of results was the same as for the experiments in GMK cells, except that the pEGFPHRV(3UTR) construct gave low fluorescence values (Figure 4.19). Infection reduced the fluorescence of the mutant construct, while the pEGFPS construct was unaltered, or even increased. However, the pEGFPeC control gave a similar result, so the pEGFPS result may not be significant. In CAV9-infected cells, there was high variability for some constructs, but the positive effect of the 3DS structure seen in HPeV1-infected cells was not observed (Figure 4.21).

Infection clearly reduced the overall level of fluorescence, which could be due to virus inhibition of transcription or translation or effects on RNA metabolism including RNase L activity (Figure 4.20). The construct with the intact 3DS structure pEGFPS seemed more resistant to this reduction.

CAV9 infection substantially reduced fluorescence for all constructs, indicating a large effect on host-cell processes (Figure 4.22). There was no productive effect of the 3DS element.

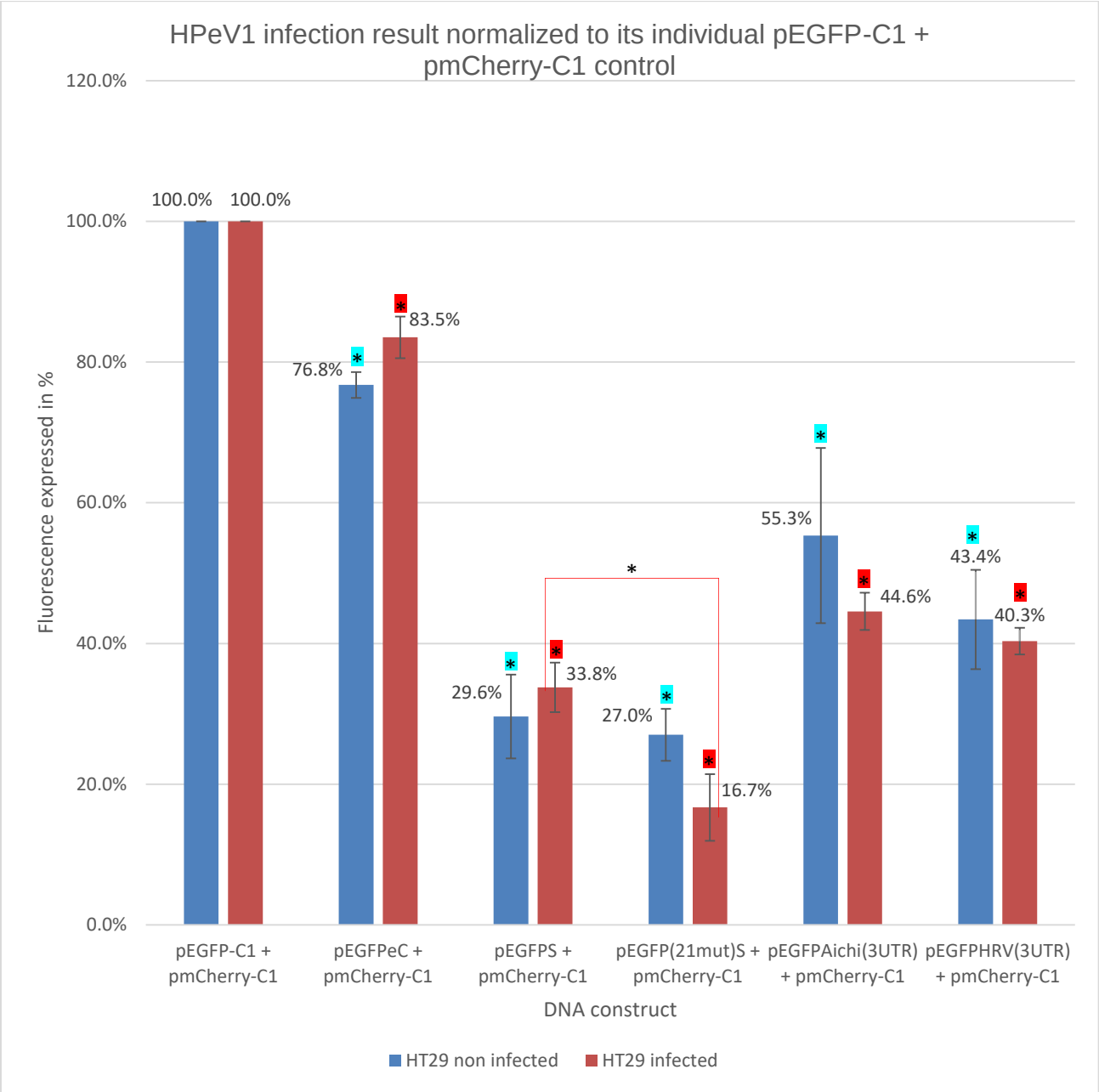


Figure 4.19: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. The result was normalized to (pEGFP-C1 + pmCherry-C1) of each independent experiment i.e. non-infected cells and infected cells. HT29 cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr infection with HPeV1. The blue bars show non-infected cells while the red bars show cells infected with 2.49 x 10^4 pfu/ml of HT29. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates p < 5.0E-03, * indicates p < 2.3E-02 and * indicates p = 4.5E-02.

* indicates significant difference of stressed cells due to infection with HPeV1

* indicates significant difference under no infection

* indicates significant difference when comparing pEGFPS and pEGFP(21mut)S.

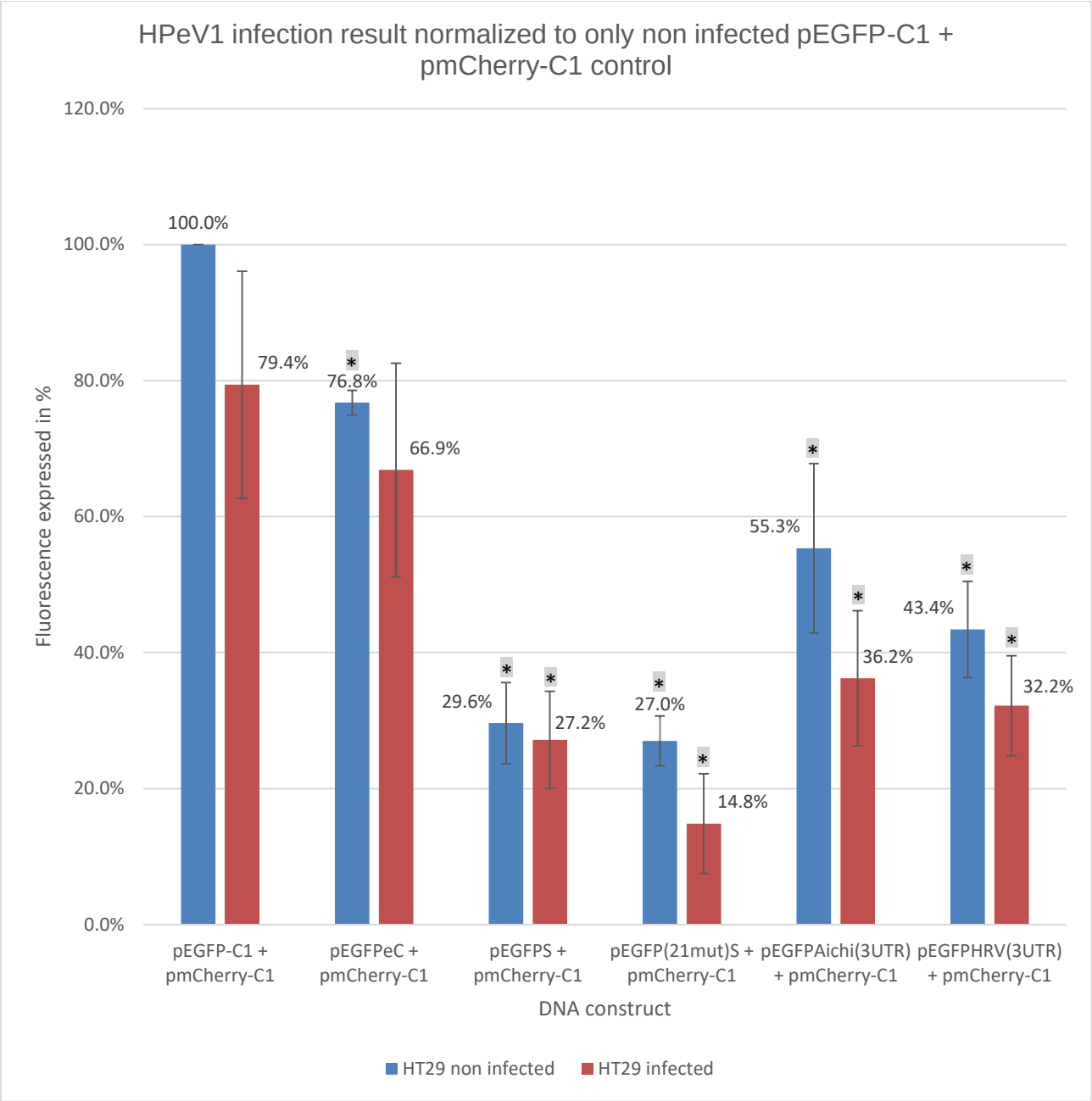


Figure 4.20: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. All results were normalized to only non-infected (pEGFP-C1 + pmCherry-C1). HT29 cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr infection with HPeV1. The blue bars show non-infected cells while the red bars show cells infected with 2.49 x 10^4 pfu/ml of HT29. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates $p < 2.3E-02$.

* indicates significant difference of all DNA constructs (infected and non-infected) to the non-infected control, pEGFP-C1.

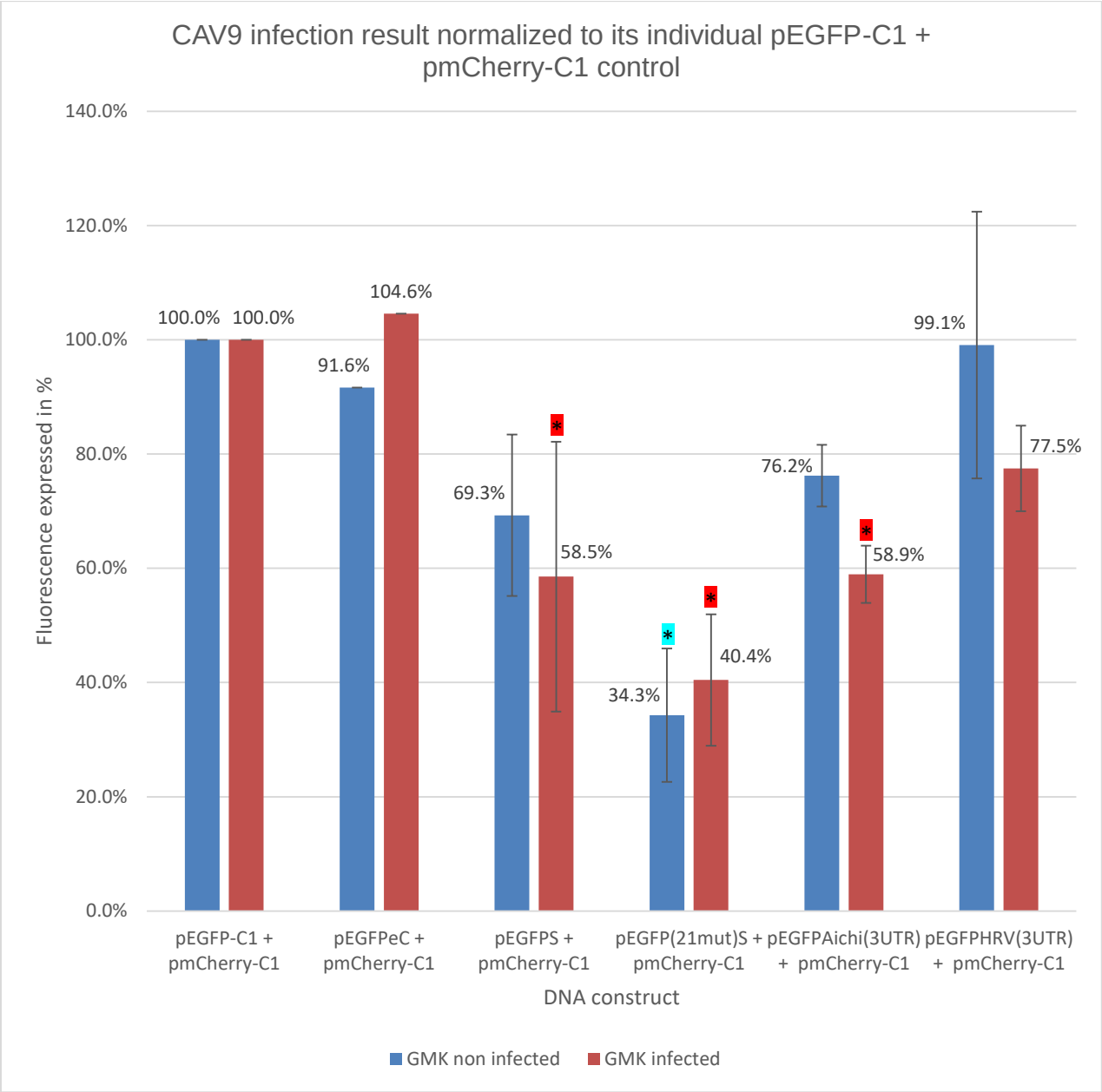


Figure 4.21: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. The result was normalized to (pEGFP-C1 + pmCherry-C1) of each independent experiment i.e. non-infected cells and infected cells. GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr infection with CAV9. The blue bars show non-infected cells while the red bars show cells infected with 1.26 x 10^5 pfu/ml of CAV9. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates p < 2.3E-02 and * indicates p = 2.6E-04

* indicates significant difference of stressed cells due to infection with CAV9

* indicates significant difference under no infection.

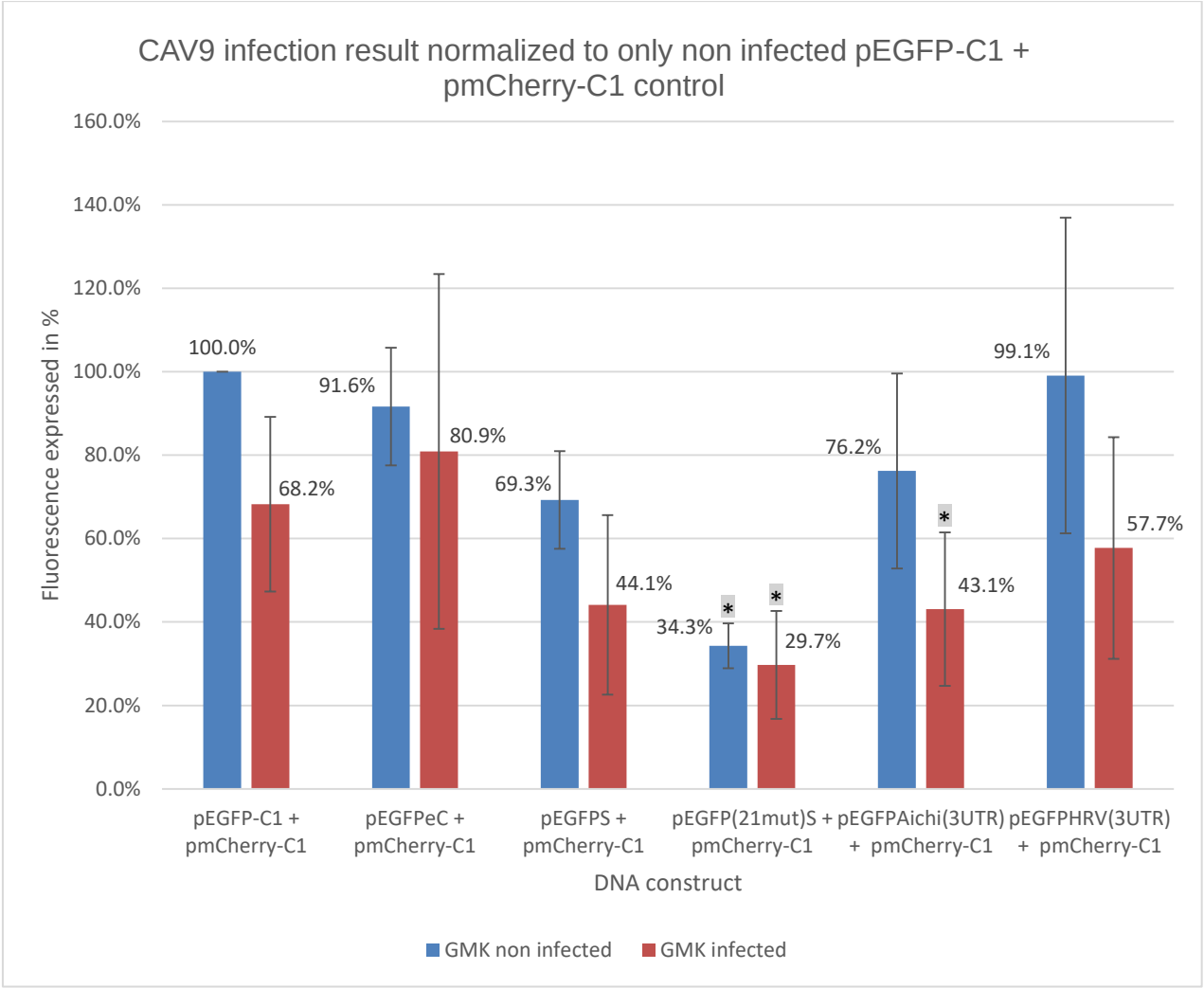


Figure 4.22: Mean fluorescence of EGFP constructs gated on the mCherry fluorescence. All results were normalized to only non-infected (pEGFP-C1 + pmCherry-C1). GMK cells were co-transfected with1.8 µg of EGFP construct and 3.7 µg pmCherry-C1 and analysed by flow cytometry using a FACS machine (BD Accuri™ C6 Cytometer) after 6 hr of transfection, 14 hr expression of the DNA and 6 hr infection with CAV9. The blue bars show non-infected cells while the red bars show cells infected with 1.26 x 10^5 pfu/ml of CAV9. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates $p < 3.6E-02$.

* indicates significant difference of all DNA constructs (infected and non-infected) to the non-infected control, pEGFP-C1.

4.5 Discussion

The results in chapter 3 indicate that there is a consistent reduction in signal in the EGFP fluorescence assay, when the 21 mutation construct was used, compared with the intact 3DS construct. To investigate if this difference can be seen more clearly and to investigate the conditions under which the structure could play a positive role, we decided to use a number of stress conditions.

Initially though we wished to find if the assay can be made more sensitive, by blocking de novo RNA synthesis in transfected cells to allow the effect of RNA decay to be more obvious. This approach is used in several similar studies e.g. Leppek *et al*., 2013 (Leppek *et al.*, 2013). These authors used an Actinomycin D concentration of 5 µg/ml and this concentration is said to be needed to block all three types of RNA transcription (Bensaude, 2011). A range of concentrations of Actinomycin-D; 5 µg/ml (Hewlett *et al.*, 1976; Welnowska *et al.*, 2011), 2 µg/ml (Krupina *et al*., 2000), 50 µg/ml (Herold & Andino, 2000) and 3 µg/ml (Romanova *et al.*, 2009) have been used over the years to carry out experiments to inhibit the transcription of RNAs. In this experiment, an optimized concentration of 0.01 µg/ml of Actinomycin-D was used because only 16.78% of the cells were lost. The next higher concentration, 0.1 µg/ml had 52.03% of cells lost while all the concentrations higher than 0.1 µg/ml also had a range of 48.29% - 95.06% loss of cells (Figure 4.1). The concentration of Actinomycin D used was therefore much lower than that often used for blocking transcription and may not have been effective. The lack of any effect was seen in Figure 4.3 where there was no reduction in fluorescence when the results are normalised to the untreated control. The toxicity of the treatment did not allow higher Actinomycin D concentrations to be used and it seems that GMK cells are

particularly sensitive to this drug. It may be worth trying to use other cell lines for these experiments, but those available in the laboratory tended to have poor transfection rates.

4.5.1 The effect of secondary structures on RNA stability in cells responding to stress

This chapter represents a novel way of analysing if the 3DS structure gives an advantage in different cell stress conditions. Viral infections cause severe stress to cells for instance they may induce ER stress (He *et al.*, 2006) and oxidative stress (Schwarz, 1996; Beck *et al.*, 2000). Each of the 3 conditions studied are well established stresses to which cells respond. In addition, oxidative stress (Dewey *et al.*, 2011), osmotic stress (Nilsson & Sunnerhagen, 2011) and ER stress (Walker *et al.*, 2013) are known to induce stress granules. These are cytoplasmic aggregates of RNA whose translation is not required. As this is an indication of major reorganization of RNA within the cell, RNA viruses could be vulnerable to the effects of these stresses and the 3DS structure could be part of a response to overcome them. It is also known that stress-inducing conditions, such as ER, oxidative and osmotic stress, can induce RNase L (Pandey *et al.*, 2004). This is the protein inhibited by the RNA structure within the 3C-encoding region of poliovirus and the HPeV 3DS structure could act in a similar manner (Han *et al.*, 2007).

5 DNA constructs (Table 4.2) and pEGFP-C1 + pmCherry-C1 were co-transfected individually to GMK cells and exposed to each stress condition/ infection for 6 hr each after expression of the transfected constructs for 14 hr prior to the examination with the FACS. Three independent experiments of each stress concentration were carried out. A low and high stress condition was chosen for each treatment.

For each experiment, the graphs show the relationship between the expression of EGFP and mCherry between stressed/infected and unstressed/non-infected cells. A set of graphs show two independent results of the expression of EGFP plus mCherry in each construct in unstressed cells normalized to the control, pEGFP-C1 + pmCherry-C1, while the expression of EGFP plus mCherry in the stressed cells is also normalized to pEGFP-C1 + pmCherry-C1 in the stressed cells only.

Neither of the osmotic stress conditions affected the overall fluorescence, while the higher levels of the oxidative and ER stress gave a substantially reduced fluorescence compared to the untreated control. The osmotic stress may not have been great enough and a higher concentration of NaCl could be used or a chemical such as sorbitol may have a different effect (Dewey *et al.*, 2011). The oxidative and ER stress condition appears to be having a significant effect on the cell.

In all the experiments there was a consistent difference between the 3DS construct and its 21 mutation version in the untreated control samples. The mutant gave around 60% of the fluorescence of the intact 3DS construct. This implies that the difference between these constructs is genuine and that the 3DS structure is having an effect. However, the two constructs did not behave as expected under the stress conditions. Under the low levels of oxidative and ER stress, the fluorescence from the two constructs tends to be changed in a similar way. At higher stress levels, however, the intact 3DS structure construct seems to be affected more severely. This can be seen for oxidative stress in Figures 4.10 and 4.12 and for ER stress in Figures 4.15 and 4.17. It might be expected that the 3DS structure could have a protective effect, but the results do not seem to support this. What could be happening is that the RNA stability/translation benefits from the presence of the structure, giving higher fluorescence in the untreated samples, but at

high stress conditions the benefit of the structure is lost. This adds further support to the idea that the structure is important.

The other test constructs, containing the Aichivirus 3'UTR stem-loop and the HRV 3'UTR, also did not show any significant difference from the controls in stressed cells.

4.5.2 **Infection with HPeV1**

In non-infected cells, the 21 mutations of the 3DS structure gave less fluorescence when compared with the 3DS structure, while in infected cells, there was a much greater loss of fluorescence in the 21 mutations of the 3DS structure when also compared to the 3DS structure. The transfection of the DNA constructs to the HT29 cells were generally low when compared with the transfection to GMK cells (Figure 4.19). This made the results more variable, but there is a clear difference between the intact 3DS structure construct and the 21 mutation construct. In the infected cells, there is a reduction in fluorescence compared to uninfected cells (Figure 4.20), suggesting that there is some interference with transcription/translation. It was previously reported that HPeV1 does not shut-off host cell protein synthesis, but this may depend on the cell line (Stanway *et al.*, 1994).

The results suggest that the 3DS structure does have a positive effect in cells infected with HPeV1, which could mean that it helps to overcome some of the host cell defences. In addition, infection with picornaviruses is known to affect the nucleus of the cell and can release several RNA binding proteins (Flather & Semler, 2015). At least one of these, Auf1, can increase turnover of picornaviruses RNA, leading to reduced virus replication. It is possible that the 3DS structure could help to increase HPeV RNA stability and combat this antivirus effect.

4.5.3 Infection with CAV9

The situation in cells infected with CAV9 is less clear. Overall, there is a more severe reduction in fluorescence in the infected cells than seen for HPeV infection, which is probably due to enteroviruses such as CAV9 being able to shut-off host cell protein synthesis. The protection given by the 3DS structure seen in HPeV infection did not occur. HPeV and CAV9 are diverse picornaviruses (Hyypiä *et al.*, 1992) and may induce different host cell effects in infection. Little is known about the intracellular events in HPeV infection, but it is clear that there are some differences from enteroviruses such as CAV9 in, for instance, replication complexes, which are derived from different cellular membranes (Gazina *et al.*, 2002). This could make a large difference to the environment in cells infected with CAV9 and HPeV. The 3DS structure may not give any advantage in CAV9 infection.

Conclusion

The many experiments performed confirmed that the 3SD construct does have a positive effect in unstressed cells. At higher levels of stress this effect seems to be lost, which was unexpected, but seemed to be a consistent observation and could be further evidence that the structure is significant. It also seemed to have a positive effect in HPeV-infected cells, compared to the mutant. As the structure needs to work in HPeV-infected cells, this result is an important indication that the structure does have a function.

Chapter 5

Assembly of a the *Renilla* Luciferase reporter construct to analyse RNA structures in HPeV1

5.1 Introduction

Reporter systems are important tools for finding the role of proteins or RNA structures in virus replication. Replicons can be produced that contain all the genome elements needed for the RNA replication without the need to generate infectious virus. This means that translation/RNA replication can be studied independently of events such as receptor binding, entry and assembly. Firefly and *Renilla* luciferases are popular reporters for replicons as they allow easily read outputs and quantitative results (Sherf *et al.*, 1996).

An HPeV1 replicon was previously used to analyse the role of the 5' terminal structures in virus replication. Two cis-acting signals (stem loops), SL-A and SL-B were observed to be involved in the replication of HPeV1. These two stem loops are located within the 5' terminal. It was seen that it was indeed the SL-A and SL-B stems that are vital for replication rather than their primary sequence, as RNA replication and growth of the virus was disrupted when there was a significant disruption of the structures and restoration of the growth properties when the structures were repaired. It was also shown that in vitro translation was not impaired in the mutants, hence suggesting that they may not be involved in translation but in RNA replication (Nateri *et al.*, 2002). However, the replicon was based on chloramphenicol acetyl transferase (CAT) as the reporter and this needs a cumbersome assay and is difficult to fully quantitate. It would be useful to have a replicon based on luciferase as the reporter, to analyse the structure described in Chapters 3 and 4 as well as other structures such as the cre. The HPeV cre has been predicted but there is no direct evidence that the prediction is correct (Al-sunaidi *et al.*, 2007). This Chapter describes the construction of a Renilla Luciferase replicon for HPeV and attempts to use it to study cre activity.

5.2 The generation of an HPeV1 replicon using a luciferase reporter

5.2.1 Backbone Preparation (pHPeV1Luc)

The cre is thought to be positon-independent (Goodfellow *et al.*, 2000). As it is assumed to be in the VP0-encoding region (Al-sunaidi *et al.*, 2007), it should be possible to remove all of the capsid region of HPeV1 and just add back the cre to give a functional replicon, because the capsid proteins are likely not to be involved in replication. The size of the pHPeV1 DNA could therefore be reduced, which may increase replicon efficiency as the probable efficiency of translation of the final RNA product in cells may be increased because shorter lengths of RNA would be translated easily than longer lengths of RNA. The same is true of replication of the RNA.

pHPeV1, a complete cDNA copy of the HPeV1 Harris genome (Nateri *et al.*, 2002) was used to make the replicon (pHPeV1Luc). The full pHPeV1 sequence was run through Webcutter2 (Heiman, 1997) to ascertain what restriction enzymes were absent in the DNA that could be introduced as unique sites for cloning and to find a unique site in the HPeV cDNA where pieces of DNA could be added. A unique AflII site was found in the VP1-encoding region. The map of the desired replicon construct is shown in Figure 5.1. The *Renilla* Luciferase gene is located downstream of the HPeV1 5'UTR and is followed by the HPeV cre. To allow this to be manipulated easily to make mutants, this was flanked by two unique restriction enzyme sites, NheI, which was artificially introduced, and the VP1 AflII site. The cre region was also surrounded by sequences encoding HPeV1 $3C^{pro}$ cleavage sites, so that the protein encoded by the cre region would cleaved from the Luc protein in case it interfered with Luc activity.

A total of 200 bp containing the cre region (51 bp), extra sequences before and after the cre (51 bp), the $3C^{pro}$ cleavage sites sandwiching the entire cre region (43 bp), and restriction enzymes (NheI and AflII) was designed. This was synthesized by Epoch (Appendix B). XhoI was a site to be introduced along with AflII, but unfortunately when a single digestion was carried out on pHPeV1, it was revealed that XhoI was already present in the DNA, hence, NheI was used instead because the DNA could not be cut by the enzyme (data not shown). Figure 5.2 shows uncut pHPeV1, and a single digestion of DNA with EcoRI, XhoI, MluI, AflII and SacI respectively. This was also done to test for the presence of a single MluI site (which is present after the poly(A) tail), a single AflII site (which is at position 1930 of the VP3 region of the virus) and also, a single SacI site (added just before the 5' UTR in the HPeV1 cDNA plasmid).

5.2.2 **PCR**

The HPeV1 5'UTR needed to be fused to the Luc gene and this was done by overlap PCR. The primers used for this experiment are listed in Table (2.4). 2 PCR products (SacI.5UTR.Luc and 5UTR.Luc.NheI, 728 bp and 954 bp respectively) were made with primers OL2259 (Forward) + OL2244 (Reverse) and OL2245 (Forward) + OL2260 (Reverse). These overlap at the junction between the 5'UTR and Luc gene. The fragments made are shown in Figure 5.3A. The PCR products were then joined in a second PCR reaction with primers OL2259 and OL2260 to give SacI.5UTR Luc.NheI (1682 bp) (Figure 5.3B). The PCR product was ligated into pGEMT-Easy and colonies with the correctly inserted DNA were identified by colony PCR. The plasmid DNA from one (named pGEM5UTR.LUC) was isolated and sequenced. The expected sequence was found with no mutations.

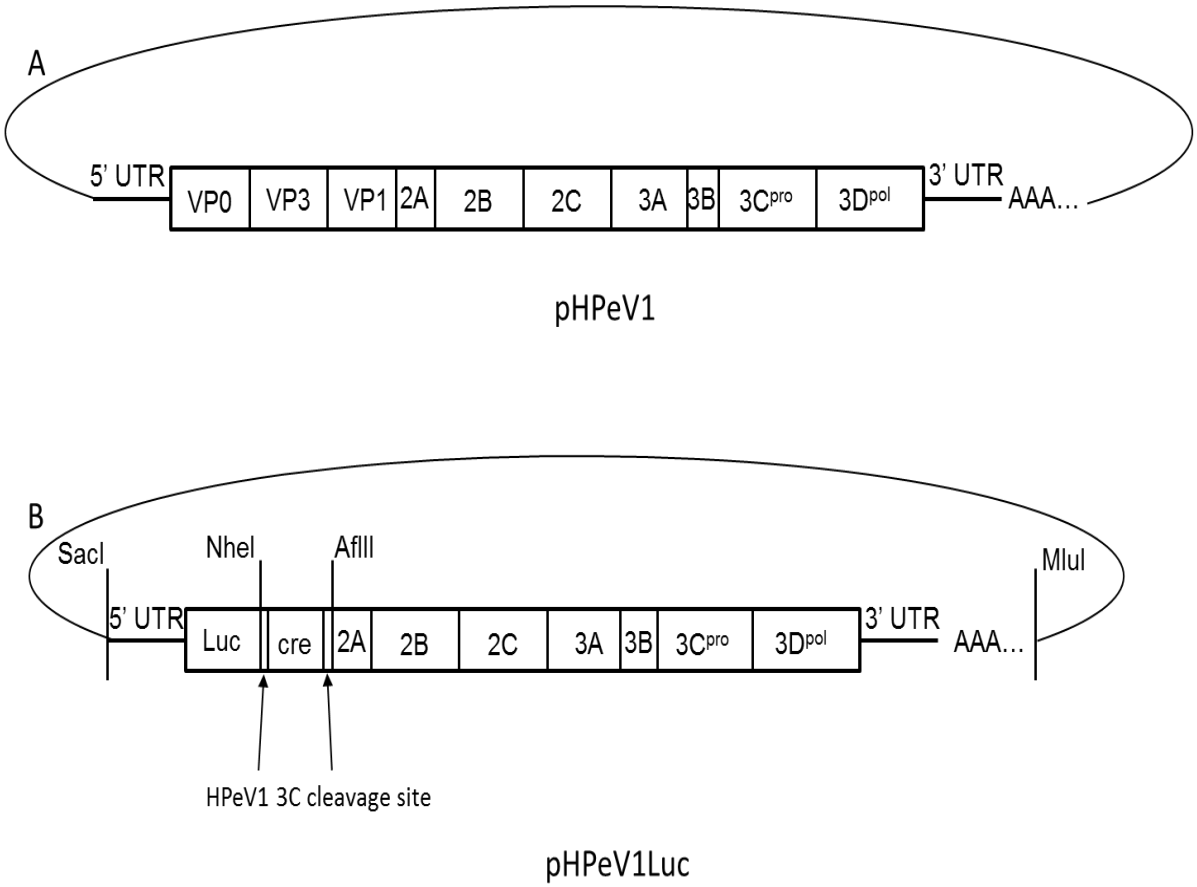


Figure 5.1: Schematic representation of pHPeV1 (A) and the desired replicon construct pHPeV1Luc (B). The restriction enzymes SacI, AflII and MluI were already present in the same sites the original pHPeV1 while NheI was built into pHPeV1Luc to aid the addition of mutated cre regions to pHPeV1Luc when cut with NheI and AflII.

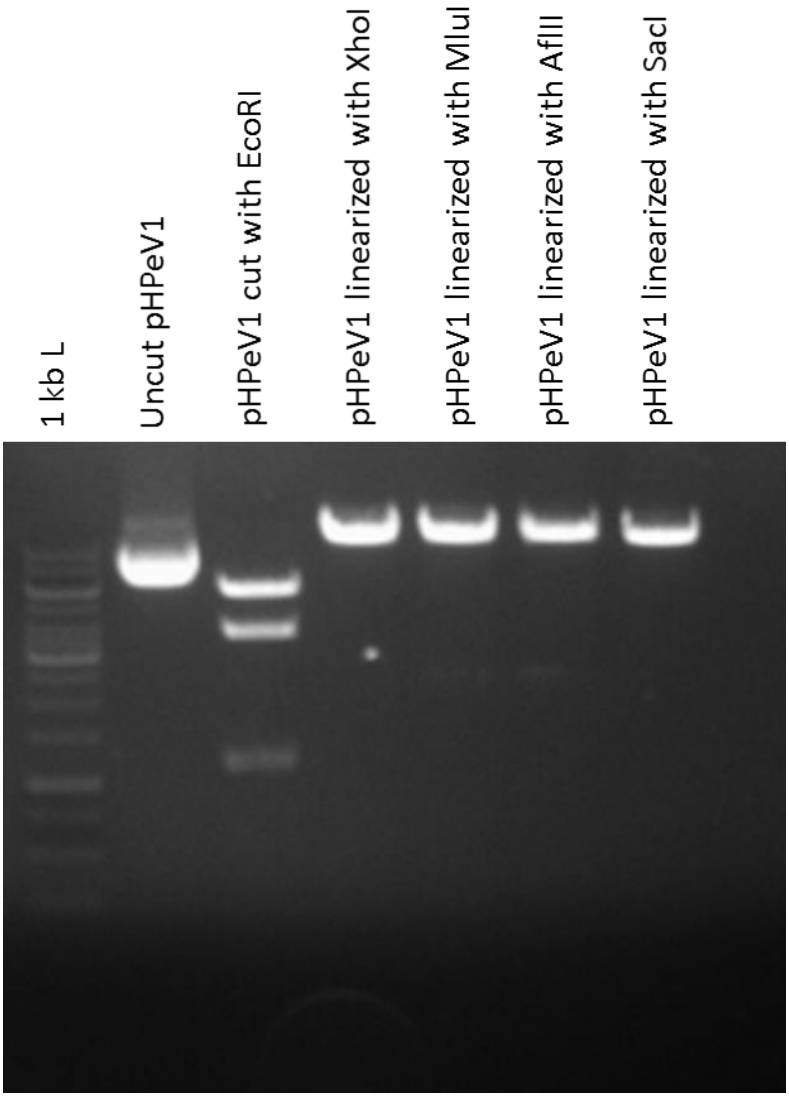


Figure 5.2: Uncut pHPeV1 and single digestion of pHPeV1 with EcoRI, XhoI, MluI, AflII and SacI.

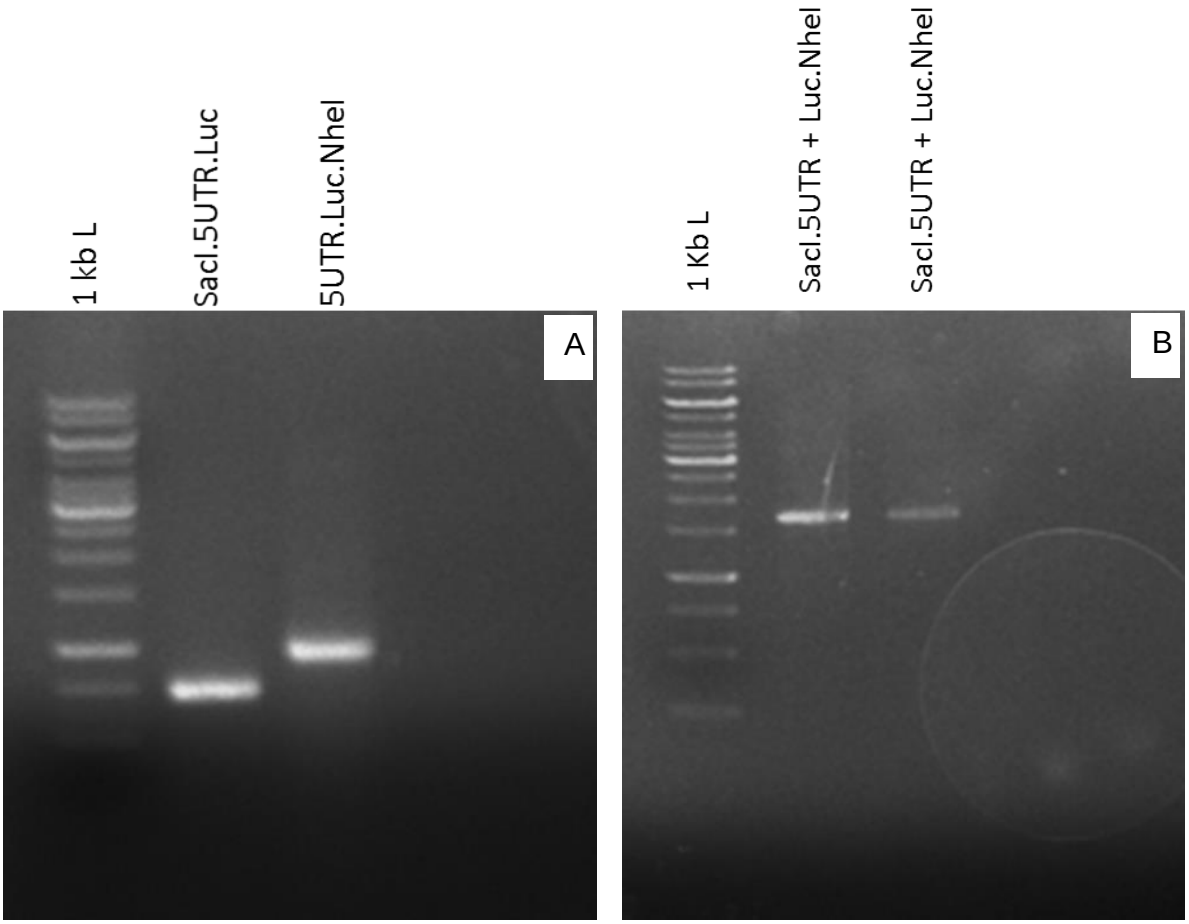


Figure 5.3: Gel picture showing A) two PCR products; SacI.5UTR.Luc and 5UTR.Luc.NheI respectively, purified and B) the joined PCR product (two separate samples); SacI.5UTR + Luc.NheI.

5.2.3 **Final ligation**

The sequence synthesized by Epoch (the clone pBSKcre) was also checked and found to be correct. Three clones, pHPeV1, pGEM5UTR.LUC and pBSKcre were then cut using the restriction enzymes shown in Table 5.1 to generate the fragments needed for a 3 fragment ligation. Ligation was then carried out in the ratio of 1:1:1 using T4 DNA ligase and transformed into *E. coli* cells. Random colonies were picked and then digested with restriction enzymes (data not shown) to confirm that the new construct pHPeV1Luc had been made. The region introduced into pHPeV1 was sequenced to confirm that there were no mutations present and that the junctions were correct. This data is shown in Appendix B.

Table 5.1: Plasmids used to give the fragments needed to make pHPeV1Luc and the restriction enzymes used to cut each.

DNA	Enzymes
pHPeV1	SacI and AflII
pGEM5UTR.LUC	SacI and NheI
pBSKCre	NheI and AflII

5.3 Use of the construct to study the putative HPeV cre

5.3.1 Production of pHPeV1LucLoopcreMut and pHPeV1LucRScreMut

The predicted HPeV cre structure is shown in Figure 5.4. It is a typical stem-loop as seen in other picornaviruses. However, the loop is relatively large and the mfold program has predicted further base-pairing in this loop. The loop contains a conserved CAAAC motif and the As of this motif are known to be critical for cre function in other viruses (Paul *et al.*, 2000). It was decided to mutate the AAA sequence to UUU and also to try to destabilize the stem, to demonstrate that the predicted structure is the cre. It was thought that mutations should be silent if possible and this limited how much the stem structure could be disrupted. It was not possible to mutate the AAA silently. Primers were designed which would be used to produce two different fragments, one containing 3 mutations at the loop of the cre “AAA” to “UUU” and the other fragment contained 2 mutations at the right stem of the cre “G” to “A” and “A” to “C” (Figure 5.4). Overlap PCR (Table 5.2) was carried out using the proof-reading enzyme, *pfu*

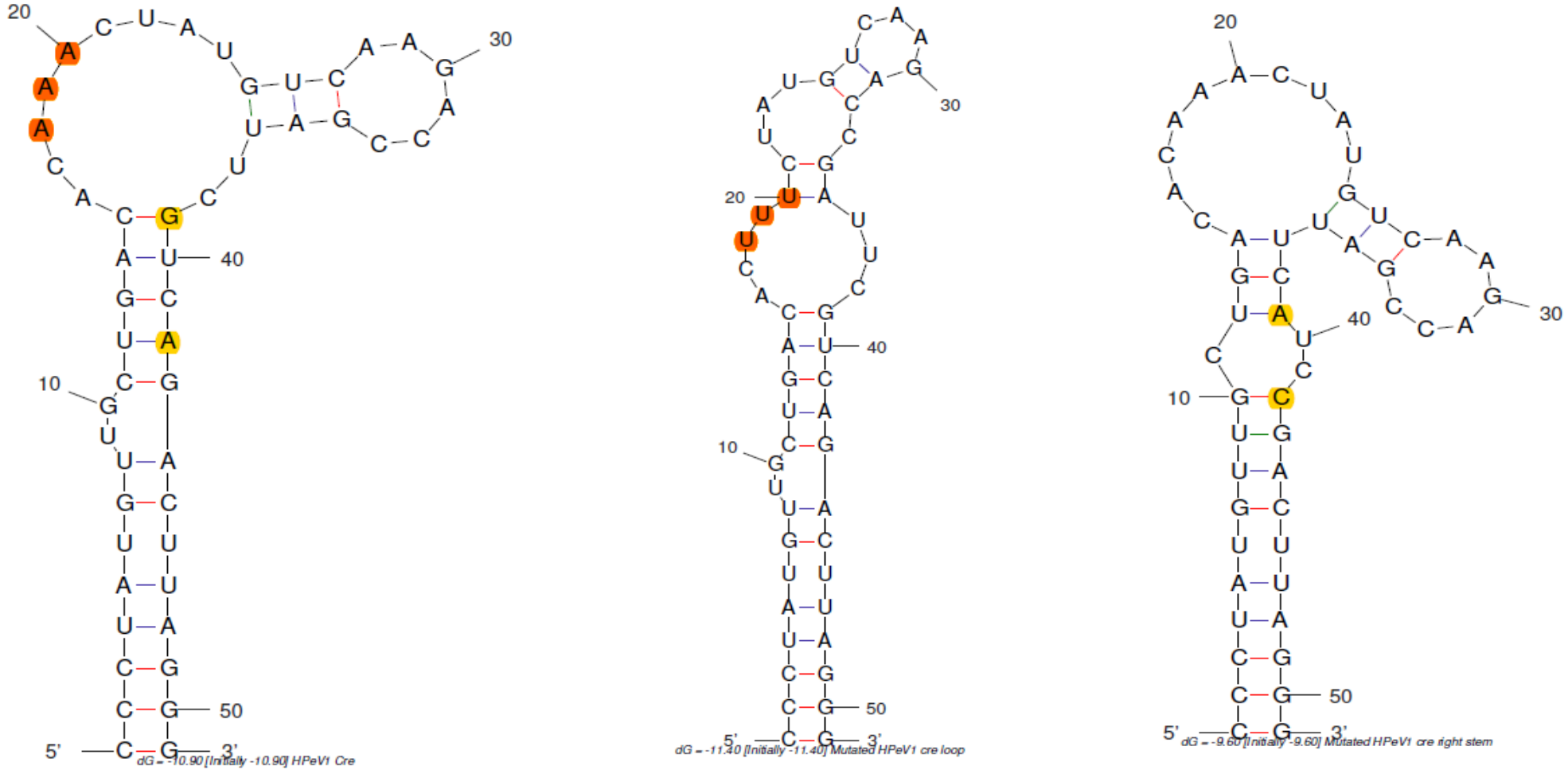


Figure 5.4: Schematic representation of A) the putative HPeV1 cre structure showing the target areas for mutation at the loop (red highlights) and right stem (yellow highlights), B) folding of the mutated loop with mutated bases highlighted in red and C) folding of the mutated right stem with mutated bases highlighted in yellow.

Table 5.2: Summary of how overlap PCR was carried out to produce pHPeV1Lucloop*cre*Mut and pHPeV1LucRS*cre*Mut.

Initial product	Template	Oligonucleotides	Final DNA construct
Fragment 1	pHPeV1Luc	GeneralFcre+ AngR	
Fragment 2	pHPeV1Luc	AngF + GeneralRcre	
	Fragment 1 + fragment 2	GeneralFcre + GeneralRcre	pHPeV1Lucloop*cre*Mut
Fragment 1	pHPeV1Luc	GeneralFcre + LouR	
Fragment 2	pHPeV1Luc	LouF+ GeneralRcre	
	Fragment 1 + fragment 2	GeneralFcre + GeneralRcre	pHPeV1LucRS*cre*Mut

polymerase, to amplify the fragments. The description of the primers used are shown in Table 2.4. The final PCR products were cloned into pGEMT-Easy and the sequences were confirmed. The cloned fragments were cut with NheI and AflII restriction enzymes. The final products were then ligated to pHPeV1Luc also cut with the same restriction enzymes.

Colonies were picked from each of the supposed new constructs, pHPeV1Lucloop*cre*Mut and pHPeV1LucRS*cre*Mut, and then analyzed using colony PCR to ascertain the presence of the mutations in the DNA. DNA constructs presumed to be correct after colony PCR were then sent for sequencing and the results confirmed that all mutations were present (data is shown in Appendix B).

5.3.2 **Analysis of the cre**

Figure 5.5 shows a schematic of the methodology to analyse the effect of mutating the putative HPeV cre in pHPeV1LucLoop*cre*Mut and pHPeV1LucRS*cre*Mut using *Renilla* Luciferase as the reporter gene. RNA was produced in vitro from the constructs and transfected into cells. If the replicon RNA can replicate, then the levels of the RNA in the cell will increase and this will be translated giving more reporter protein thus more luciferase activity. If the cre is not active (as predicted for pHPeV1LucLoop*cre*Mut) then the replicon RNA cannot replicate and so luciferase levels should remain low.

5.3.3 **Transcription**

In vitro transcription of the pHPeV1Luc, pHPeV1Lucloop*cre*Mut and pHPeV1LucRS*cre*Mut was carried out (Section 2.2.12) to give the corresponding RNA to be analysed via transfection in GMK cells. It can be seen that the RNA gives a distinct band, but there is also quite a lot of smaller RNA, possibly due to degradation (Figure

5.6). However, it was thought that the RNA quantity and quality were enough to continue.

5.3.4 **Analysis of the potential cre**

The RNA products of transcription from pHPeV1Luc, pHPeV1Lucloop*cre*Mut and pHPeV1LucRS*cre*Mut, together with pGL4.73[hRluc/SV40] (the *Renilla* Luciferase plasmid DNA) and pCDNA3, were transfected into GMK cells. Transfection was carried out using Lipofectin after 24 hr of transfection and 24 hr of expression of the DNA/RNA constructs. Figure 5.7 shows the results of the transfection from each construct analysed. There was no luminescence expressed by any of the constructs (except pGL4.73[hRluc/SV40]) analysed. This suggests that the RNA replicon is not functional and this needs to be further investigated.

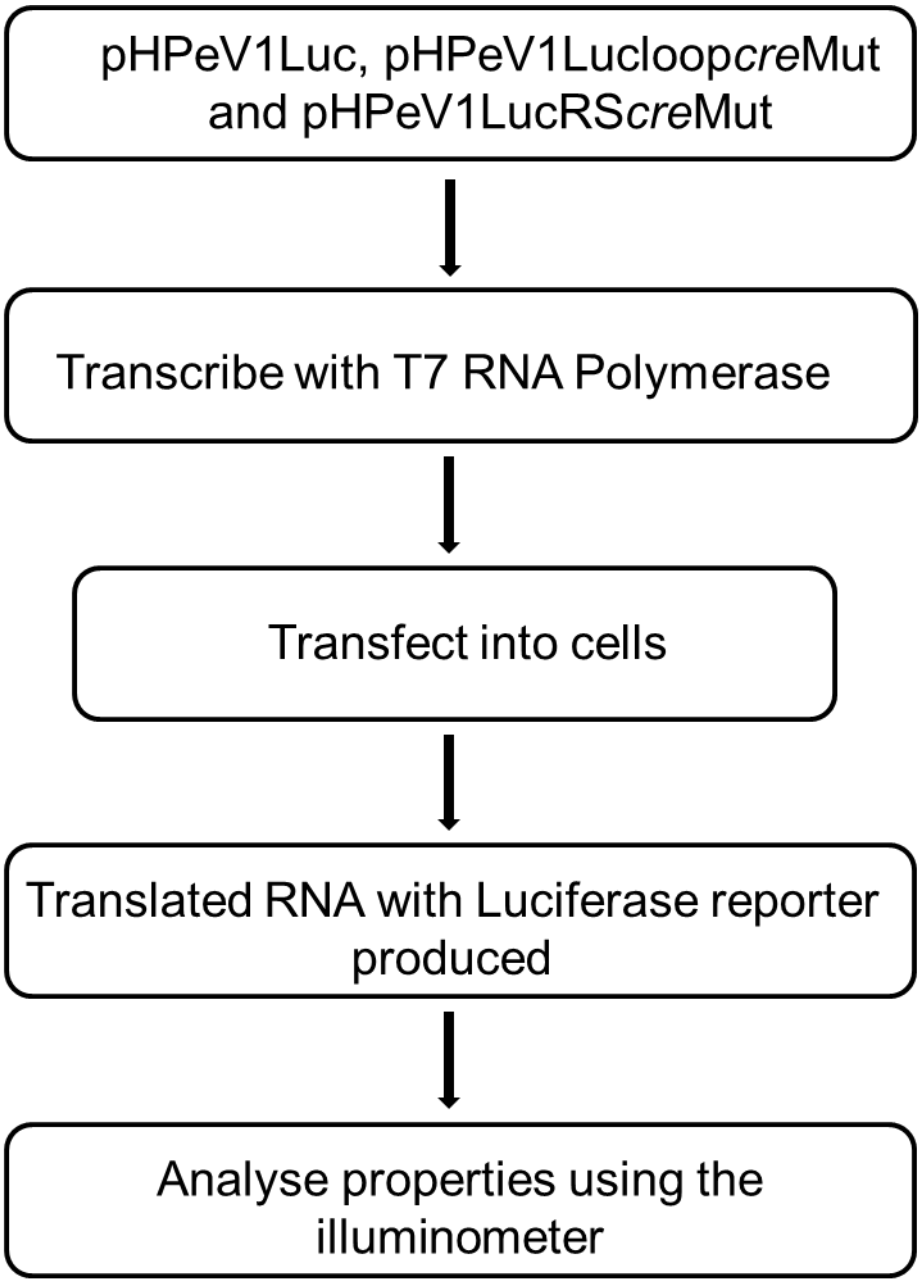


Figure 5.5: Schematic methodology diagram to illustrate how the HPeV replicon and its mutants were analysed using *Renilla* Luciferase as a reporter.

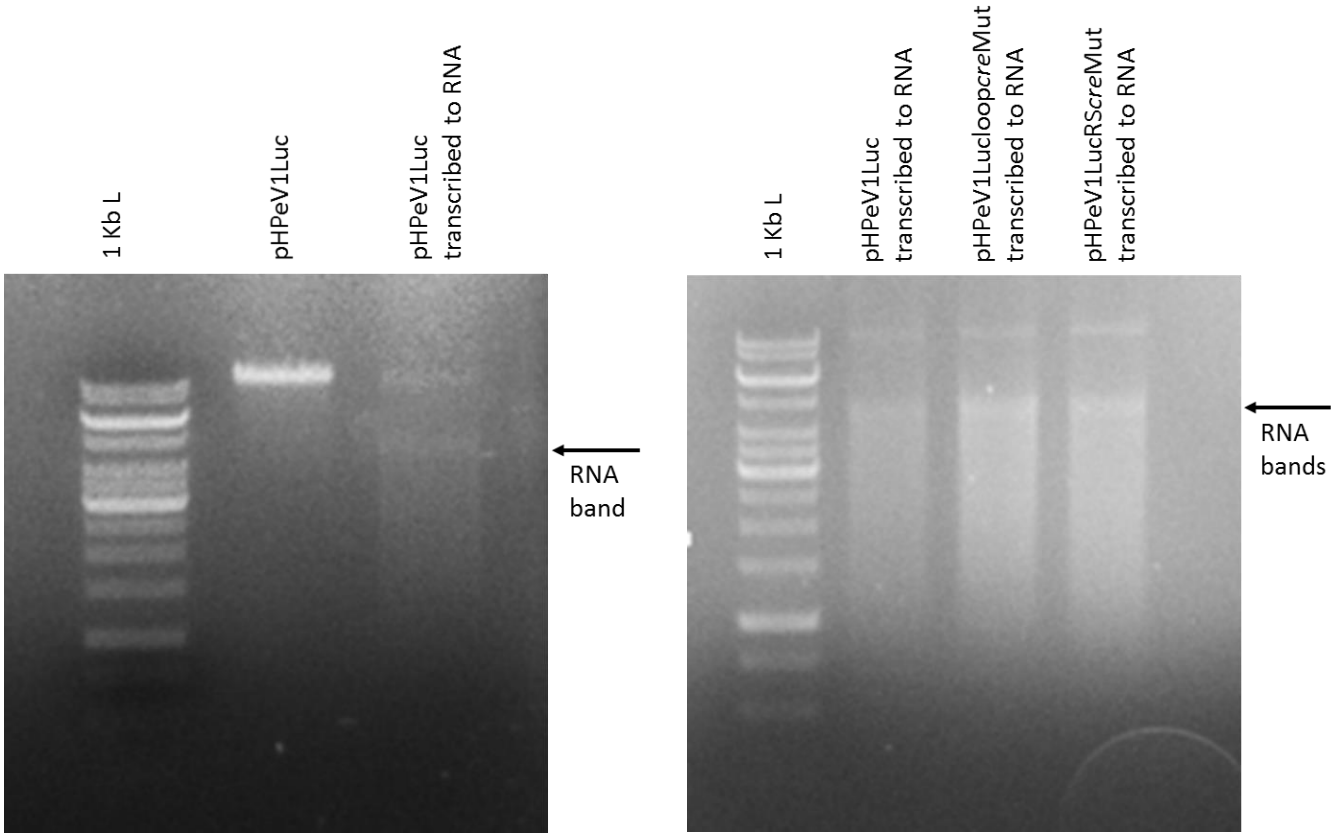


Figure 5.6: *In vitro* RNA transcription of the DNA samples; pHPeV1Luc, pHPeV1Lucloop*cre*Mut and pHPeV1LucRS*cre*Mut using T7 RNA polymerase. The RNA bands are shown by the black arrows.

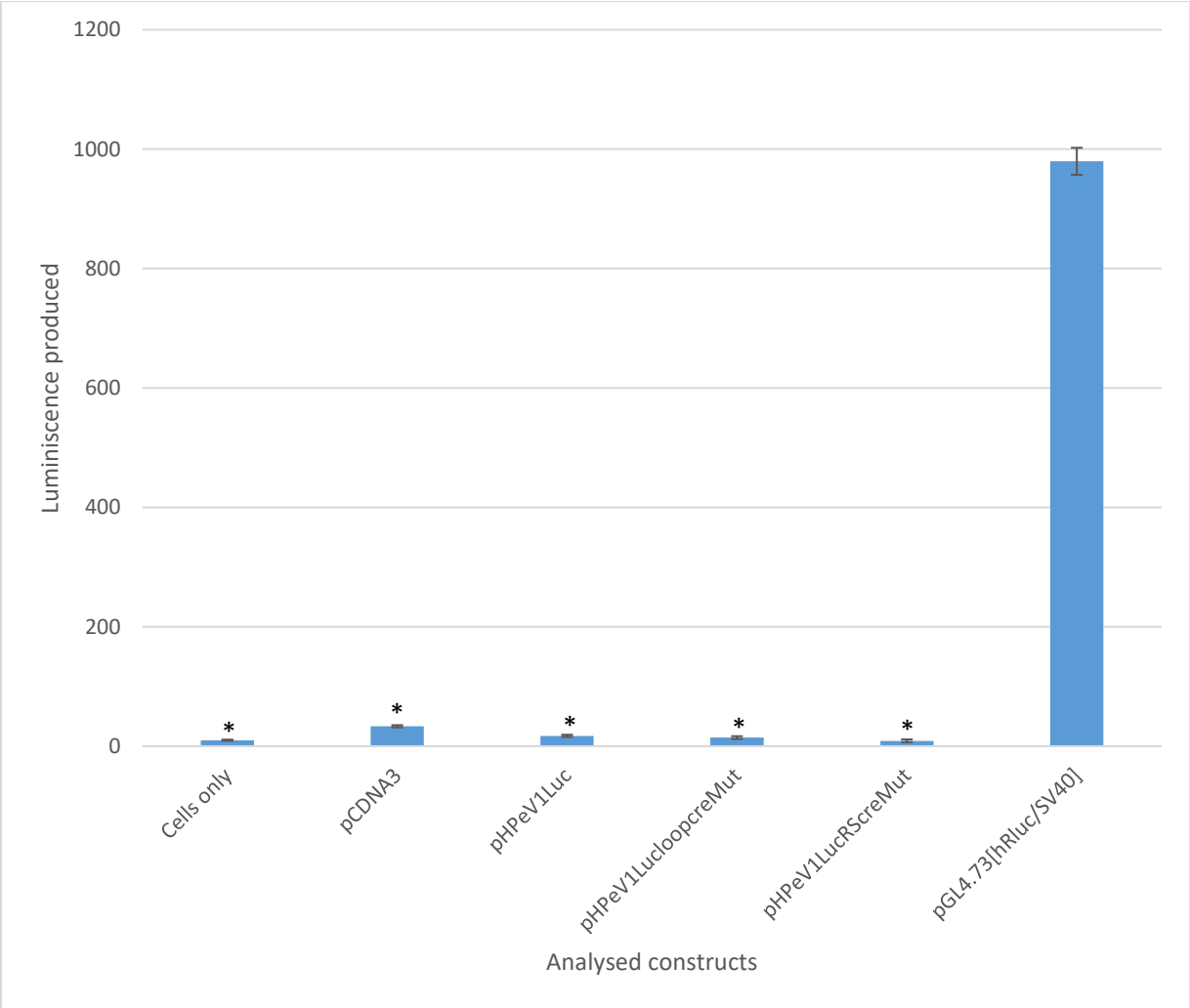


Figure 5.7: Mean luminescence of constructs analysed. GMK cells were transfected with1.8 µg of all constructs and analysed using an Illuminometer (BMG LABTECH) after 24 hr of transfection and 24 hr expression of the DNA/RNA. Error bars show the Standard Error (SE) from 3 independent experiments. * indicates $p < 2.0E-06$.

* indicates significant difference of all constructs to the control, pGL4.73[hRluc/SV40].

5.4 The 3DS RNA secondary structure in RNA replication

5.4.1 Production of pHPeV1(21mut)Luc

The potential role of 3DS RNA secondary structure involved in RNA replication can also be investigated using the pHPeV1Luc replicon. While the replicon was being produced, the 3DS 21 mutation EGFP construct described in chapter 3 was used to transfer the 21 mutations to pHPeV1 to give pHPeV1(21mut).

5.4.2 pHPeV1(21mut)

Three overlapping PCR fragments; WX, XY and YZ were produced from pHPeV1 or pEGFP(21mut)3DS. These spanned a region of pHPeV1 flanked by unique Mph1103I and MluI sites. The overlapping fragments were joined using overlap PCR to produce a single fragment containing the 21 mutations in the 3DS structure. This was carried out using some of the primers shown on Table 2.5. The purified DNA was ligated to pGEMT-Easy vector to give pGEM21mut and clones produced were confirmed by sequencing (Appendix C). Figure 5.8 gives a schematic representation of how each of the fragments was produced before being joined by overlap PCR. The reaction for the overlap PCR is shown in Table 5.3. pGEM21mut and pHPeV1 were cut with Mph1103I and MluI and the purified fragments were ligated to give the construct pHPeV1(21mut) (Figure 5.9A).

The final product, pHPeV1(21mut)Luc (Figure 5.9B), was to be made by transferring the AflII/MluI fragment from pHPeV(21mut) to pHPeV1Luc. However, pHPeV1(21mut)Luc could not be made because several attempts to ligate the fragments proved abortive. Also the experiment was paused because of the falied results from the RNA constructs; pHPeV1Luc, pHPeV1loop*cre*Mut and

pHPeV1RS*cre*Mut, analysed using the Luciferase reporter. The pHPeV(21mut) construct will itself be useful though, to analyse the effect of the mutations on the virus which could be recovered from this construct.

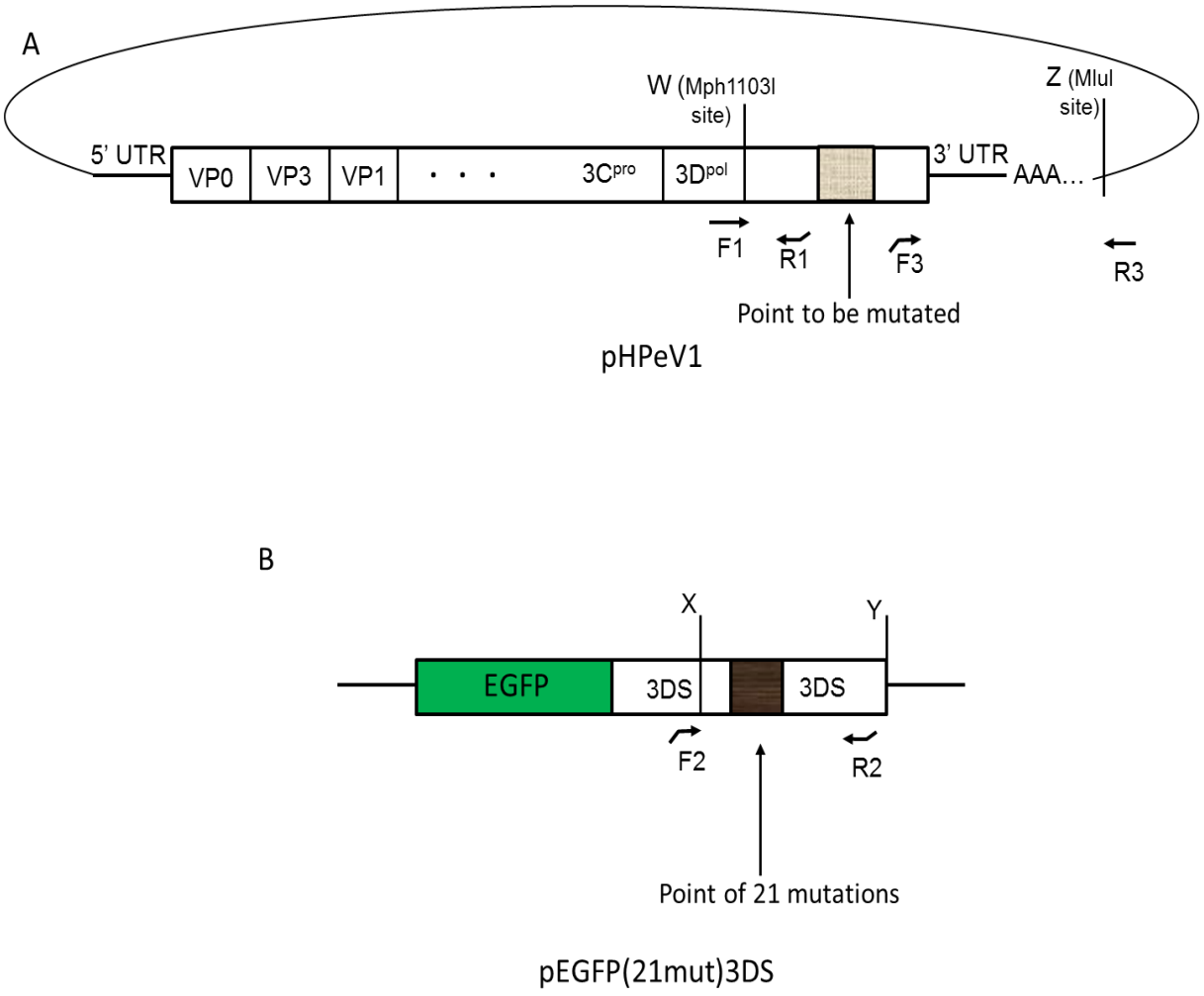


Figure 5.8: Schematic diagram showing A) primer target sites in the pHPeV1 cDNA which produces fragments 1 and 3 also known as fragments WX and YZ respectively, and B) primer target sites in the pEGFP(21mut)3DS construct (which contains the 21 mutations to be cloned to pHPeV1(21mut)) which produces fragment 2 also known as fragment XY. W and Z are the restriction enzyme sites; Mph1103I and MluI respectively which would be referred to in the later parts of the Chapter.

Table 5.3: Summary of how overlap PCR was carried out to produce the final product, pHPeV1(21mut)

Initial product	Template	Oligonucleotides	Final DNA construct
Fragment 1 (WX)	pHPeV1	OL2263 + OL2264	
Fragment 2 (XY)	pEGFP(21mut)3DS	OL2265 + OL2266	
Fragment 3 (YZ)	pHPeV1	OL2267 + OL2268	
	Fragment 1 + fragment 2 + Fragment 3	OL2263 + OL2268	pHPeV1(21mut)

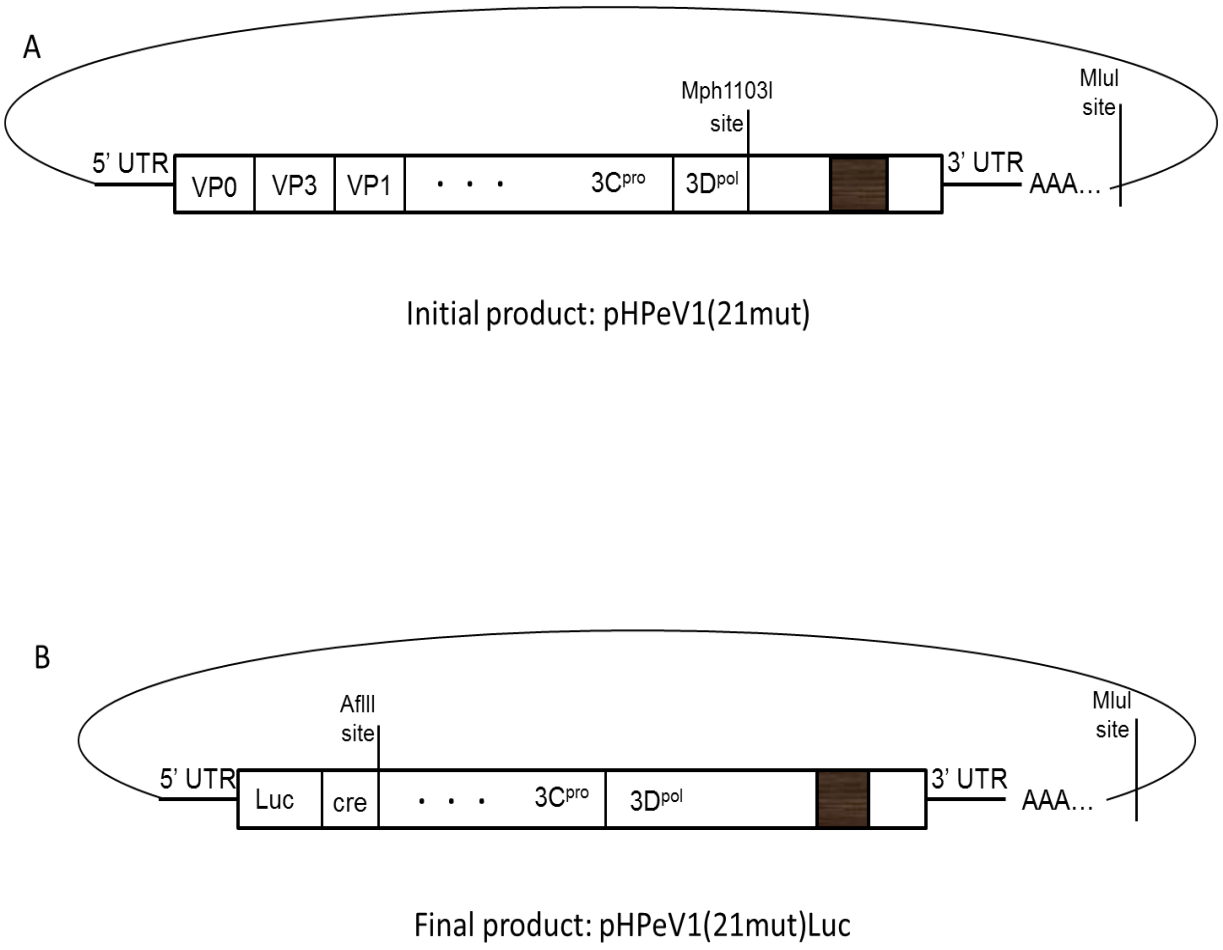


Figure 5.9: Schematic representation of A) the initial product; pHPeV1(21mut) made by adding the DNA fragment; pGEM21mut cut with Mph1103I and MluI to pHPeV1 cut with the same enzymes and B). The final product; pHPeV1(21mut)Luc, intended to be made by adding pHPeV1(21mut) cut with AflII and MluI to pHPeV1Luc cut with the same enzymes. The dark brown square shows the point of the 21 mutations.

5.4.3 **pHPeV1YGAA**

To control for the levels of luciferase produced by transfection of RNA without replication, a replication incompetent replicon is needed. The $3D^{pol}$ region of HPeV1 contains a YGDD motif active site found in all picornaviruses (Ghazi *et al.*, 1998; Stanway & Hyypiä, 1999). The YGDD motif was mutated to a YGAA motif (Appendix C), which should destroy polymerase activity.

Essentially the same process used to make pHPeV(21mut) was repeated. Overlap PCR was used to produce a PCR fragment containing the YGAA mutation in the $3D^{pol}$-encoding region of HPeV1 using primers shown on Table 2.5. The reaction for the overlap PCR is shown in Table 5.4. The purified DNA was ligated to pGEMT-Easy vector to give pGEMYGAA. Mph1103I and MluI were then used to introduce this into pHPeV1. Again the idea was to use AflII and MluI to transfer the mutated region into pHPeV1Luc (Figure 5.10), but this was not successful and was not repeated until the replicon could be shown to be active.

Table 5.4: Summary of how overlap PCR was carried out to produce the final product, pHPeV1(YGAA)

Initial product	Template	Oligonucleotides	Final DNA construct
Fragment 1	pHPeV1	OL2263 + OL2270	
Fragment 2	pHPeV1	OL2269 + OL2268	
	Fragment 1 + fragment 2	OL2263 + OL2268	pHPeV1(YGAA)

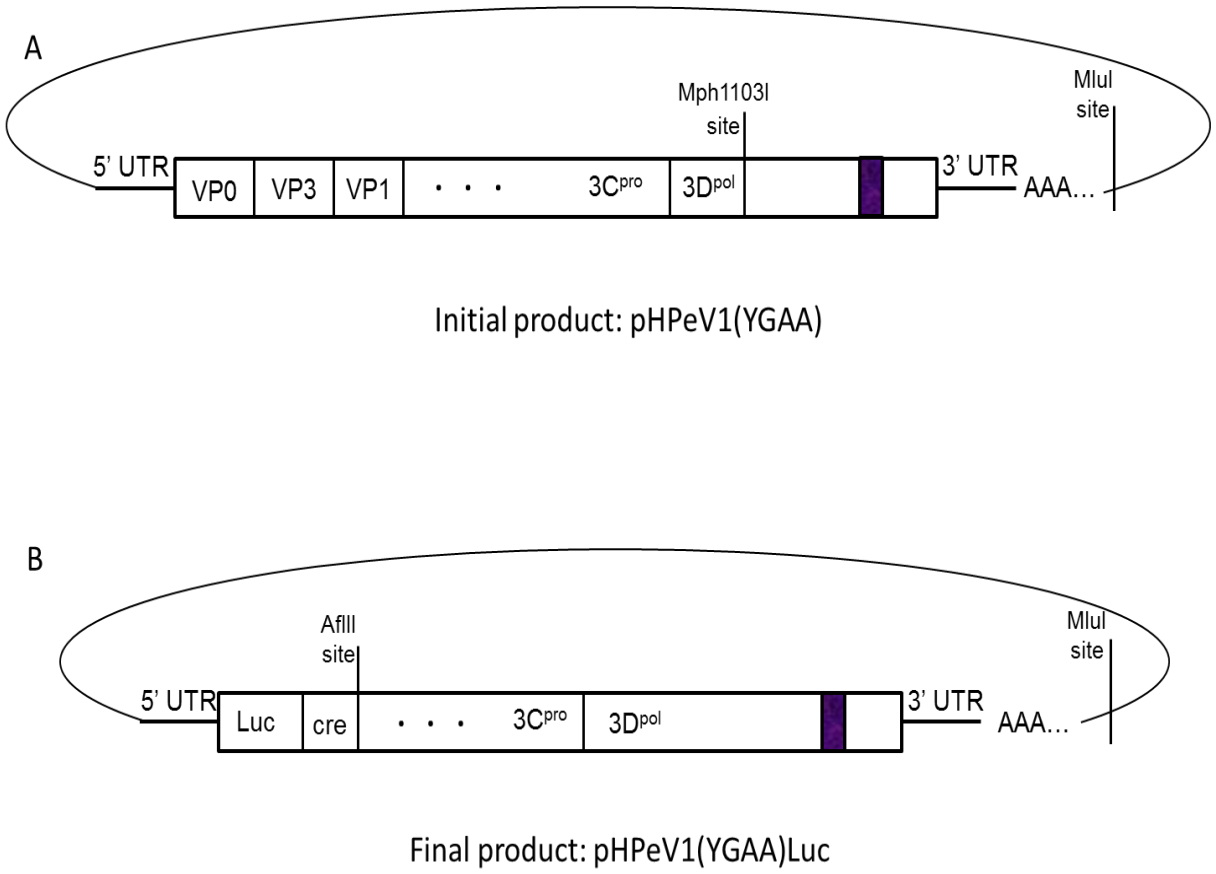


Figure 5.10: Schematic representation of A) the initial product; pHPeV1(YGAA) made by adding the DNA fragment; pGEMYGAA cut with Mph1103I and MluI to pHPeV1 cut with the same enzymes and B). The intended final product; pHPeV1(YGAA)Luc made by adding pHPeV1(YGAA) cut with AflII and MluI to pHPeV1Luc cut with the same enzymes. The dark blue rectangle shows the point of the YGAA mutation.

5.5 Discussion

The HPeVs are common human pathogens with 95% of people becoming seropositive between the ages of two to five years (Stanway *et al.*, 2000; Harvala *et al.*, 2010), however recent studies have linked HPeV1, HPeV3 and HPeV4 to serious cases of neonatal sepsis and meningitis (Harvala *et al.*, 2008; Wildenbeest *et al.*, 2010; Kolehmainen *et al.*, 2014b). The large number of HPeV types (16) means that vaccines are not practical and we need to understand more about how these viruses replicate to provide information which could be useful for producing antivirus drug. The HPeV cre is very important in replication of the virus because the AA dinucleotide in the loop serves as a template for UU synthesis which binds to VPg and works as a primer for negative sense RNA synthesis (Pathak *et al.*, 2007). The cre of HPeV has been predicted to be located in the VP0-encoding region (Al-sunaidi *et al.*, 2007) and it was hoped to test this prediction via mutational analysis.

To test cre function and also open up the possibility of analyzing the 3DS structure in terms of the effect on RNA replication, an HPeV luciferase replicon pHPeV1Luc) was constructed (Figure 5.1). This introduced the *Renilla* luciferase coding sequence downstream of the HPeV1 5'UTR. A region containing the predicted cre, surrounded by sites for NheI and AflII was commercially synthesized and introduced downstream of the luciferase gene. PCR was used to link the 5'UTR/luciferase/cre regions together and then this section of DNA was used to replace the corresponding region in pHPeV1 from the 5'UTR to a unique AflII site in the VP1-encoding region. This gives a construct with all the HPeV1 non-structural region present and this should provide everything needed for replication, provided the cre is active. PCR to make the construct and also for mutagenesis of the cre was performed using *pfu* polymerase, because it is a

proofreading enzyme (Lawyer *et al.*, 1993b; Cline *et al.*, 1996) and after the manipulations, the sequences of the fragments were exactly as expected. It was easy to mutate the cre region in the new construct, pHPeV1Luc, because it was designed to be part of a 200 bp region between two unique sites. This should also be useful for further studies on the cre.

It would have been interesting to find out the level of luminescence produced by the wild type cre and also each of the mutants. After transfection, there was no luminescence produced by the RNA constructs, suggesting that there was no translation and/or replication of the constructs in the cells. pGL4.73[hRluc/SV40], a *Renilla* Luciferase DNA construct, was used as a control to test that the assay system was functional and as expected, this gave a high luminescence value. To trouble-shoot for the reason(s) why the RNA constructs were not translated and/or replicated in cells, different approaches could be taken and they include; transcription of pGL4.73[hRluc/SV40] to RNA to check that the RNA transfection to cells is working adequately. Although the lipofectin method has been used for transfection many times in the laboratory for RNA, other transfection reagents could be used to enhance the transfection efficiency. Also, the 5UTR.Luc DNA fragment could be cloned into a new vector, to check that the 5' UTR indeed drives the translation of the Luciferase in the pHPeV1Luc construct. The failure to see a luciferase signal may also be due to technical problems. The RNA quality was not high and this should be improved by optimizing the transcription reaction.

There are a number of other reasons why the replicon may not work. One of these is that the experiment was performed in GMK cells, but HPeV does not replicate efficiently in these cells (data not shown). They were chosen as they give good transfection rates

(for DNA) but the poor growth could be due to an intracellular step in replication, translation or RNA replication, and this could affect the observed luciferase activity. Cells such as HT29, where HPeV1 replicates more efficiently can be used, although the lower transfection rates could be a problem.

In the case of an FMDV replicon based on GFP as the reporter, it was found that the virus protease L^{pro} caused cleavage of the GFP (Tulloch *et al.*, 2014b) and removal of L^{pro} gave a much higher level of the signal. It is possible that the HPeV $3C^{pro}$ could cleave the luciferase produced by our replicon and this could be studied by Western Blotting. As $3C^{pro}$ is essential for virus replication and cannot be removed from the HPeV replicon, if $3C^{pro}$ does cleave luciferase, an alternative reporter, such as EGFP, would need to be investigated.

The experiment relies on the cre prediction being correct (Al-sunaidi *et al.*, 2007). This is in a well-conserved region and there is good covariance data to suggest it is the cre, but if the actual cre is in the part of the capsid-encoding region deleted in pHPeV1Luc, then the construct would not work. The previously used CAT replicon had a deletion of about 900 nucleotides less that pHPeV1Luc and potentially this region could contain the real cre or have some function needed for RNA replication (Nateri *et al.*, 2002). This could be investigated by making exactly the same replicon, but replacing CAT with luciferase.

It was also decided to test the effect of mutation of the 3DS structure in the replicon. The final transfer of the mutant fragment into pHPeV1Luc was not successful. The entire size of pHPeV1Luc was about 10000 bp and making the construct needed the ligation of two large pieces of DNA which could be a possible reason why the ligation was very difficult and could not be achieved. It is easier to ligate smaller fragments of

DNA than larger ones. Attempts to complete a YGAA mutant which should be replication negative, to act as a control in the luciferase assay, was also unsuccessful. However, the 3DS mutant was placed into pHPeV1 and it should be possible to recover virus from this construct for further experiments to understand the significance of the structure.

Chapter 6

General discussion and future work

6.1 General discussion and future work

6.1.1 General discussion

The work in this thesis has helped to broaden the knowledge of the 3DS structure in the $3D^{pol}$-encoding region of Human Parechovirus (HPeV) (Williams *et al.*, 2009). The experiments involved mainly mutational analysis as well as useful bioinformatics tools to analyse sequences and validate experimental works.

Chapter 3 contained results of extensive analysis which also involved mutations of the 3DS structure in the HPeV1 genome. The analysis began with an alignment of the 3DS structure in all strains of HPeV sequenced in this region to refine the previously predicted structure (Williams *et al.*, 2009). This was then followed by the investigation of the role of the 3DS structure via microscopy.

A novel experiment was conducted to study the possible effect of the 3DS structure on translation in a bicistronic construct with dsRed as the reporter of the IRES-driven translation. The first cistron was expressed when the cells were successfully transfected while the second cistron was expressed only when the IRES in the 5' UTR functions to drive the translation. If the 3DS structure enhances IRES-driven translation, then the dsRed signal should be increased. The results were inconclusive because of the poor transfection efficiency of the dsRed constructs produced.

The investigation of the 3DS structure was further investigated by ligation into pEGFP-C1 between XhoI and BamHI sites and using pmCherry-C1 for co-transfection. Several other

constructs (made in pEGFP-C1 between XhoI and BamHI sites) were also investigated as controls. The constructs included those from different virus origins, the mutation of two conserved 'CG' dinucleotides (Williams *et al.*, 2009) in the $3D^{pol}$-encing region of HPeV1, 21 mutations of the conserved $3D^{pol}$ RNA structure as well as different sizes of the DNA constructs. The results from the analysis with the fluorescent microscopy for all the constructs were validated by using flow cytometry.

Chapter 4 investigated the stability of the 3DS structure when compared with the 21 mutations in the 3DS structure (which completely destroys the structure), under different stress conditions. Transfected GMK cells were stressed 6 hr prior to analysis with the FACS. The induced stress was carried out separately using such as NaCl for osmotic stress (Brocker *et al.*, 2010), H_2O_2 for oxidative stress (Halliwell *et al.*, 2000), Thapsigargin for ER stress (Ruan *et al.*, 2007) as well as viral stress on the cells being transfected using HPeV1 and CAV9 on HT29 and GMK cells respectively. Different concentrations of each stress condition were tested to select a relatively low and relatively high concentration that would be used to carry out the experiment. The relatively low stress concentrations selected were those in which the cells were relatively healthy when compared with the unstressed cells while the relatively high stress concentrations selected were those in which the cells were relatively unhealthy when compared with the unstressed cells. The respective relatively low and high stress concentration for each material used for the experiment include: 140 mM and 400 mM of NaCl, 50 µM and 500 µM of H_2O_2 and 0.3 µM and 1 µM of Thapsigargin.

The results show that in unstressed cells and under all stress conditions, the 3DS structure allows a higher level of fluorescence than the mutant, suggesting an effect on RNA stability/translation. Some of this effect was lost at high stress levels, and the fluorescence was reduced closer to the mutant value, which suggests that the benefit of the structure was overcome by these conditions.

Chapter 5 focused on investigating the HPeV1 putative *cre* structure. A replicon construct based on Renilla luciferase was made. Analysis of RNA transcribed from the replicon showed that no luminescence was produced. This suggests that the RNA was either not translated and/or replicated in the cells. Further studies of the 21 mutations in the 3DS structure also proved abortive due to difficulties in ligation to the final Luciferase replicon to be used for the investigation. Different approaches for trouble-shooting on the reason(s) behind the failed experiments have already been discussed in the chapter.

The work provided in this thesis have shown that the 3DS structure is indeed important in the biology of HPeVs. The results therefore gives a basis for a better understanding of how HPeVs interacts with the cells and this information could be used for the development of vaccines or anti-viral therapy in the future.

6.1.2 **Future work**

Now that an HPeV construct containing the mutant 3DS structure has been made, it should be possible to generate mutant virus, as long as the mutations are non-lethal. It would be interesting to perform competition experiments with the wild type virus and the virus with 21 mutations in the 3DS structure, to see if the structure gives an advantage to the virus. This will be done by infecting cells with the viruses and then sequencing the mixture to determine the prevailing strain(s). From time to time, a number of passages and sequencing will be carried out to conclude on which strain(s) mostly dominated the cells. Other measures of growth, plaque size and growth curves, may also be useful.

The EGFP assay is novel, convenient and seemed to give clear results. It should be validated though using other methods, such as Northern blotting, to analyse RNA levels (Leppek *et al.*, 2013). It can also be validated by introducing different sequences/structures into the constructs that promote RNA stability or instability, to ensure that the expected effect on the levels of EGFP fluorescence is observed.

It would also be useful to identify if the structure works by binding to cell or virus proteins. These could be identified by mixing RNA with extracts of infected cells and then carrying out mass spectrometry. This should give clearer insight into the role of the RNA structure in the virus life-cycle.

The work described in this thesis provides novel information on HPeV biology. A further understanding using the approaches and materials generated in this work could provide the resources to generate antivirus agents which can treat infections with these important pathogens.

Chapter 7

References

Al-sunaidi M, Williams ÇH, Hughes PJ, Schnurr DP, Stanway G (2007) Analysis of a New Human Parechovirus Allows the Definition of Parechovirus Types and the Identification of RNA Structural Domains. *Journal of virology*, **81**, 1013–1021.

Atkinson NJ, Witteveldt J, Evans DJ, Simmonds P (2014) The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic acids research*, **42**, 4527–45.

Baltimore D (1971) Expression of animal virus genomes. *Bacteriological reviews*, **35**, 235–241.

Banerjee R, Dasgupta A (2001) Interaction of picornavirus 2C polypeptide with the viral negative-strand RNA. *The Journal of general virology*, **82**, 2621–2627.

Bannwarth S, Gatignol A (2005) HIV-1 TAR RNA: The Target of Molecular Interactions Between the Virus and its Host. *Current HIV Research*, **3**, 61–71.

Barco A, Carrasco L (1995) A human virus protein, poliovirus protein 2BC, induces membrane proliferation and blocks the exocytic pathway in the yeast Saccharomyces cerevisiae. *the The European Molecular Biology Organization Journal*, **14**, 3349–3364.

Barr JN, Fearns R (2010) How RNA viruses maintain their genome integrity. *The Journal of general virology*, **91**, 1373–87.

Beck MA, Handy J, Levander OA (2000) The role of oxidative stress in viral infections. *Annals of the New York Academy of Sciences*, **917**, 906–12.

Bedard KM, Semler BL (2004) Regulation of picornavirus gene expression. *Microbes and Infection*, **6**, 702–713.

Bedard KM, Walter BL, Semler BL (2004) Multimerization of poly (rC) binding protein 2 is required for translation initiation mediated by a viral IRES. *RNA (New York, N.Y.)*, **10**, 1266–1276.

Belsham GJ (2009) Divergent picornavirus IRES elements. *Virus Research*, **139**, 183–192.

Benjamin D, Colombi M, Stoecklin G et al. (2006) A GFP-based assay for monitoring post-transcriptional regulation of ARE-mRNA turnover. *Molecular BioSystems*, **2**, 561.

Bensaude O (2011) Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription*, **2**, 103–108.

Benschop KSM, Williams CH, Wolthers KC, Stanway G, Simmonds P (2008) Widespread recombination within human parechoviruses: analysis of temporal dynamics and constraints. *The Journal of general virology*, **89**, 1030–1035.

Bergamini G, Preiss T, Hentze MW (2000) Picornavirus IRESes and the poly(A) tail jointly promote cap-independent translation in a mammalian cell-free system. *RNA (New York, N.Y.)*, **6**, 1781–1790.

Bird L (1997) Characterization and crystallization of the helicase domain of bacteriophage T7 gene 4 protein. *Nucleic Acids Research*, **25**, 2620–2626.

Bochkov YA, Watters K, Ashraf S et al. (2015) Cadherin-related family member 3, a childhood asthma susceptibility gene product, mediates rhinovirus C binding and replication. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 5485–90.

Boros Á, Nemes C, Pankovics P, Kapusinszky B, Delwart E, Reuter G (2012) Identification and complete genome characterization of a novel picornavirus in turkey (Meleagris gallopavo). *The Journal of general virology*, **93**, 2171–2182.

Boros Á, Nemes C, Pankovics P, Kapusinszky B, Delwart E, Reuter G (2013) Genetic characterization of a novel picornavirus in turkeys (Meleagris gallopavo) distinct from turkey galliviruses and megriviruses and distantly related to the members of the genus Avihepatovirus. *Journal of General Virology*, **94**, 1496–1509.

Bracht AJ, O'Hearn ES, Fabian AW, Barrette RW, Sayed A (2016) Real-time reverse transcription PCR assay for detection of senecavirus a in swine vesicular diagnostic specimens. *PLoS ONE*, **11**, 1–13.

Breitbart M, Rohwer F (2005) Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology*, **13**, 278–284.

Brocker C, Lassen N, Estey T et al. (2010) Aldehyde dehydrogenase 7A1 (ALDH7A1) is a novel enzyme involved in cellular defense against hyperosmotic stress. *Journal of Biological Chemistry*, **285**, 18452–18463.

Bryksin A V, Matsumura I (2010) Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques*, **48**, 463–5.

Burg MB, Ferraris JD, Dmitrieva NI (2007) Cellular Response to Hyperosmotic Stresses. *Physiological Reviews*, **87**, 1441–74.

Burrill CP, Westesson O, Schulte MB, Strings VR, Segal M, Andino R (2013) Global RNA structure analysis of poliovirus identifies a conserved RNA structure involved in viral replication and infectivity. *Journal of virology*, **87**, 11670–83.

Bushell M, Sarnow P (2002) Hijacking the translation apparatus by RNA viruses. *Journal of Cell Biology*, **158**, 395–399.

Cabrerizo M, Trallero G, Pena MJ et al. (2015) Comparison of epidemiology and clinical characteristics of infections by human parechovirus vs. those by enterovirus during the first month of life. *European Journal of Pediatrics*, **174**, 1511–1516.

Calvert J, Chieochansin T, Benschop KS et al. (2010) Recombination dynamics of

human parechoviruses: Investigation of type-specific differences in frequency and epidemiological correlates. *Journal of General Virology*, **91**, 1229–1238.

Campanini G, Rovida F, Meloni F, Cascina A, Ciccocioppo R, Piralla A, Baldanti F (2013) Persistent human cosavirus infection in lung transplant recipient, Italy. *Emerging Infectious Diseases*, **19**, 1667–1669.

Cano-Gomez C, Jimenez-Clavero M. (2016) Complete Coding Genome Sequence of a Putative Novel Teschovirus. *Genome Announcement*, **4**.

Carmona-Vicente N, Buesa J, Brown PA et al. (2013) Phylogeny and prevalence of kobuviruses in dogs and cats in the UK. *Veterinary Microbiology*, **164**, 246–252.

Carter JB, Saunders VA (2007) *Virology : principles and applications*. Wiley.

Chen X, Feng Q, Wu Z et al. (2003) RNA-dependent RNA polymerase gene sequence from foot-and-mouth disease virus in Hong Kong. *Biochemical and Biophysical Research Communications*, **308**, 899–905.

Chen YH, Du W, Hagemeijer MC et al. (2015) Phosphatidylserine vesicles enable efficient en bloc transmission of enteroviruses. *Cell*, **160**, 619–630.

Cline J, Braman J, Hogrefe H (1996) PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucl. acids res.*, **24**, 3546–51.

Cordey S, Gerlach D, Junier T, Zdobnov EM, Kaiser L, Tapparel C (2008) The cis-acting replication elements define human enterovirus and rhinovirus species. *RNA (New York, N.Y.)*, **14**, 1568–78.

Deigan KE, Li TW, Mathews DH, Weeks KM (2009) Accurate SHAPE-directed RNA structure determination. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 97–102.

Delhaye S, van Pesch V, Michiels T (2004) The leader protein of Theiler's virus interferes with nucleocytoplasmic trafficking of cellular proteins. *J Virol*, **78**, 4357–4362.

Dewey CM, Cenik B, Sephton CF et al. (2011) TDP-43 is directed to stress granules by sorbitol, a novel physiological osmotic and oxidative stressor. *Molecular and cellular biology*, **31**, 1098–108.

Dimmock NJ, Easton AJ (Andrew J., Leppard KN (Keith N. (2007) *Introduction to modern virology*. Blackwell Pub.

Diviney S, Tuplin A, Struthers M, Armstrong V, Elliott RM, Simmonds P, Evans DJ (2008) A hepatitis C virus cis-acting replication element forms a long-range RNA-RNA interaction with upstream RNA sequences in NS5B. *Journal of virology*, **82**, 9008–22.

Doedens JR, Kirkegaard K (1995) Inhibition of cellular protein secretion by poliovirus proteins 2B and 3A. *The EMBO journal*, **14**, 894–907.

Donin DG, de Arruda Leme R, Alfieri AF, Alberton GC, Alfieri AA (2014) First report of Porcine teschovirus (PTV), Porcine sapelovirus (PSV) and Enterovirus G (EV-G) in pig herds of Brazil. *Tropical Animal Health and Production*, **46**, 523–528.

Donin DG, Leme R de A, Alfieri AF, Alberton GC, Alfieri AA (2015) Molecular survey of porcine teschovirus, porcine sapelovirus, and enterovirus G in captive wild boars (Sus scrofa scrofa) of Paraná state, Brazil. *Pesquisa Veterinária Brasileira*, **35**, 403–408.

Donnelly MLL, Hughes LE, Luke G, Mendoza H, Ten Dam E, Gani D, Ryan MD (2001) The “cleavage” activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring “2A-like” sequences. *Journal of General Virology*, **82**, 1027–1041.

Drexler JF, Corman VM, Lukashev AN et al. (2015) Evolutionary origins of hepatitis A virus in small mammals. *Pnas*, **112**, 15190–15195.

Van Dung N, Anh PH, Van Cuong N et al. (2016) Large-scale screening and characterization of enteroviruses and kobuviruses infecting pigs in Vietnam. *Journal of General Virology*, **97**, 378–388.

van Etten JL (2011) Another really, really big virus. *Viruses*, **3**, 32–46.

Farkas T, Fey B, Hargitt E, Parcells M, Ladman B, Murgia M, Saif Y (2012) Molecular detection of novel picornaviruses in chickens and turkeys. *Virus Genes*, **44**, 262–272.

Fauquet CM (2010) *Taxonomy, Classification and Nomenclature of Viruses* (eds Mahy BW., Van Regenmortel MH V). Academic Press: Elsevier, 80-94 pp.

Ferre-D’Amare a R, Doudna J a (2001) Methods to crystallize RNA. *Curr Protoc Nucleic Acid Chem*, **Chapter 7**, Unit 7.6.

Flather D, Semler BL (2015) Picornaviruses and nuclear functions: Targeting a cellular compartment distinct from the replication site of a positive-strand RNA virus. *Frontiers in Microbiology*, **6**, 1–27.

Flint SJ, L. W. Enquist, V. R. Racaniello, A. M. Skalka (2008) *Principles of virology*.

Gan L, Jensen GJ (2012) Electron tomography of cells. *Quarterly Reviews of Biophysics*, **45**, 27–56.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, **31**, 3784–3788.

Gazina E V, Mackenzie JM, Gorrell RJ, Anderson DA (2002) Differential requirements for COPI coats in formation of replication complexes among three genera of Picornaviridae. *Journal of virology*, **76**, 11113–22.

Geisberg JV, Moqtaderi Z, Fan X, Ozsolak F, Struhl K (2014) Global Analysis of mRNA Isoform Half-Lives Reveals Stabilizing and Destabilizing Elements in Yeast. *Cell*, **156**, 812–824.

Gershon AA (2006) Varicella-Zoster Virus. In: *Congenital and Perinatal Infections*, pp. 91–100. Humana Press, Totowa, NJ.

Ghazi F, Hughes PJ, Hyypia T, Stanway G (1998) Molecular analysis of human parechovirus type 2 (formerly echovirus 23). *Journal of General Virology*, **79**, 2641–2650.

Global Polio Eradication Initiative (2016) Global Polio Eradication Initiative.

Gong P, Peersen OB (2010) Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 22505–10.

Gong Z, Schwieters CD, Tang C (2015) Conjoined use of EM and NMR in RNA structure refinement. *PLoS ONE*, **10**, 1–9.

Goodarzi H, Elemento O, Tavazoie S (2012) Revealing global regulatory perturbations across human cancers. *Molecular cell*, **36**, 900–11.

Goodfellow I, Chaudhry Y, Richardson A, Meredith J, Almond JW, Barclay W, Evans DJ (2000) Identification of a cis-acting replication element within the poliovirus coding region. *Journal of virology*, **74**, 4590–600.

Halliwell B, Clement M V, Ramalingam J, Long LH (2000) Hydrogen peroxide. Ubiquitous in cell culture and in vivo? *IUBMB life*, **50**, 251–257.

Han J-Q, Townsend HL, Jha BK, Paranjape JM, Silverman RH, Barton DJ (2007) A phylogenetically conserved RNA structure in the poliovirus open reading frame inhibits the antiviral endoribonuclease RNase L. *Journal of virology*, **81**, 5561–5572.

Harris KS, Xiang W, Alexander L, Lane WS, Paul A V, Wimmer E (1994) Interaction of poliovirus polypeptide 3CDpro with the 5’ and 3’ termini of the poliovirus genome: Identification of viral and cellular cofactors needed for efficient binding. *Journal of Biological Chemistry*, **269**, 27004–27014.

Harvala H, Robertson I, McWilliam Leitch EC, Benschop K, Wolthers KC, Templeton K, Simmonds P (2008) Epidemiology and clinical associations of human parechovirus respiratory infections. *Journal of Clinical Microbiology*, **46**, 3446–3453.

Harvala H, Wolthers KC, Simmonds P (2010) Parechoviruses in children: understanding a new infection. *Current Opinion in Infectious Diseases*, **23**, 224–230.

He H, Liu M, Zheng Z et al. (2006) Synthesis and analgesic activity evaluation of some agmatine derivatives. *Molecules*, **11**, 393–402.

Headey SJ, Huang H, Claridge JK et al. (2007) NMR structure of stem-loop D from

human rhinovirus-14. *RNA (New York, N.Y.)*, **13**, 351–60.

Heiman M (1997) Webcutter 2.0.

Herold J, Andino R (2000) Poliovirus requires a precise 5’ end for efficient positive-strand RNA synthesis. *Journal of virology*, **74**, 6394–6400.

Hershan AA (2012) *Identification and analysis of conserved structures in RNA viruses*. University of Essex.

Hewlett MJ, Rose JK, Baltimore D (1976) 5’-terminal structure of poliovirus polyribosomal RNA is pUp. *Proceedings of the National Academy of Sciences of the United States of America*, **73**, 327–30.

Hofacker IL (2003) Vienna RNA secondary structure server. *Nucleic Acids Research*, **31**, 3429–3431.

Hofacker IL, Stadler PF, Stocsits RR (2004) Conserved RNA secondary structures in viral genomes: A survey. *Bioinformatics*, **20**, 1495–1499.

Howard CR, Fletcher NF (2012) Emerging virus diseases: can we ever expect the unexpected? *Emerging microbes & infections*, **1**, e46.

Hughes PJ, North C, Jellis CH, Minor PD, Stanway G (1988) The nucleotide sequence of human rhinovirus 1B: Molecular relationships within the rhinovirus genus. *Journal of General Virology*, **69**, 49–58.

Hughes PJ, Horsnell C, Hyypiä T, Stanway G (1995) The coxsackievirus A9 RGD motif is not essential for virus viability. *Journal of virology*, **69**, 8035–8040.

Hyypiä T, Horsnell C, Maaronen M et al. (1992) A distinct picornavirus group identified by sequence analysis. *Proceedings of the National Academy of Sciences of the United States of America*, **89**, 8847–8851.

ICTV Virus Taxonomy (2012) ICTV Virus Taxonomy.

Jääskeläinen AJ, Voutilainen L, Lehmusto R et al. (2015) Serological survey in the Finnish human population implies human-to-human transmission of Ljungan virus or antigenically related viruses. *Epidemiology and infection*, **21**, 1–8.

Jackson T, Sheppard D, Denyer M, Blakemore W, King AM (2000) The epithelial integrin alphavbeta6 is a receptor for foot-and-mouth disease virus. *Journal of virology*, **74**, 4949–56.

Jang CJ, Jan E (2010) Modular domains of the Dicistroviridae intergenic internal ribosome entry site. *RNA (New York, N.Y.)*, **16**, 1182–95.

Jonassen CM, Jonassen TØ, Grinde B (1998a) Short Communication A common RNA motif in the 3’ end of the genomes of astroviruses , avian infectious bronchitis virus and an equine rhinovirus. *Journal of General Virology*, **79**, 715–718.

Jonassen CM, Jonassen TO, Grinde B (1998b) A common RNA motif in the 3' end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. *The Journal of general virology*, **79**, 715–718.

Jones MS, Lukashov V V., Ganac RD, Schnurr DP (2007) Discovery of a novel human picornavirus in a stool sample from a pediatric patient presenting with fever of unknown origin. *Journal of Clinical Microbiology*, **45**, 2144–2150.

Kapoor A, Victoria J, Simmonds P et al. (2008) A highly prevalent and genetically diversified Picornaviridae genus in South Asian children. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 20482–7.

Kapusinszky B, Phan TG, Kapoor A, Delwart E (2012) Genetic diversity of the genus cosavirus in the family picornaviridae: A new species, recombination, and 26 new genotypes. *PLoS ONE*, **7**, 3–9.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.

Ke A, Doudna JA (2004) Crystallization of RNA and RNA–protein complexes. *Methods*, **34**, 408–414.

Kerkvliet J, Edukulla R, Rodriguez M (2010) Novel Roles of the Picornaviral 3D Polymerase in Viral Pathogenesis. *Advances in Virology*, **2010**.

Kew O, Morris-Glasgow V, Landaverde M et al. (2002) Outbreak of Poliomyelitis in Hispaniola Associated with Circulating Type 1 Vaccine-Derived Poliovirus. *Science*, **296**.

King AM., Adams MJ, Carsten EB, Lefkowitz EJ (2012) *Virus Taxonomy: Classification and Nomenclature of Viruses. Ninth Report of the International Committee on Taxonomy of Viruses.* (eds King AM., Adams M., Carstens E., Lefkowitz E.). Elsevier, San Diego, CA., 1-5 pp.

Kitajima M, Rachmadi AT, Iker BC, Haramoto E (2015) Occurrence and genetic diversity of human cosavirus in influent and effluent of wastewater treatment plants in Arizona , United. *Archives of Virology*, 1775–1779.

Knowles NJ (2013) Human parechovirus.

Kolehmainen P, Jääskeläinen A, Blomqvist S et al. (2014a) Human parechovirus type 3 and 4 associated with severe infections in young children. *The Pediatric infectious disease journal*, **33**, 1109–13.

Kolehmainen P, Jääskeläinen A, Blomqvist S et al. (2014b) Human Parechovirus Type 3 and 4 Associated with Severe Infections in Young Children. *The Pediatric Infectious Disease Journal*, **33**.

Krupina KA, Sheval E V., Lidsky P V. (2010) Variability in inhibition of host RNA

synthesis by entero- and cardioviruses. *Journal of General Virology*, **91**, 1239–1244.

Kühlbrandt W (2014) Cryo-EM enters a new era. *eLife*, **3**.

Laloo B, Simon D, Veillat V et al. (2009) Analysis of post-transcriptional regulations by a functional, integrated, and quantitative method. *Molecular & cellular proteomics : MCP*, **8**, 1777–88.

Lan D, Ji W, Yang S, Cui L, Yang Z, Yuan C, Hua X (2011) Isolation and characterization of the first Chinese porcine sapelovirus strain. *Archives of virology*, **156**, 1567–74.

Lange J, Groth M, Fichtner D et al. (2014) Virus isolate from carp: Genetic characterization reveals a novel picornavirus with two aphthovirus 2A-like sequences. *Journal of General Virology*, **95**, 80–90.

Larkin MA, Blackshields G, Brown NP et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, **23**, 2947–8.

Lawyer FC, Stoffel S, Saiki RK, Chang S, Landre PA, Abrarnson RD, Gelfand DH (1993a) High-level Expression , Purification , and Enzymatic Characterization of Full-length Thermus aquatlcus DNA Polymerase and a Truncated Form Deficient in 5 ’ to 3 ’ Exonuclease Activity. In: *Genome Research*, pp. 275–287.

Lawyer FC, Stoffel S, Saiki RK, Chang SY, Landre PA, Abramson RD, Gelfand DH (1993b) High-level expression, purification, and enzymatic characterization of full-length Thermus aquaticus DNA polymerase and a truncated form deficient in 5′ to 3′ exonuclease activity. *Genome Research*, **2**, 275–287.

Van Leer-Buter CC, Poelman R, Borger R, Niesters HGM (2016) Newly identified enterovirus c genotypes, identified in the netherlands through routine sequencing of all enteroviruses detected in clinical materials from 2008 to 2015. *Journal of Clinical Microbiology*, **54**, 2306–2314.

Leppek K, Schott J, Reitter S, Poetz F, Hammond MC, Stoecklin G (2013) Roquin promotes constitutive mrna decay via a conserved class of stem-loop recognition motifs. *Cell*, **153**, 869–881.

Li D, Lei C, Xu Z, Yang F, Liu H, Zhu Z, Li S (2016) Foot-and-mouth disease virus non-structural protein 3A inhibits the interferon- β signaling pathway. *Nature Publishing Group*, 1–12.

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

Lin T-L, Lin T-H, Chiu S-C et al. (2015) Molecular epidemiological analysis of Saffold cardiovirus genotype 3 from upper respiratory infection patients in Taiwan. *Journal of Clinical Virology*, **70**, 7–13.

Lindon JC, Nicholson JK, Holmes E, Everett JR (2000) Metabonomics: Metabolic processes studied by NMR spectroscopy of biofluids. *Concepts in Magnetic*

Resonance, **12**, 289–320.

Liu Y, Wimmer E, Paul A V (2009a) *Cis-acting RNA elements in human and animal plus-strand RNA viruses.*, Vol. 1789. 495-517 pp.

Liu Y, Wimmer E, Paul A V (2009b) Cis-acting RNA elements in human and animal plus-strand RNA viruses. *Biochimica et Biophysica Acta - Molecular Cell Research*, **1789**, 495–517.

Liu Y, Wang C, Mueller S, Paul A V., Wimmer E, Jiang P (2010) Direct interaction between two viral proteins, the nonstructural protein 2CATPase and the capsid protein VP3, is required for enterovirus morphogenesis. *PLoS Pathogens*, **6**, 75–76.

Lobert P-E, Escriou N, Ruelle J, Michiels T (1999) A coding RNA sequence acts as a replication signal in cardioviruses. *Microbiology*, **96**, 11560–11565.

Lodder WJ, Wuite M, de Roda Husman AM, Rutjes SA (2013) Environmental surveillance of human parechoviruses in sewage in the Netherlands. *Applied and Environmental Microbiology*, **79**, 6423–6428.

Low JT, Weeks KM (2010) SHAPE-directed RNA secondary structure prediction. *Methods (San Diego, Calif.)*, **52**, 150–8.

Lozano G, Fernandez N, Martinez-Salas E (2014) Magnesium-dependent folding of a picornavirus IRES element modulates RNA conformation and eIF4G interaction. *FEBS Journal*, **281**, 3685–3700.

Lukashev AN (2010) Recombination among picornaviruses. *Reviews in Medical Virology*, **20**, 327–337.

Macadam AJ, Stone DM, Almond JW, Minor PD (1994) The 5’ noncoding region and virulence of poliovirus vaccine strains. *Trends in microbiology*, **2**, 449–454.

Mager WH, de Boer AH, Siderius MH, Voss HP (2000) Cellular responses to oxidative and osmotic stress. *Cell stress & chaperones*, **5**, 73–5.

Martín J, Samoilovich E, Dunn G et al. (2002) Isolation of an intertypic poliovirus capsid recombinant from a child with vaccine-associated paralytic poliomyelitis. *Journal of virology*, **76**, 10921–8.

Mason PW, Bezborodova S V, Tina M, Henry TM (2002) Identification and Characterization of a cis -Acting Replication Element (cre) Adjacent to the Internal Ribosome Entry Site of Foot-and-Mouth Disease Virus. *Journal of virology*, **76**, 9686–9694.

Mattick JS (2003) Challenging the dogma: The hidden layer of non-protein-coding RNAs in complex organisms. *BioEssays*, **25**, 930–939.

McKnight KL, Lemon SM (1998) The rhinovirus type 14 genome contains an internally located RNA structure that is required for viral replication. *RNA (New York, N.Y.)*, **4**,

1569–84.

McLeish NJ, Williams CH, Kaloudas D, Roivainen MM, Stanway G (2012) Symmetry-Related Clustering of Positive Charges Is a Common Mechanism for Heparan Sulfate Binding in Enteroviruses. *Journal of Virology*, **86**, 11163–11170.

Mirmomeni MH, Hughes PJ, Stanway G (1997) An RNA Tertiary Structure in the 3 J Untranslated Region of Enteroviruses Is Necessary for Efficient Replication. *Journal of virology*, **71**, 2363–2370.

Moffat K, Howell G, Knox C et al. (2005) Effects of Foot-and-Mouth Disease Virus Nonstructural Proteins on the Structure and Function of the Early Secretory Pathway : 2BC but Not 3A Blocks Endoplasmic Reticulum-to-Golgi Transport Effects of Foot-and-Mouth Disease Virus Nonstructural Proteins on . *Society*, **79**, 4382–4395.

Molla A, Harris KS, Paul A V, Shin SH, Mugavero J, Wimmer E (1994) Stimulation of poliovirus proteinase 3Cpro-related proteolysis by the genome-linked protein VPg and its precursor 3AB. *The Journal of biological chemistry*, **269**, 27015–20.

Morrison JM, Racaniello VR (2009) Proteinase 2Apro is essential for enterovirus replication in type I interferon-treated cells. *Journal of virology*, **83**, 4412–4422.

Nateri AS, Hughes PJ, Stanway G (2002) Terminal RNA replication elements in human parechovirus 1. *J.Virol.*, **76**, 13116–13122.

Niklasson B, Heller KE, Schønecker B et al. (2003) Development of Type 1 Diabetes in Wild Bank Voles Associated With Islet Autoantibodies and the Novel Ljungan Virus. *International journal of experimental diabesity research*, **4**, 35–44.

Nilsson D, Sunnerhagen P (2011) Cellular stress induces cytoplasmic RNA granules in fission yeast. *RNA (New York, N.Y.)*, **17**, 120–33.

Nugent CI, Johnson KL, Sarnow P, Kirkegaard K (1999) Functional coupling between replication and packaging of poliovirus replicon RNA. *Journal of virology*, **73**, 427–35.

Oberste MS, Maher K, Flemister MR, Kilpatrick DR, Pallansch M a, Marchetti G (2000) Comparison of Classic and Molecular Approaches for the Identification of Untypeable Enteroviruses Comparison of Classic and Molecular Approaches for the Identification of Untypeable Enteroviruses †. *Journal of Clinical Microbiology*, **38**, 1170–1174.

Olarte-Castillo XA, Heeger F, Mazzoni CJ et al. (2015) Molecular characterization of canine kobuvirus in wild carnivores and the domestic dog in Africa. *Virology*, **477**, 89–97.

Oldstone MBA (2010) *Viruses, plagues, and history : past, present, and future*. Oxford University Press.

De Palma AM, Vliegen I, De Clercq E, Neyts J (2008) Selective inhibitors of picornavirus replication. *Medicinal research reviews*, **28**, 823–84.

Palmenberg AC, Spiro D, Kuzmickas R et al. (2009) Sequencing and Analyses of All Known Human Rhinovirus Genomes Reveal Structure and Evolution. *Science*, **324**.

Pandey M, Bajaj GD, Rath PC (2004) Induction of the interferon-inducible RNA-degrading enzyme, RNase L, by stress-inducing agents in the human cervical carcinoma cells. *RNA biology*, **1**, 21–7.

Pankovics P, Boros Á, Kiss T, Reuter G (2015) Identification and complete genome analysis of kobuvirus in faecal samples of European roller (Coracias garrulus): for the first time in a bird. *Archives of Virology*, **160**, 345–351.

Pathak HB, Arnold JJ, Wiegand PN, Hargittai MRS, Cameron CE (2007) Picornavirus genome replication: assembly and organization of the VPg uridylylation ribonucleoprotein (initiation) complex. *The Journal of biological chemistry*, **282**, 16202–13.

Paul A V, Rieder E, Kim DW, van Boom JH, Wimmer E (2000) Identification of an RNA hairpin in poliovirus RNA that serves as the primary template in the in vitro uridylylation of VPg. *Journal of virology*, **74**, 10359–70.

Paul A V, Peters J, Mugavero J, Yin J, van Boom JH, Wimmer E (2003) Biochemical and genetic studies of the VPg uridylylation reaction catalyzed by the RNA polymerase of poliovirus. *Journal of virology*, **77**, 891–904.

Pelletier J, Sonenberg N (1988) Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature*, **334**, 320–325.

Pham KTN, Trinh QD, Takanashi S et al. (2010) Novel human parechovirus, Sri Lanka. *Emerging infectious diseases*, **16**, 130–132.

Pounder KC, Watts PC, Niklasson B et al. (2015) Genome characterisation of two Ljungan virus isolates from wild bank voles (Myodes glareolus) in Sweden. *Infection, Genetics and Evolution*, **36**, 156–164.

Purcell RH, Emerson SU (2001) Animal Models of Hepatitis A and E. *ILAR Journal*, **42**, 161–177.

Reddy PS, Burroughs KD, Hales LM et al. (2007) Seneca Valley virus, a systemically deliverable oncolytic picornavirus, and the treatment of neuroendocrine cancers. *Journal of the National Cancer Institute*, **99**, 1623–1633.

Van Regenmortel MH V, Mahy BWJ (2004) Emerging Issues in Virus Taxonomy. *Emerging Infectious Diseases*, **10**, 8–13.

Rieder E, Wimmer E (2002) Cellular receptors of picornaviruses: an overview. In: *Molecular Biology of Picornaviruses* (eds Selmer B, Wimmer E), pp. 61–70. ASM Press, Washington.

Robertson MP, Igel H, Baertsch R, Haussler D, Ares M, Scott WG (2004) The Structure of a Rigorously Conserved RNA Element within the SARS Virus Genome. *PLoS Biology*, **3**, 86–94.

Rohll JB, Moon DH, Evans DJ, Almond JW (1995) The 3' untranslated region of picornavirus RNA: features required for efficient genome replication. *Journal of virology*, **69**, 7835–7844.

Romanova LI, Lidsky P V, Kolesnikova MS et al. (2009) Antiapoptotic activity of the cardiovirus leader protein, a viral "security" protein. *Journal of virology*, **83**, 7273–7284.

Romeo C, Ferrari N, Rossi C et al. (2014) Ljungan virus and an adenovirus in Italian squirrel populations. *Journal of Wildlife Diseases*, **50**, 409–411.

Ruan Q, Quintanilla RA, Johnson GVW (2007) Type 2 transglutaminase differentially modulates striatal cell death in the presence of wild type or mutant huntingtin. *Journal of Neurochemistry*, **102**, 25–36.

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

Sanders BP, de los Rios Oakes I, van Hoek V et al. (2016) Cold-Adapted Viral Attenuation (CAVA): Highly Temperature Sensitive Polioviruses as Novel Vaccine Strains for a Next Generation Inactivated Poliovirus Vaccine. *PLoS Pathogens*, **12**, 1–23.

Sano R, Reed JC (2013) ER stress-induced cell death mechanisms. *Biochimica et Biophysica Acta - Molecular Cell Research*, **1833**, 3460–3470.

Sasaki J, Taniguchi K (2003a) The 5'-end sequence of the genome of Aichi virus, a picornavirus, contains an element critical for viral RNA encapsidation. *Journal of virology*, **77**, 3542–8.

Sasaki J, Taniguchi K (2003b) The 5′-End Sequence of the Genome of Aichi Virus , a Picornavirus , Contains an Element Critical for Viral RNA Encapsidation. *Journal of virology*, **77**, 3542–3548.

Saunus JM, Edwards SL, French JD, Smart CE, Brown MA (2007) *Regulation of BRCA1 messenger RNA stability in human epithelial cell lines and during cell cycle progression*, Vol. 581. 3435-3442 pp.

Schieber M, Chandel NS (2014) ROS function in redox signaling and oxidative stress. *Current Biology*, **24**, R453–R462.

Schneider-Schaulies J (2000) Cellular receptors for viruses : links to tropism and pathogenesis. *Journal of General Virology*, **81**, 1413–1429.

Schwarz KB (1996) Oxidative stress during viral infection: a review. *Free radical biology & medicine*, **21**, 641–9.

Seal LA, Jamison RM (1984) Evidence for secondary structure within the virion RNA of echovirus 22. *Journal of virology*, **50**, 641–4.

Seitsonen JJT, Shakeel S, Susi P et al. (2012) Structural analysis of coxsackievirus A7 reveals conformational changes associated with uncoating. *Journal of virology*, **86**, 7207–15.

Semler BL, Wimmer E (2002) *Molecular Biology Of Picornaviruses*. ASM Press, Washington.

Severi F, Conticello SG (2015) Flow-cytometric visualization of C>U mRNA editing reveals the dynamics of the process in live cells. *RNA Biology*, **12**, 389–397.

Shafren DR, Dorahy DJ, Ingham RA, Burns GF, Barry RD (1997) Coxsackievirus A21 binds to decay-accelerating factor but requires intercellular adhesion molecule 1 for cell entry. *Journal of virology*, **71**, 4736–43.

Sharp J, Harrison CJ, Puckett K et al. (2015) during Sepsis Evaluations. *The Pediatric infectious disease journal*, **32**, 213–216.

Shen M, Wang Q, Yang Y et al. (2007) Human rhinovirus type 14 gain-of-function mutants for oriI utilization define residues of 3C(D) and 3Dpol that contribute to assembly and stability of the picornavirus VPg uridylylation complex. *Journal of virology*, **81**, 12485–95.

Sherf BA, Navarro SL, Hannah RR, Wood K V (1996) Dual-Luciferase TM Reporter Assay: An Advanced Co-Reporter Technology Integrating Firefly and Renilla Luciferase Assays. *Promega Notes Magazine Number*, **57**.

Smith KM (1943) *Beyond the Microscope:* (eds Beales HL, Williams WE). Penguin Books, 9-17 pp.

Smits SL, Raj VS, Oduber MD et al. (2013) Metagenomic Analysis of the Ferret Fecal Viral Flora. *PLoS ONE*, **8**.

Somarowthu S, Legiewicz M, Chillón I, Marcia M, Liu F, Pyle AM (2015) HOTAIR Forms an Intricate and Modular Secondary Structure. *Molecular Cell*, **58**, 353–361.

Son KY, Kim DS, Matthijnssens J et al. (2014) Molecular epidemiology of Korean porcine sapeloviruses. *Archives of Virology*, **159**, 1175–1180.

Song Y, Liu Y, Ward CB et al. (2012) Identification of two functionally redundant RNA elements in the coding sequence of poliovirus using computer-generated design. *Proceedings of the National Academy of Sciences*, **109**, 14301–14307.

Stanway G (1990) Structure , function and evolution of picornaviruses r [] I I illi II. *Journal of General Virology*, **71**, 2483–2501.

Stanway G, Hyypiä T (1999) Parechoviruses MINIREVIEW. *Journal of virology*, **73**, 5249–5254.

Stanway G, Kalkkinen N, Roivainen M et al. (1994) Molecular and biological characteristics of echovirus 22, a representative of a new picornavirus group. *Journal of virology*, **68**, 8232–8.

Stanway G, Joki-Korpela P, Hyypiä T (2000) Human parechoviruses-biology and clinical significance. *Reviews in Medical Virology*, **13**, 57–69.

Steil BP, Barton DJ (2009) Cis-active RNA elements (CREs) and picornavirus RNA replication. *Virus Research*, **139**, 240–252.

Stothard P (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques*, **28**, 1102, 1104.

Strauss E, Strauss J (2007) *Viruses and Human Disease*. Elsevier / Academic Press.

Strauss DM, Wuttke DS (2007) Characterization of protein-protein interactions critical for poliovirus replication: analysis of 3AB and VPg binding to the RNA-dependent RNA polymerase. *Journal of virology*, **81**, 6369–78.

Suhy DA, Giddings TH, Kirkegaard K (2000) Remodeling the endoplasmic reticulum by poliovirus infection and by individual viral proteins: an autophagy-like origin for virus-induced vesicles. *Journal of virology*, **74**, 8953–65.

Svitkin Y V, Herdy B, Costa-Mattioli M, Gingras A-C, Raught B, Sonenberg N (2005) Eukaryotic translation initiation factor 4E availability controls the switch between cap-dependent and internal ribosomal entry site-mediated translation. *Molecular and cellular biology*, **25**, 10556–10565.

Sweeney TR, Dhote V, Yu Y, Hellen CUT (2012) A distinct class of internal ribosomal entry site in members of the Kobuvirus and proposed Salivirus and Paraturdivirus genera of the Picornaviridae. *Journal of virology*, **86**, 1468–86.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Tapia G, Bøås H, De Muinck EJ et al. (2015) Saffold virus, a human cardiovirus, and risk of persistent Islet autoantibodies in the longitudinal birth cohort study MIDIA. *PLoS ONE*, **10**, 1–11.

Thiviyanathan V, Yang Y, Kaluarachchi K, Rijnbrand R, Gorenstein DG, Lemon SM (2004) High-resolution structure of a picornaviral internal cis-acting RNA replication element (cre). *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 12688–93.

Tijerina P, Mohr S, Russell R (2007) DMS Footprinting of Structured RNAs and RNA-Protein Complexes. *Biochemistry*, **2**, 2608–2623.

Todd S, Semler BL (1996) Structure- infectivity analysis of the human rhinovirus genomic

RNA 3′ non-coding region. *Mutagenesis*, **24**, 2133–42.

Tolf C, Gullberg M, Johansson ES, Tesh RB, Andersson B, Lindberg AM (2009) Molecular characterization of a novel Ljungan virus (Parechovirus; Picornaviridae) reveals a fourth genotype and indicates ancestral recombination. *Journal of General Virology*, **90**, 843–853.

Townsend HL, Jha BK, Han J-Q, Maluf NK, Silverman RH, Barton DJ (2008) A viral RNA competitively inhibits the antiviral endoribonuclease domain of RNase L. *RNA (New York, N.Y.)*, **14**, 1026–36.

Tsai ATH, Kuo CC, Kuo YC, Yang JL, Chang CY, Wang FI (2016) The urinary shedding of porcine teschovirus in endemic field situations. *Veterinary Microbiology*, **182**, 150–155.

Tulloch F, Atkinson NJ, Evans DJ, Ryan MD, Simmonds P (2014a) RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. *Elife*, **3**, e04531.

Tulloch F, Pathania U, Luke GA et al. (2014b) FMDV replicons encoding green fluorescent protein are replication competent. *Journal of Virological Methods*, **209**, 35–40.

Tuthill TJ, Groppelli E, Hogle JM, Rowlands DJ (2010) Picornaviruses. *Current topics in microbiology and immunology*, **343**, 43–89.

Tycowski KT, Shu M-D, Borah S, Shi M, Steitz JA (2012) Conservation of a triple-helix-forming RNA stability element in noncoding and genomic RNAs of diverse viruses. *Cell reports*, **2**, 26–32.

Walker AK, Soo KY, Sundaramoorthy V et al. (2013) ALS-associated TDP-43 induces endoplasmic reticulum stress, which drives cytoplasmic TDP-43 accumulation and stress granule formation. *PloS one*, **8**.

Wang RYL, Kuo R-L, Ma W-C et al. (2013) Heat shock protein-90-beta facilitates enterovirus 71 viral particles assembly. *Virology*, **443**, 236–47.

Welnowska E, Sanz MA, Redondo N, Carrasco L (2011) Translation of viral mRNA without active eIF2: The case of picornaviruses. *PLoS ONE*, **6**.

Wen XJ, Cheng AC, Wang MS et al. (2014) Detection, differentiation, and VP1 sequencing of duck hepatitis A virus type 1 and type 3 by a 1-step duplex reverse-transcription PCR assay. *Poultry science*, **93**, 2184–92.

Wessels E, Duijsings D, Notebaart RA, Melchers WJG, van Kuppeveld FJM, Melchers JG, Kuppeveld FJM Van (2005) A proline-rich region in the coxsackievirus 3A protein is required for the protein to inhibit endoplasmic reticulum-to-golgi transport. *Journal of virology*, **79**, 5163–73.

White JP, Lloyd RE, Franks TM et al. (2012) Regulation of stress granules in virus

systems. *Trends in microbiology*, **20**, 175–83.

Wildenbeest JG, Harvala H, Pajkrt D, Wolthers KC (2010) The need for treatment against human parechoviruses: how, why and when? *Expert Review of Anti-infective Therapy*, **8**, 1417–1429.

Wildenbeest JG, Wolthers KC, Straver B, Pajkrt D (2013) Successful IVIG treatment of human parechovirus-associated dilated cardiomyopathy in an infant. *Pediatrics*, **132**, e243-7.

Wilkinson KA, Gorelick RJ, Vasa SM et al. (2008) High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biology*, **6**, 883–899.

Willcocks MM, Locker N, Gomwalk Z et al. (2011) Structural features of the Seneca Valley virus internal ribosome entry site (IRES) element: a picornavirus with a pestivirus-like IRES. *Journal of virology*, **85**, 4452–4461.

Williams CH, Panayiotou M, Girling GD, Peard CI, Oikarinen S, Hyöty H, Stanway G (2009) Evolution and conservation in human parechovirus genomes. *The Journal of general virology*, **90**, 1702–1712.

Withers JB, Beemon KL (2011) The structure and function of the rous sarcoma virus RNA stability element. *Journal of cellular biochemistry*, **112**, 3085–92.

Witwer C, Rauscher S, Hofacker IL, Stadler PF (2001) Conserved RNA secondary structures in Picornaviridae genomes. *Nucleic Acids Research*, **29**, 5079–5089.

Woo PCY, Lau SKP, Choi GKY et al. (2016) Equine rhinitis B viruses in horse fecal samples from the Middle East. *Virology Journal*, **13**, 94.

Yamashita T, Sakae K, Kobayashi S, Ishihara Y, Miyake T, Mubina A, Isomura S (1995) Isolation of Cytopathic Small Round Virus (Aichi Virus) from Pakistani Children and Japanese Travelers from Southeast Asia. *Microbiology and inmunology*, **39**, 433–435.

Yang Y, Yi M, Evans DJ, Simmonds P, Lemon SM (2008) Identification of a conserved RNA replication element (cre) within the 3Dpol-coding sequence of hepatoviruses. *Journal of virology*, **82**, 10118–10128.

Yang S, Zhang W, Shen Q, Yang Z, Zhu J, Cui L, Hua X (2009) Aichi Virus Strains in Children with Gastroenteritis, China. *Emerging Infectious Diseases*, **15**, 1703–1705.

Yip CCY, Lo K, Que T et al. (2014) Epidemiology of human parechovirus , Aichi virus and salivirus in fecal samples from hospitalized children with gastroenteritis in Hong Kong. *Virology Journal*, **11**, 1–10.

Yokota T, Sugawara K, Ito K, Takahashi R, Ariga H, Mizusawa H (2003) Down regulation of DJ-1 enhances cell death by oxidative stress, ER stress, and proteasome inhibition. *Biochemical and Biophysical Research Communications*, **312**, 1342–

1348.

Ypma-Wong MF, Dewalt PG, Johnson VH, Lamb JG, Semler BL (1988) Protein 3CD is the major poliovirus proteinase responsible for cleavage of the P1 capsid precursor. *Virology*, **166**, 265–70.

Zhu L, Wang X, Ren J et al. (2015) Structure of Ljungan virus provides insight into genome packaging of this picornavirus. *Nature Communications*, **6**, 8316.

Zuker M (1989) On Finding All Suboptimal Foldings of an RNA Molecule. *Science*, **244**.

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, **31**, 3406–3415.

Zuker M, Jaeger JA, Turner DH (1991) A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures determined by phylogenetic comparison. *Nucleic Acids Res.*, **19**, 2707–2714.

Appendices

Appendix A

Desired sequence of the constructs, T5dsRed, T5dsREd3D and T5dsRed3DS, used for the IRES driven translation experiment

gaagcttaaACCCCGACTTGCTGAGCTTCTCTAGGAGAGTCCCTTTCCCAGCCTTGGGGTGGCTGGTCAATAAAAACC
CCATATGTAACCAACACCTAAGACAATTTGGTCAACCCTATGCCTGGTCCCCACTATTCGAAGGCAACTTGCAATAAG
AAGAGTGGAACAAGGATGCTTAAAGCATAGTGTAAATGATCTTTTCTAACCTGTATTATGTACAGGGTGGCAGATGGC
GTGCCATAAATCTATTAGTGGGATACCACGCTTGTGGACCTTATGCCCACACAGCCATCCTCTAGTAAGTTTGTAAAA
TGTCTGGTGAGATGTGGGAACTTATTGGAAACAACAATTTGCTTAATAGCATCCTAGTGCCAGCGGAACAACATCTGG
TAACAGATGCCTCTGGGGCCAAAAGCCAAGGTTTGACAGACCCATTAGGATTGGTTTCAAAACCTGAATTGTTGTGGA
AGATATTCAGTACCTATCAATCTGGTAGTGGTGCAAACACTAGTTGTAAGGCCCACGAAGGATGCCCAGAAGGTACCC
GCAGGTAACAAGAGACACTGTGGATCTGATCTGGGGCCAACTACCTCTATCAGGTGAGTTAGTTAAAAAACGTCTAGT
GGGCCAAACCCAGGGGGGATCCCTGGTTTCCTTTTATTGTTAATATTGACATTATGGTGCGCTCCTCCAAGAACGTCA
TCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCACCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCG
AGGGCCGCCCCTACGAGGGCCACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACA
TCCTGTCCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCCGACATCCCCGACTACAAGAAGCTGT
CCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCT
CCCTGCAGGACGGCTGCTTCATCTACAAGGTGAAGTTCATCGGCGTGAACTTCCCCTCCGACGGCCCCGTAATGCAGA
AGAAGACCATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGGGCGAGATCCACAAGG
CCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGTCCATCTACATGGCCAAGAAGCCCGTGCAGCTGC
CCGGCTACTACTACGTGGACTCCAAGCTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAGCAGTACGAGC
GCACCGAGGGCCGCCACCACCTGTTCCTGtaaccgcgga

GTGTATTTACACTACAAATTTAATCAGTCCT GGAATTGATTGTCTACCAATCGTATATGG
CGATGATGTTATTCTTTCATTGGACAAAGAAATTGAACCAGAGAAACTGCAAAGTATCAT
GGCAGATTCATTTGGAGCCGAAGTGACTGGGTCACGCAAAGATGAACCTCCTTCACTTAA
ACCAAGAATGGAGGTTGAATTCTTAAAGCGGAAACCTGGTTATTTCCCAGAGTCTACTTT
TATAGTAGGTAAATTAGACACTGAAAATATGATACAACATTTAATGTGGATGAAAAACTT
TAGCACATTTAAGCAGCAGCTCCAATCCTATTTAATGGAACTATGCCTCCATGGAAAAGA
CACTTATCAACACTACATTAAGATCTTGGAACCATATCTACAGGAATGGAATATCACTGT
GGATGATTATGATGTGGTTATAACTAAGTTGATGCCCATGGTGTTTGATTAAGATTAATG
TTTTGTTTTTCTTTTGCTATGGACTATGTGGTTAGAACACTAATTAATTAGGA CACTAAC
TTGGAACATTGGACATGAC

Figure 1: Nucleotide sequence of the DNAs that form the constructs; T5dsRed, T5dsRed3D and T5dsRed 3DS. The beginning of the sequence (highlighted in green) to the end of the sequence (highlighted in red) forms the constructs T5dsRed. The green highlight is the forward primer for 5' UTR joined to dsRed which works with the reverse primer highlighted in yellow to form T5dsRed3D, while the green highlighted primer also works with the reverse primer highlighted in blue to form T5dsRed3DS. The white font colour is the forward primer for 2nd fragment (dsRed overhang with part of the $3D^{pol}$ of HPeV1) while the pink highlight is the target site for the reverse primer that was used with 1st fragment (dsRed overhang and part of the 3Dpol + 3'UTR of HPeV1 (OL2093). The red colour shows the region of dsRed.

Appendix B

Production of pHPeV1Luc and the cre mutations in pHPeV1Luc

THE sequence of the insert in the vector of the desired pHPeV1Luc construct

1 gagctctttgaaaggggtctcctagagagcttggccgtcgggccttataccccga
cttgctgagcttctctaggagagtccctttcccagccttggggtggctgg
tcaataaaaaccccatatgtaaccaacacctaagacaatttggtcaaccc
tatgcctggtccccactattcgaaggcaactt
gcaataagaagagtggaacaaggatgcttaaagcatagtgtaaatgatcttttctaacct
gtattatgtacagggtggcagatggcgtgccataaatctattagtgggataccacgcttg
tggaccttatgcccacacagccatcctctagtaagtttgtaaaatgtctggtgagatgtg
ggaacttattggaaacaacaatttgcttaatagcatcctagtgccagcggaacaacatct
ggtaacagatgcctctggggccaaaagccaaggtttgacagacccattaggattggtttc
aaaacctgaattgttgtggaagatattcagtacctatcaatctggtagtggtgcaaacac
tagttgtaaggcccacgaaggatgcccagaaggtacccgcaggtaacaagagacactgtg
gatctgatctggggccaactacctctatcaggtgagttagttaaaaaacgtctagtgggc
caaacccaggggggatccctggtttccttttattgttaatattgacatta tggcttccaag gtgtacgacc
ccgagcaacg caaacgcatg
541 atcactgggc ctcagtggtg ggctcgctgc aagcaaatga acgtgctgga ctccttcatc
601 aactactatg attccgagaa gcacgccgag aacgccgtga tttttctgca tggtaacgct
661 gcctccagct acctgtggag gcacgtcgtg cctcacatcg agcccgtggc tagatgcatc
721 atccctgatc tgatcggaat gggtaagtcc ggcaagagcg ggaatggctc atatcgcctc
781 ctggatcact acaagtacct caccgcttgg ttcgagctgc tgaaccttcc aaagaaaatc
841 atctttgtgg gccacgactg gggggcttgt ctggcctttc actactccta cgagcaccaa
901 gacaagatca aggccatcgt ccatgctgag agtgtcgtgg acgtgatcga gtcctgggac
961 gagtggcctg acatcgagga ggatatcgcc ctgatcaaga gcgaagaggg cgagaaaatg
1021 gtgcttgaga ataacttctt cgtcgagacc atgctcccaa gcaagatcat gcggaaactg
1081 gagcctgagg agttcgctgc ctacctggag ccattcaagg agaagggcga ggttagacgg
1141 cctaccctct cctggcctcg cgagatccct ctcgttaagg gaggcaagcc cgacgtcgtc
1201 cagattgtcc gcaactacaa cgcctacctt cgggccagcg acgatctgcc taagatgttc
1261 atcgagtccg accctgggtt cttttccaac gctattgtcg agggagctaa gaagttccct
1321 aacaccgagt tcgtgaaggt gaagggcctc cacttcagcc aggaggacgc tccagatgaa
1381 atgggtaagt acatcaagag cttcgtggag cgcgtgctga agaacgagca g

gctagcggttct gttgttacat tccagaattcatggggttca cagatggaaacc acacaggctg acttatgtat
1381 cccctatgtt gctgacacaa actatgtcaa gaccgattcg tcagacttag ggcagttaaa
1441 agtctatgtc tggactggttct gttgttacat tccagaattcatggggttca cagatgtacttaag actat
ttccaaattt gaatgttttt gttaacagtt attcatactt
1981 taggggttca ttagttttaa gattgagcgt ctacgccagt actttcaata gaggtcgatt
2041 gaggatgggc ttcttcccca atgccacaac ggactcaact tccaccttgg acaatgctat
2101 atacacaata tgtgatatag gtagtgacaa tagttttgaa attacaatcc cttactcttt
2161 ttccacttgg atgaggaaaa caaatggcca cccaattgga ttgtttcaaa ttgaagtctt
2221 aaacaggctc acatacaaca gttctagtcc atcggaggtt tattgtatag tccaaggtaa
2281 aatgggacaa gatgcca ggt tcttctgccc aactggttct gttgttacat tccagaattc
2341 atggggttca cagatg gatt taaccgaccc tctttgtatt gaagatgaca cagaaaattg
2401 caaacaaaca atgtctccaa atgaactagg actcacttca gcccaagatg atggcccact

2461 tggtcaagaa aagccaaatt attttctcaa ttttaggtcg atgaatgtgg acatttttac
2521 tgtatcacat actaaagtag ataacctatt tgggcgggca tggtttttta tggagcatac
2581 tttcaccaat gagggacaat ggagagtgcc attggaattt ccaaaacaag gtcatgggtc
2641 cttatcactg ttgtttgctt attttactgg tgaactgaat atccatgttc tgttcctaag
2701 tgagaggggg tttctgaggg ttgcacacac atatgacact agtaatgatc gagtcaattt
2761 tctgtcatcg aacggtgtaa taactgtacc agccggagag cagatgacac tttcagctcc
2821 ctactattca aacaaaccat taagaactgt cagagataac aatagtcttg gttatttgat
2881 gtgcaagccc ttcttgactg gaacctctac tggtaaaatt gaggtttatc ttagcctgag
2941 atgtccaaat ttctttttcc ctcttcctgc ccctaaggtt acgagtagtc gtgcactacg
3001 gggtgatatg gcaaacctta caaatcagag tccatatggt caacaaccac agaatcgtat
3061 gatgaaacta gcatatttag atagaggttt ctacaaacat tatggcatta tagttggaga
3121 ccatgtgtac caactagatt cagatgacat tttcaagaca gcattgacag gaaaagccaa
3181 atttactaaa actaagttga cttcagattg ggttattgaa gaagagtgtg aattagacta
3241 ctttagaatc aaatatcttg aatcagctgt ggattcagag cacatttttt cagtagacaa
3301 aaactgtgaa accattgcca aggatatttt tggaacccat acactaagcc aacatcaagc
3361 aatagggtta gtgggaacca ttttattgac tgctggcctt atgtcaacta taaaaacccc
3421 agttaatgct gtgacaataa aagaattctt taatcatgcc atagatggcg atgagcaagg
3481 tctttccttg cttgtacaaa aatgtactac attcttttct tctgctgcaa ctgagatttt
3541 ggataatgat ttagtcaagt tcatagtgaa aatcctggtg agaatacttt gctacatggt
3601 actttattgc cataaaccta acattttgac tactgcttgc ttatccactc ttttaataat
3661 ggatgttact tcctcttcag tgctgtcacc ttcttgtaaa gccttgatgc agtgtctgat
3721 ggatggtgat gttaagaaac tggctgaaat tgtggcagaa tcaatgtcca acactgatga
3781 tgatgaagtt aaggagcaaa tatgtgacac agtaaagtat actaagacca tcctatcaaa
3841 tcaaggacct tttaaaggat tcaatgaagt ttccacagca tttaggcata tagactggtg
3901 gatacataca ttactcaaaa tcaaagatat ggttcttagt gttttcaaac ccagcataga
3961 aagcaaggct attcagtggt tggagagaaa taaggaacat gtttgtagta tccttgatta
4021 tgcctcagac attattgttg agtcaaaaga ccaatctaag atgaagactc aagatttcta
4081 ccaaaggtat tcagattgtt tggccaagtt caagccaatc atggcaattt gctttagaag
4141 ctgccacaac agtatcagca acacagttta cagacttttc caggagctgg ccaggatccc
4201 caaccgaatc agtaccaata atgatttaat tagaattgag cccattggca tttggatcca
4261 aggtgagcca ggacaaggaa aatctttctt gacccacacc ctctccagac aattacaaaa
4321 atcatgtaag cttaatggtg tttttactaa tccaactgct agtgagttca tggatggtta
4381 tgataatcag gacattcatt tgattgatga cttgggtcaa acaagaaaag agaaagatat
4441 tgaaatgttg tgcaattgca tttcttctgt tccctttatt gtgccaatgg cacatcttga
4501 agaaaaaggg aaattttata ctagcaagtt agttgttgcc accactaaca agtcagattt
4561 ttctagcacg gtcctccagg attctggggc actgaagagg agattccctt acattatgca
4621 cattcgagca gcaaaggcct acagcaaggc tggaaagctc aatgtaagcc aggccatggc
4681 aacaatgtca actggtgagt gctgggaagt atcaaagaat ggaagggatt gggaaacact
4741 aaaattgaaa gacctagttg acaaaattac aatcgattat aatgagaggg ttaagaatta
4801 taatgcttgg aaacaacaat tggaaaatca gactcttgat gatctagatg atgctgtatc
4861 atacattaag cacaatttcc cagatgccat accatatgtt gatgagtatc ttaatattga

4921 aatgtcaaca ttaattgaac aaatggaggc tttcatagaa ccaaagccta gtgtcttcaa
4981 atgctttgct aacaaaattg gttcaaaaat ttctaaagct tctagggaag ttgtggactg
5041 gttttcagat aaaattaaat ccatgctcag ttttgttgag aggaataagg cttggctcac
5101 agttgtatct gctgtcacca gtgctattag tatacttttg ttggtaacaa agatcttcaa
5161 gaaagaggaa tctaaagatg agagagctta taatcctact ttaccggttg ccaaacccaa
5221 gggcactttc ccagtttccc agcgggagtt caaaaatgaa gctccttatg atggacaatt
5281 ggaacacatc atttctcaaa tggcatatat tactggttca acaactggcc atatgactca
5341 ttgcgccggt tatcaacatg atgaaattat actccatggt cattccatta agtatttaga
5401 acaagaagat gaattgacac tacattacaa gaataaagtt ttcccaattg aacaaccatc
5461 tgttacccaa gtgacacttg gtggtaaacc aatggatttg gctattctca agtgtaagtt
5521 gcctttcagg ttcaagaaaa actctaaata ttacaccaac aagattggaa ctgaaagtat
5581 gctaatctgg atgactgagc aaggcattat taccaaggaa gtccaaagag tgcaccattc
5641 aggtggcatc aagacaaggg aaggaactga aagcacaaag actatcagtt atactgttaa
5701 atcttgcaaa ggaatgtgtg gtggcctact tatttcaaaa gtagaaggta acttcaaaat
5761 cctgggtatg catattgctg gtaatggtga aatgggagta gctataccct ttaattttct
5821 taaaaatgac atgtctgatc aaggcattgt tactgagatc actccaatcc agcccatgta
5881 tataaacact aaaactcaga tccacaagag ccctgtctat ggtgctgttg aggtgaaaat
5941 gggtccagct gtcctaagta agtcagacac caggttggag gaacctgtag aatgtcttat
6001 taaaaaatcg gcttccaagt acagagttaa caaattccag gtgaacaatg aactctggca
6061 aggtgtcaaa gcatgtgtta aatctaaatt cagagaaatc tttggaatga atggtattgt
6121 tgacatgaaa actgccattt taggaacttc tcatgtaaat tctatggatt tgagcacttc
6181 agctggttat agttttgtta aatctggcta caaaaagaaa gatctaattt gcttggaacc
6241 attctcagta gccccattgt tagagaggct tgtacaggac aaattccaca acttactaaa
6301 agggaatcaa ataactacaa ccttcaacac ttgccttaaa gatgagttaa gaaagctaga
6361 taagattgcc tcaggaaaaa ctaggtgtat agaggcttgt gaagttgatt attgcattgt
6421 ttacagaatg atcatgatgg aaatttatga caagatttat caaactccat gttactactc
6481 aggactagct gttggaatca acccatacaa agattggcac ttcatgatta atgctctcaa
6541 tgattacaat tatgaaatgg actattctca atatgatggc tcccttagtt caatgttgct
6601 atgggaggct gtggaagttc tagcttactg tcatgattca cctgatctcg tgatgcaact
6661 gcacaaacca gtcattgact ctgaccatgt tgttttcaac gagagatggt tgattcatgg
6721 tggtatgcca tcaggatctc catgtactac tgtgttaaat tcactatgca atttgatgat
6781 gtgtatttac actacaaatt taatcagtcc tggaattgat tgtctaccaa tcgtatatgg
6841 cgatgatgtt attctttcat tggacaaaga aattgaacca gagaaactgc aaagtatcat
6901 ggcagattca tttggagccg aagtgactgg gtcacgcaaa gatgaacctc cttcacttaa
6961 accaagaatg gaggttgaat tcttaaagcg gaaacctggt tatttcccag agtctacttt
7021 tatagtaggt aaattagaca ctgaaaatat gatacaacat ttaatgtgga tgaaaaactt
7081 tagcacattt aagcagcagc tccaatccta tttaatggaa ctatgcctcc atggaaaaga
7141 cacttatcaa cactacatta agatcttgga accatatcta caggaatgga atatcactgt
7201 ggatgattat gatgtggtta taactaagtt gatgcccatg gtgtttgatt aagattaatg
7261 ttttgttttt cttttgctat ggactatgtg gttagaacac taattaatta ggacactaac
7321 ttggaacatt ggacatgac

6661 gcacaaacca gtcattgact ctgaccatgt tgttttcaac gagagatggt tgattcatgg

6721 tggtatgcca tcaggatctc catgtactac tgtgttaaat tcactatgca atttgatgat

6781 gtgtatttac actacaaatt taatcagtcc tggaattgat t gtctaccaa tcgtatatgg

6841 cgatgatgtt attctttcat tggacaaaga aattgaacca gagaaactgc aaagtatcat

6901 ggcagattca tttggagccg aagtgactgg gtcacgcaaa gatgaacctc cttcacttaa

6961 accaagaatg gaggttgaat tcttaaagcg gaaacctggt tatttcccag agtctacttt

7021 tatagtaggt aaattagaca ctgaaaatat gatacaacat ttaatgtgga tgaaaaactt

7081 tagcacattt aagcagcagc tccaatccta tttaatggaa ctatgcctcc atggaaaaga

7141 cacttatcaa cactacatta agatcttgga accatatcta caggaatgga atatcactgt

7201 ggatgattat gatgtggtta taactaagtt gatgcccatg gtgtttgat **taa**gattaatg

7261 ttttgttttt cttttgctat ggactatgtg gttagaacac taattaatta ggacactaac

7321 ttggaacatt ggacatgac

Figure 2: Complete sequence of the insert DNA contained in the vector giving the pHPeV1Luc construct. The red highlights show sites of the restriction enzymes used (Table 5.1), green highlight is the start codon which begins the translation of the Luciferase as well as the rest of the RNA while the pink highlight shows the portion of the synthesized cre now placed in the VP1 region of the RNA gene. The red font shows the position of the of the 3DS structure

A. Original cre sequence in pHPeV1Luc using the forward primer (Table 2.4)

```
NNNNNNNNNTCNNNTCNANCNGGGTTCTTTTCCAACGCTATTGTCGAGGGAGCTAAGAAGTTCCCTAACACCGAGTTCGT
GAAGGTGAAGGGCCTCCACTTCAGCCAGGAGGACGCTCCAGATGAAATGGGTAAGTACATCAAGAGCTTCGTGGAGCGCG
TGCTGAAGAACGAGCAGGCTAGCGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGGAAACCACACAGGCT
GACTTATGTATCCCCTATGTTGCTGACACAAACTATGTCAAGACCGATTCGTCAGACTTAGGGCAGTTAAAAGTCTATGT
CTGGACTGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGTACTTAAGACTATTTCCAAATTTGAATGTTT
TTGTTAACAGTTATTCATACTTTAGGGGTTCATTAGTTTTAAGATTGAGCGTCTACGCCAGTACTTTCAATAGAGGTCGA
TTGAGGATGGGCTTCTTCCCCAATGCCACAACGGACTCAACTTCCACCTTGGACAATGCTATATACACAATATGTGATAT
AGGTAGTGACAATAGTTTTGAAATTACAATCCCTTACTCTTTTTCCACTTGGATGAGGAAAACAAATGGCCACCCAATTG
GATTGTTTCAAATTGAAGTCTTAAACAGGCTCACATACAACAGTTCTAGTCCATCGGAGGTTTATTGTATAGTCCAAGGT
AAAATGGGACAAGATGCCAGGTTCTTCTGCCCAACTGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGGA
TTTAACCGACCCTCTTTGTATTGAAGATGACACAGAAAATTGCANCAAACAATGTCTCCAAATGAACTAGGACTCACTTC
AGCCCAAGATGATGGCCCACTTGGTCAAGAAAAGCCAAATTATTTTCTCAATTTTAGGTCGATGAATGTGGACATTTTTA
CTGTATCACATACTAAAGTAGATAACCTATTTGGGCGGGNATGNNTTTTTATGGAGCATACTTTCACCNNGNNNNNANGN
NANTGCCATTGGAATTNCAAANNAGNCATGGGNCNNTCNCTNNNTTGCTNATTTTNCTGNNACTGANNNCNTNTNTGTNN
ANNGNNANGGNNNTCTGANNNGCNNNNNCNNNNNNNCNNNNAANNATNNNNNNATTNNNNNNTCNNNNNNNNANNNNTAC
NNNNNGNNANCANNNNNNCNNNANNNNCCNNNNNNNN
```

Sequence aligned with the original:

```
seq             NTCNNNTCNANCN--GGGTTCTTTTCCAACGCTATTGTCGAGGGAGCTAAGAAGTTCCCT
ori             ATCGAGTCCGACCCTGGGTTCTTTTCCAACGCTATTGTCGAGGGAGCTAAGAAGTTCCCT
                 **   **   *   ******************************************

seq             AACACCGAGTTCGTGAAGGTGAAGGGCCTCCACTTCAGCCAGGAGGACGCTCCAGATGAA
ori             AACACCGAGTTCGTGAAGGTGAAGGGCCTCCACTTCAGCCAGGAGGACGCTCCAGATGAA
                ************************************************************

seq             ATGGGTAAGTACATCAAGAGCTTCGTGGAGCGCGTGCTGAAGAACGAGCAGGCTAGCGGT
ori             ATGGGTAAGTACATCAAGAGCTTCGTGGAGCGCGTGCTGAAGAACGAGCAGGCTAGCGGT
                ************************************************************

seq             TCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGGAAACCACACAGGCTGACTTA
ori             TCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGGAAACCACACAGGCTGACTTA
                ************************************************************

seq             TGTATCCCCTATGTTGCTGACACAAACTATGTCAAGACCGATTCGTCAGACTTAGGGCAG
ori             TGTATCCCCTATGTTGCTGACACAAACTATGTCAAGACCGATTCGTCAGACTTAGGGCAG
                ************************************************************

seq             TTAAAAGTCTATGTCTGGACTGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAG
ori             TTAAAAGTCTATGTCTGGACTGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAG
                ************************************************************

seq             ATGTACTTAAGACTATTTCCAAATTTGAATGTTTTTGTTAACAGTTATTCATACTTTAGG
ori             ATGTACTTAAGACTATTTCCAAATTTGAATGTTTTTGTTAACAGTTATTCATACTTTAGG
                ************************************************************

seq             GGTTCATTAGTTTTAAGATTGAGCGTCTACGCCAGTACTTTCAATAGAGGTCGATTGAGG
ori             GGTTCATTAGTTTTAAGATTGAGCGTCTACGCCAGTACTTTCAATAGAGGTCGATTGAGG
                ************************************************************

seq             ATGGGCTTCTTCCCCAATGCCACAACGGACTCAACTTCCACCTTGGACAATGCTATATAC
ori             ATGGGCTTCTTCCCCAATGCCACAACGGACTCAACTTCCACCTTGGACAATGCTATATAC
                ************************************************************

seq             ACAATATGTGATATAGGTAGTGACAATAGTTTTGAAATTACAATCCCTTACTCTTTTTCC
ori             ACAATATGTGATATAGGTAGTGACAATAGTTTTGAAATTACAATCCCTTACTCTTTTTCC
                ************************************************************

seq             ACTTGGATGAGGAAAACAAATGGCCACCCAATTGGATTGTTTCAAATTGAAGTCTTAAAC
ori             ACTTGGATGAGGAAAACAAATGGCCACCCAATTGGATTGTTTCAAATTGAAGTCTTAAAC
                ************************************************************
```

```
seq             AGGCTCACATACAACAGTTCTAGTCCATCGGAGGTTTATTGTATAGTCCAAGGTAAAATG
ori             AGGCTCACATACAACAGTTCTAGTCCATCGGAGGTTTATTGTATAGTCCAAGGTAAAATG
                ************************************************************

seq             GGACAAGATGCCAGGTTCTTCTGCCCAACTGGTTCTGTTGTTACATTCCAGAATTCATGG
ori             GGACAAGATGCCAGGTTCTTCTGCCCAACTGGTTCTGTTGTTACATTCCAGAATTCATGG
                ************************************************************

seq             GGTTCACAGATGGATTTAACCGACCCTCTTTGTATTGAAGATGACACAGAAAATTGCAN-
ori             GGTTCACAGATGGATTTAACCGACCCTCTTTGTATTGAAGATGACACAGAAAATTGCAAA
                ************************************************************

seq             CAAACAATGTCTCCAAATGAACTAGGACTCACTTCAGCCCAAGATGATGGCCCACTTGGT
ori             CAAACAATGTCTCCAAATGAACTAGGACTCACTTCAGCCCAAGATGATGGCCCACTTGGT
                ************************************************************

seq             CAAGAAAAGCCAAATTATTTTCTCAATTTTAGGTCGATGAATGTGGACATTTTTACTGTA
ori             CAAGAAAAGCCAAATTATTTTCTCAATTTTAGGTCGATGAATGTGGACATTTTTACTGTA
                ************************************************************

seq             TCACATACTAAAGTAGATAACCTATTTGGGCGGGNATGNNTTTTTATGGAGCATACTTTC
ori             TCACATACTAAAGTAGATAACCTATTTGGGCGGGCATGGTTTTTTATGGAGCATACTTTC
                ********************************** ***  ********************

seq             ACCNNGNNNNNAN----GNNANTGCCATTGGAATTNCAAANN---AGNCATGGGNCNNTC
ori             ACCAATGAGGGACAATGGAGAGTGCCATTGGAATTTCCAAAACAAGGTCATGGGTCCTTA
                ***        *     *  * ************ * **      * ****** *  *
```

B. Original cre sequence in pHPeV1Luc using the reverse primer (Table 2.4)

```
NNNNNNNNNNNNNNGNNANAGCCCATCCTCATCGACCTCTATTGAAAGTACTGGCGTAGACGCTCAATCTTAAAACTAAT
GAACCCCTAAAGTATGAATAACTGTTAACAAAAACATTCAAATTTGGAAATAGTCTTAAGTACATCTGTGAACCCCATGA
ATTCTGGAATGTAACAACAGAACCAGTCCAGACATAGACTTTTAACTGCCCTAAGTCTGACGAATCGGTCTTGACATAGT
TTGTGTCAGCAACATAGGGGATACATAAGTCAGCCTGTGTGGTTTCCATCTGTGAACCCCATGAATTCTGGAATGTAACA
ACAGAACCGCTAGCCTGCTCGTTCTTCAGCACGCGCTCCACGAAGCTCTTGATGTACTTACCCATTTCATCTGGAGCGTC
CTCCTGGCTGAAGTGGAGGCCCTTCACCTTCACGAACTCGGTGTTAGGGAACTTCTTAGCTCCCTCGACAATAGCGTTGG
AAAAGAACCCAGGGTCGGACTCGATGAACATCTTAGGCAGATCGTCGCTGGCCCGAAGGTAGGCGTTGTAGTTGCGGACA
ATCTGGACGACGTCGGGCTTGCCTCCCTTAACGAGAGGGATCTCGCGAGGCCAGGAGAGGGTAGGCCGTCTAACCTCGCC
CTTCTCCTTGAATGGCTCCAGGTAGGCAGCGAACTCCTCAGGCTCCAGTTTCCGCATGATCTTGCTTGGGAGCATGGTCT
CGACGAAGAAGTTATTCTCAAGCACCATTTTCTCGCCCTCTTCGCTCTTGATCAGGGCGATATCCTCCTCGATGTCAGGC
CACTCGTCCCAGGACTCGATCACGTCCACGACACTCTCAGCATGGACGATGGCCTTGATCTTGTCTTGGNGCTCGTAGNA
GTAGTGAAAGGCCAGACAAGCCCCCCAGTCGTGGNCCACAAGATGATTTTCTTTGGAAGGTTCAGCAGCTCGAACCAAGC
GGTGANGTACTTGTAGTGATCCAGNAGGNGATATGAGCCATTCCCGCTCTTGCCGGACTTACCCATTCCGATCAGATCAN
GGNTGATGCATCTANCNACGGNNCGATNTGNNGNANNACNNNNNCNNNGNAGCTGNGNNNNNNNCNTGCANAAAAANCAN
GNNTNNNNGNNNNCTNCTCGNNNNNNNNNTANNNATNANNANNCNNNNNNNTNNNNNNNNTGNNNNNNNNNNNNNANNNNN
NNANNNNGNNNNNNNNNNCNNGNNNNNNNNNNCNNNNN
```

C. Reverse complemented cre sequence (from B) in pHPeV1Luc to show the forward sequence

```
NNNNNGNNNNNNNNNNCNNGNNNNNNNNNNCNNNNTNNNNNNNTNNNNNNNNNNNNNNCANNNNNNNNANNNNNNNGNNTNNTNATNN
NTANNNNNNNNCGAGNAGNNNNCNNNNANNCNTGNTTTTTNTGCANGNNNNNNNCNCAGCTNCNNNGNNNNNGTNNTNCNNCANATC
GNNCCGTNGNTAGATGCATCANCCNTGATCTGATCGGAATGGGTAAGTCCGGCAAGAGCGGGAATGGCTCATATCNCCTNCTGGATC
ACTACAAGTACNTCACCGCTTGGTTCGAGCTGCTGAACCTTCCAAAGAAAATCATCTTGTGGNCCACGACTGGGGGGCTTGTCTGGC
CTTTCACTACTNCTACGAGCNCCAAGACAAGATCAAGGCCATCGTCCATGCTGAGAGTGTCGTGGACGTGATCGAGTCCTGGGACGA
GTGGCCTGACATCGAGGAGGATATCGCCCTGATCAAGAGCGAAGAGGGCGAGAAAATGGTGCTTGAGAATAACTTCTTCGTCGAGAC
CATGCTCCCAAGCAAGATCATGCGGAAACTGGAGCCTGAGGAGTTCGCTGCCTACCTGGAGCCATTCAAGGAGAAGGGCGAGGTTAG
ACGGCCTACCCTCTCCTGGCCTCGCGAGATCCCTCTCGTTAAGGGAGGCAAGCCCGACGTCGTCCAGATTGTCCGCAACTACAACGC
CTACCTTCGGGCCAGCGACGATCTGCCTAAGATGTTCATCGAGTCCGACCCTGGGTTCTTTTCCAACGCTATTGTCGAGGGAGCTAA
GAAGTTCCCTAACACCGAGTTCGTGAAGGTGAAGGGCCTCCACTTCAGCCAGGAGGACGCTCCAGATGAAATGGGTAAGTACATCAA
GAGCTTCGTGGAGCGCGTGCTGAAGAACGAGCAGGCTAGCGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGGAAAC
CACACAGGCTGACTTATGTATCCCCTATGTTGCTGACACAAACTATGTCAAGACCGATTCGTCAGACTTAGGGCAGTTAAAAGTCTA
TGTCTGGACTGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGTACTTAAGACTATTTCCAAATTTGAATGTTTTTGT
```

TAACAGTTATTCATACTTTAGGGGTTCATTAGTTTTAAGATTGAGCGTCTACGCCAGTACTTTCAATAGAGGTCGATGAGGATGGGC
TNTNNCNNNNNNNNNNNNNN

Sequence aligned with the original:

```
seq             NANN--CNTGNTTTTTNTGCANGNNNNNNNCNC----AGCTNCNNN---GNNNNNGTNNT
ori             GAACGCCGTGATTTTTCTGCATGGTAACGCTGCCTCCAGCTACCTGTGGAGGCACGTCGT
                 *    * ** ***** **** *         *    **** *            **  *

seq             NCNNCANATCGNNCC-GTNGNTAGATGCATCANCCNTGATCTGATCGGAATGGGTAAGTC
ori             GCCTCACATCGAGCCCGTGGCTAGATGCATCATCCCTGATCTGATCGGAATGGGTAAGTC
                 *  ** ****  ** ** * ********** ** ************************

seq             CGGCAAGAGCGGGAATGGCTCATATCNCCTNCTGGATCACTACAAGTACNTCACCGCTTG
ori             CGGCAAGAGCGGGAATGGCTCATATCGCCTCCTGGATCACTACAAGTACCTCACCGCTTG
                ************************** *** ***************** **********

seq             GTTCGAGCTGCTGAACCTTCCAAAGAAAATCATCTT-GTGGNCCACGACTGGGGGGCTTG
ori             GTTCGAGCTGCTGAACCTTCCAAAGAAAATCATCTTTGTGGGCCACGACTGGGGGGCTTG
                ********************************** **** ******************

seq             TCTGGCCTTTCACTACTNCTACGAGCNCCAAGACAAGATCAAGGCCATCGTCCATGCTGA
ori             TCTGGCCTTTCACTACTCCTACGAGCACCAAGACAAGATCAAGGCCATCGTCCATGCTGA
                ***************** ******** ********************************

seq             GAGTGTCGTGGACGTGATCGAGTCCTGGGACGAGTGGCCTGACATCGAGGAGGATATCGC
ori             GAGTGTCGTGGACGTGATCGAGTCCTGGGACGAGTGGCCTGACATCGAGGAGGATATCGC
                ************************************************************

seq             CCTGATCAAGAGCGAAGAGGGCGAGAAAATGGTGCTTGAGAATAACTTCTTCGTCGAGAC
ori             CCTGATCAAGAGCGAAGAGGGCGAGAAAATGGTGCTTGAGAATAACTTCTTCGTCGAGAC
                ************************************************************

seq             CATGCTCCCAAGCAAGATCATGCGGAAACTGGAGCCTGAGGAGTTCGCTGCCTACCTGGA
ori             CATGCTCCCAAGCAAGATCATGCGGAAACTGGAGCCTGAGGAGTTCGCTGCCTACCTGGA
                ************************************************************

seq             GCCATTCAAGGAGAAGGGCGAGGTTAGACGGCCTACCCTCTCCTGGCCTCGCGAGATCCC
ori             GCCATTCAAGGAGAAGGGCGAGGTTAGACGGCCTACCCTCTCCTGGCCTCGCGAGATCCC
                ************************************************************

seq             TCTCGTTAAGGGAGGCAAGCCCGACGTCGTCCAGATTGTCCGCAACTACAACGCCTACCT
ori             TCTCGTTAAGGGAGGCAAGCCCGACGTCGTCCAGATTGTCCGCAACTACAACGCCTACCT
                ************************************************************

seq             TCGGGCCAGCGACGATCTGCCTAAGATGTTCATCGAGTCCGACCCTGGGTTCTTTTCCAA
ori             TCGGGCCAGCGACGATCTGCCTAAGATGTTCATCGAGTCCGACCCTGGGTTCTTTTCCAA
                ************************************************************

seq             CGCTATTGTCGAGGGAGCTAAGAAGTTCCCTAACACCGAGTTCGTGAAGGTGAAGGGCCT
ori             CGCTATTGTCGAGGGAGCTAAGAAGTTCCCTAACACCGAGTTCGTGAAGGTGAAGGGCCT
                ************************************************************

seq             CCACTTCAGCCAGGAGGACGCTCCAGATGAAATGGGTAAGTACATCAAGAGCTTCGTGGA
ori             CCACTTCAGCCAGGAGGACGCTCCAGATGAAATGGGTAAGTACATCAAGAGCTTCGTGGA
                ************************************************************

seq             GCGCGTGCTGAAGAACGAGCAGGCTAGCGGTTCTGTTGTTACATTCCAGAATTCATGGGG
ori             GCGCGTGCTGAAGAACGAGCAGGCTAGCGGTTCTGTTGTTACATTCCAGAATTCATGGGG
                ************************************************************
```

```
seq             TTCACAGATGGAAACCACACAGGCTGACTTATGTATCCCCTATGTTGCTGACACAAACTA
ori             TTCACAGATGGAAACCACACAGGCTGACTTATGTATCCCCTATGTTGCTGACACAAACTA
                ************************************************************

seq             TGTCAAGACCGATTCGTCAGACTTAGGGCAGTTAAAAGTCTATGTCTGGACTGGTTCTGT
ori             TGTCAAGACCGATTCGTCAGACTTAGGGCAGTTAAAAGTCTATGTCTGGACTGGTTCTGT
                ************************************************************

seq             TGTTACATTCCAGAATTCATGGGGTTCACAGATGTACTTAAGACTATTTCCAAATTTGAA
ori             TGTTACATTCCAGAATTCATGGGGTTCACAGATGTACTTAAGACTATTTCCAAATTTGAA
                ************************************************************

seq             TGTTTTTGTTAACAGTTATTCATACTTTAGGGGTTCATTAGTTTTAAGATTGAGCGTCTA
ori             TGTTTTTGTTAACAGTTATTCATACTTTAGGGGTTCATTAGTTTTAAGATTGAGCGTCTA
                ************************************************************

seq             CGCCAGTACTTTCAATAGAGGTCGAT-GAGGATGGGCTNTNNCNNNNNNNNNNNNNN---
ori             CGCCAGTACTTTCAATAGAGGTCGATTGAGGATGGGCTTCTTCCCCAATGCCACAACGGA
                ************************* ***********    *

seq             ------------------------------------------------------------
ori             CTCAACTTCCACCTTGGACAATGCT
```

Figure 3: Desired sequence of the pHPeV1Luc construct (ori) compared to the obtained results sequenced A) using a forward primer and B) a reverse primer (Table 2.4). C) Shows the reverse complement sequence of B) which gives the sequence read in the 5'-3' orientation. The restriction enzymes NheI and AflII (Table 5.1) are highlighted in red.

A. SacI.5' UTR + Luc.NheI in pGemt Easy sequenced with T7 forward primer

```
NNNNNNNNNNANGCTCCGGCCGCCATGGCGGCCGCGGGAATTCGATTGAGCTCTTTGAAAGGGGTCTCCTAGAGAGCTTG
GCCGTCGGGCCTTATACCCCGACTTGCTGAGCTTCTCTAGGAGAGTCCCTTTCCCAGCCTTGGGGTGGCTGGTCAATAAA
AACCCCATATGTAACCAACACCTAAGACAATTTGGTCAACCCTATGCCTGGTCCCCACTATTCGAAGGCAACTTGCAATA
AGAAGAGTGGAACAAGGATGCTTAAAGCATAGTGTAAATGATCTTTTCTAACCTGTATTATGTACAGGGTGGCAGATGGC
GTGCCATAAATCTATTAGTGGGATACCACGCTTGTGGACCTTATGCCCACACAGCCATCCTCTAGTAAGTTTGTAAAATG
TCTGGTGAGATGTGGGAACTTATTGGAAACAACAATTTGCTTAATAGCATCCTAGTGCCAGCGGAACAACATCTGGTAAC
AGATGCCTCTGGGGCCAAAAGCCAAGGTTTGACAGACCCATTAGGATTGGTTTCAAAACCTGAATTGTTGTGGAAGATAT
TCAGTACCTATCAATCTGGTAGTGGTGCAAACACTAGTTGTAAGGCCCACGAAGGATGCCCAGAAGGTACCCGCAGGTAA
CAAGAGACACTGTGGATCTGAGGGCCAACTACCTCTATCAGGTGAGTTAGTTAAAAAACGTCTAGTGGGCCAAACC
CAGGGGGGATCCCTGGTTTCCTTTTATTGTTAATATTGACATTATGGCTTCCAAGGTGTACGACCCCGAGCAACGCAACG
CATGATCACTGGGCCTCAGTGGTGGGCTCGCTGCAAGCAAATGAACGTGCTGGACTCCTTCATCAACTACTATGATTCCG
AGAAGCACGCCGAGAACGCCGTGATTTTTCTGCATGGTAACGCTGCCTCCAGCTACCCTGTGGAGGCACGTCGNGCNNCA
CATCGAGCCCGNNGNTAGATGCATCATCCCTGATCTGATCGNAATGGGTAAGTCCGGNAGANCGGNNNGGCTCATATCGC
CTNCNNNNCACTACAGTACNCNCGCTNNNNNNNCTGCTGNNNNNNANAAATCATCNTTGNGGNNNCGACNGGNGGNNNNN
CTNNNNNNNNNANTCNNCNANNNNNNANNANATNANNNTCNNCANNCTNANNNNNNNNNNNGGNANNTGNNATNNNNANA
NNTNNNNNNNNNNNNCCNGNANNN
```

Sequence aligned with the original:

```
ori             ------------------------------------------------GAGCTCTTTGAAA
seq             NNNNNNNNNNANGCTCCGGCCGCCATGGCGGCCGCGGGAATTCGATTGAGCTCTTTGAAA
                                                               *************

ori             GGGGTCTCCTAGAGAGCTTGGCCGTCGGGCCTTATACCCCGACTTGCTGAGCTTCTCTAG
seq             GGGGTCTCCTAGAGAGCTTGGCCGTCGGGCCTTATACCCCGACTTGCTGAGCTTCTCTAG
                ************************************************************

ori             GAGAGTCCCTTTCCCAGCCTTGGGGTGGCTGGTCAATAAAAACCCCATATGTAACCAACA
seq             GAGAGTCCCTTTCCCAGCCTTGGGGTGGCTGGTCAATAAAAACCCCATATGTAACCAACA
                ************************************************************

ori             CCTAAGACAATTTGGTCAACCCTATGCCTGGTCCCCACTATTCGAAGGCAACTTGCAATA
seq             CCTAAGACAATTTGGTCAACCCTATGCCTGGTCCCCACTATTCGAAGGCAACTTGCAATA
                ************************************************************

ori             AGAAGAGTGGAACAAGGATGCTTAAAGCATAGTGTAAATGATCTTTTCTAACCTGTATTA
seq             AGAAGAGTGGAACAAGGATGCTTAAAGCATAGTGTAAATGATCTTTTCTAACCTGTATTA
                ************************************************************

ori             TGTACAGGGTGGCAGATGGCGTGCCATAAATCTATTAGTGGGATACCACGCTTGTGGACC
seq             TGTACAGGGTGGCAGATGGCGTGCCATAAATCTATTAGTGGGATACCACGCTTGTGGACC
                ************************************************************

ori             TTATGCCCACACAGCCATCCTCTAGTAAGTTTGTAAAATGTCTGGTGAGATGTGGGAACT
seq             TTATGCCCACACAGCCATCCTCTAGTAAGTTTGTAAAATGTCTGGTGAGATGTGGGAACT
                ************************************************************

ori             TATTGGAAACAACAATTTGCTTAATAGCATCCTAGTGCCAGCGGAACAACATCTGGTAAC
seq             TATTGGAAACAACAATTTGCTTAATAGCATCCTAGTGCCAGCGGAACAACATCTGGTAAC
                ************************************************************

ori             AGATGCCTCTGGGGCCAAAAGCCAAGGTTTGACAGACCCATTAGGATTGGTTTCAAAACC
seq             AGATGCCTCTGGGGCCAAAAGCCAAGGTTTGACAGACCCATTAGGATTGGTTTCAAAACC
                ************************************************************

ori             TGAATTGTTGTGGAAGATATTCAGTACCTATCAATCTGGTAGTGGTGCAAACACTAGTTG
seq             TGAATTGTTGTGGAAGATATTCAGTACCTATCAATCTGGTAGTGGTGCAAACACTAGTTG
                ************************************************************

ori             TAAGGCCCACGAAGGATGCCCAGAAGGTACCCGCAGGTAACAAGAGACACTGTGGATCTG
seq             TAAGGCCCACGAAGGATGCCCAGAAGGTACCCGCAGGTAACAAGAGACACTGTGGATCTG
                ************************************************************
```

```
ori             ATCTGGGGCCAACTACCTCTATCAGGTGAGTTAGTTAAAAAACGTCTAGTGGGCCAAACC
seq             ATCTGGGGCCAACTACCTCTATCAGGTGAGTTAGTTAAAAAACGTCTAGTGGGCCAAACC
                ************************************************************

ori             CAGGGGGGATCCCTGGTTTCCTTTTATTGTTAATATTGACATTATGGCTTCCAAGGTGTA
seq             CAGGGGGGATCCCTGGTTTCCTTTTATTGTTAATATTGACATTATGGCTTCCAAGGTGTA
                ************************************************************

ori             CGACCCCGAGCAACGCAAACGCATGATCACTGGGCCTCAGTGGTGGGCTCGCTGCAAGCA
seq             CGACCCCGAGCAACGCAA-CGCATGATCACTGGGCCTCAGTGGTGGGCTCGCTGCAAGCA
                ****************** ****************************************

ori             AATGAACGTGCTGGACTCCTTCATCAACTACTATGATTCCGAGAAGCACGCCGAGAACGC
seq             AATGAACGTGCTGGACTCCTTCATCAACTACTATGATTCCGAGAAGCACGCCGAGAACGC
                ************************************************************

ori             CGTGATTTTTCTGCATGGTAACGCTGCCTCCAGCTACC-TGTGGAGGCACGTCGTGCCTC
seq             CGTGATTTTTCTGCATGGTAACGCTGCCTCCAGCTACCCTGTGGAGGCACGTCGNGCNNC
                ************************************ ************** **  *

ori             ACATCGAGCCCGTGGCTAGATGCATCATCCCTGATCTGATCGGAATGGGTAAGTCCGGCA
seq             ACATCGAGCCCGNNGNTAGATGCATCATCCCTGATCTGATCGNAATGGGTAAGTCCGGNA
                ************  * ************************ *************** *

ori             AGAGCGGGAATGGCTCATATCGCCTCCTGGATCACTACAAGTACCTCACCGCTTGGTTCG
seq             -GANCGGNNN-GGCTCATATCGCCTNCNNNN-CACTACA-GTACNC--NCGCTNNNNNNN
                 ** ***    ************** *     ******* ****     ****

ori             AGCTGCTGAACCTTCCAAAGAAAATCATCTTTGTGGGCCACGACTGGGGGGCTTGTCTGG
seq             --CTGCTGNNNNNN-----ANAAATCATCNTTGNGGNNN-CGACNGGNGGNNNNNCTNNN
                  ******             ******** *** **    **** ** **
```

B. SacI.5’ UTR + Luc.NheI in pGemt Easy sequence with SP6 primer

```
NNNNNNNNNCANNCGTTGGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGCGAATTCACTAGTGATCAGAACCGCTA
GCCTGCTCGTTCTTCAGCACGCGCTCCACGAAGCTCTTGATGTACTTACCCATTTCATCTGGAGCGTCCTCCTGGCTGAA
GTGGAGGCCCTTCACCTTCACGAACTCGGTGTTAGGGAACTTCTTAGCTCCCTCGACAATAGCGTTGGAAAAGAACCCAG
GGTCGGACTCGATGAACATCTTAGGCAGATCGTCGCTGGCCCGAAGGTAGGCGTTGTAGTTGCGGACAATCTGGACGACG
TCGGGCTTGCCTCCCTTAACGAGAGGGATCTCGCGAGGCCAGGAGAGGGTAGGCCGTCTAACCTCGCCCTTCTCCTTGAA
TGGCTCCAGGTAGGCAGCGAACTCCTCAGGCTCCAGTTTCCGCATGATCTTGCTTGGGAGCATGGTCTCGACGAAGAAGT
TATTCTCAAGCACCATTTTCTCGCCCTCTTCGCTCTTGATCAGGGCGATATCCTCCTCGATGTCAGGCCACTCGTCCCAG
GACTCGATCACGTCCACGACACTCTCAGCATGGACGATGGCCTTGATCTTGTCTTGGTGCTCGTAGGAGTAGTGAAAGGC
CAGACAAGCCCCCCAGTCGTGGCCCACAAAGATGATTTTCTTTGGAAGGTTCAGCAGCTCGAACCAAGCGGTGAGGTACT
TGTAGTGATCCAGGAGGCGATATGAGCCATTCCCGCTCTTGCCGGACTTACCCATTCCGATCAGATCAGGGATGATGCAT
CTAGCCACGGGCTCGATGTGAGGCACGACGTGCCTCCACAGGTAGCTGGANGCAGCGTTACCATGCAGAAAAATCACGGC
GTTCTCGGCGTGCTTCTCGGAATCATAGTAGTTGATGAAAGGAGTCCAGCACGTTCATTTGCTTGCAGCGAGCCCACCAC
TGAGGCCCAGTGATCATGCGTTTGCGTTGCTCGGGGNCGTACACCTTGNNNGCCATAATGTCANNNTNNNNAAAAGAAAN
CAGGNTCCCCCCCNGGGNTNGNCCNACTANACNTTTTTANTANNNCNCNGATANAGNANNNGNCCCNGATNNNANNCNNC
ANNNNNNNNNNNNNNCNGNNNNNNNNNNGGNNATNNNNNGNNNNNNACNNNNNNNNNNNNNNNCNNANNNNNNNNNACNNN
ANNNNNNNTTCCNNNN
```

SacI.5’ UTR + Luc.NheI in pGemt Easy seq with SP6 primer (reverse complemented)

```
NNNNGGAANNNNNNNTNNNGTNNNNNNNNNTNNGNNNNNNNNNNNNNNNNGTNNNNNNCNNNNNATNNCCNNNNNNNNNNCNGNNNNNNNNNNNNNNTGN
NGNNTNNNATCNGGGNCNNNTNCTNTATCNGNGNNNTANTAAAAANGTNTAGTNGGNCNANCCCNGGGGGGGGANCCTGNTTTCTTTTNNNNANNNTGA
CATTATGGCNNNCAAGGTGTACGNCCCCGAGCAACGCAAACGCATGATCACTGGGCCTCAGTGGTGGGCTCGCTGCAAGCAAATGAACGTGCTGGACT
CCTTTCATCAACTACTATGATTCCGAGAAGCACGCCGAGAACGCCGTGATTTTTCTGCATGGTAACGCTGCNTCCAGCTACCTGTGGAGGCACGTCGT
GCCTCACATCGAGCCCGTGGCTAGATGCATCATCCCTGATCTGATCGGAATGGGTAAGTCCGGCAAGAGCGGGAATGGCTCATATCGCCTCCTGGATC
ACTACAAGTACCTCACCGCTTGGTTCGAGCTGCTGAACCTTCCAAAGAAAATCATCTTTGTGGGCCACGACTGGGGGGCTTGTCTGGCCTTTCACTAC
TCCTACGAGCACCAAGACAAGATCAAGGCCATCGTCCATGCTGAGAGTGTCGTGGACGTGATCGAGTCCTGGGACGAGTGGCCTGACATCGAGGAGGA
TATCGCCCTGATCAAGAGCGAAGAGGGCGAGAAAATGGTGCTTGAGAATAACTTCTTCGTCGAGACCATGCTCCCAAGCAAGATCATGCGGAAACTGG
AGCCTGAGGAGTTCGCTGCCTACCTGGAGCCATTCAAGGAGAAGGGCGAGGTTAGACGGCCTACCCTCTCCTGGCCTCGCGAGATCCCTCTCGTTAAG
GGAGGCAAGCCCGACGTCGTCCAGATTGTCCGCAACTACAACGCCTACCTTCGGGCCAGCGACGATCTGCCTAAGATGTTCATCGAGTCCGACCCTGG
GTTCTTTTCCAACGCTATTGTCGAGGGAGCTAAGAAGTTCCCTAACACCGAGTTCGTGAAGGTGAAGGGCCTCCACTTCAGCCAGGAGGACGCTCCAG
ATGAAATGGGTAAGTACATCAAGAGCTTCGTGGAGCGCGTGCTGAAGAACGAGCAGGCTAGCGGTTCTGATCACTAGTGAATTCGCGGCCGCCTGCAG
GTCGACCATATGGGAGAGCTCCCAACGNNTGNNNNNNNNN
```

Sequence aligned with the original:

```
ori             ACCCCGAGCAACGCAAACGCATGATCACTGGGCCTCAGTGGTGGG
seq             NCCCCGAGCAACGCAAACGCATGATCACTGGGCCTCAGTGGTGGG
                 ******************************************

ori             CTCGCTGCAAGCAAATGAACGTGCTGGACTCCTT-CATCAACTACTATGATTCCGAGAAG
seq             CTCGCTGCAAGCAAATGAACGTGCTGGACTCCTTTCATCAACTACTATGATTCCGAGAAG
                ********************************** ************************

ori             CACGCCGAGAACGCCGTGATTTTTCTGCATGGTAACGCTGCCTCCAGCTACCTGTGGAGG
seq             CACGCCGAGAACGCCGTGATTTTTCTGCATGGTAACGCTGCNTCCAGCTACCTGTGGAGG
                ************************************* ******************

ori             CACGTCGTGCCTCACATCGAGCCCGTGGCTAGATGCATCATCCCTGATCTGATCGGAATG
seq             CACGTCGTGCCTCACATCGAGCCCGTGGCTAGATGCATCATCCCTGATCTGATCGGAATG
                ************************************************************

ori             GGTAAGTCCGGCAAGAGCGGGAATGGCTCATATCGCCTCCTGGATCACTACAAGTACCTC
seq             GGTAAGTCCGGCAAGAGCGGGAATGGCTCATATCGCCTCCTGGATCACTACAAGTACCTC
                ************************************************************

ori             ACCGCTTGGTTCGAGCTGCTGAACCTTCCAAAGAAAATCATCTTTGTGGGCCACGACTGG
seq             ACCGCTTGGTTCGAGCTGCTGAACCTTCCAAAGAAAATCATCTTTGTGGGCCACGACTGG
                ************************************************************

ori             GGGGCTTGTCTGGCCTTTCACTACTCCTACGAGCACCAAGACAAGATCAAGGCCATCGTC
seq             GGGGCTTGTCTGGCCTTTCACTACTCCTACGAGCACCAAGACAAGATCAAGGCCATCGTC
                ************************************************************

ori             CATGCTGAGAGTGTCGTGGACGTGATCGAGTCCTGGGACGAGTGGCCTGACATCGAGGAG
seq             CATGCTGAGAGTGTCGTGGACGTGATCGAGTCCTGGGACGAGTGGCCTGACATCGAGGAG
                ************************************************************

ori             GATATCGCCCTGATCAAGAGCGAAGAGGGCGAGAAAATGGTGCTTGAGAATAACTTCTTC
seq             GATATCGCCCTGATCAAGAGCGAAGAGGGCGAGAAAATGGTGCTTGAGAATAACTTCTTC
                ************************************************************

ori             GTCGAGACCATGCTCCCAAGCAAGATCATGCGGAAACTGGAGCCTGAGGAGTTCGCTGCC
seq             GTCGAGACCATGCTCCCAAGCAAGATCATGCGGAAACTGGAGCCTGAGGAGTTCGCTGCC
                ************************************************************

ori             TACCTGGAGCCATTCAAGGAGAAGGGCGAGGTTAGACGGCCTACCCTCTCCTGGCCTCGC
seq             TACCTGGAGCCATTCAAGGAGAAGGGCGAGGTTAGACGGCCTACCCTCTCCTGGCCTCGC
                ************************************************************

ori             GAGATCCCTCTCGTTAAGGGAGGCAAGCCCGACGTCGTCCAGATTGTCCGCAACTACAAC
seq             GAGATCCCTCTCGTTAAGGGAGGCAAGCCCGACGTCGTCCAGATTGTCCGCAACTACAAC
                ************************************************************

ori             GCCTACCTTCGGGCCAGCGACGATCTGCCTAAGATGTTCATCGAGTCCGACCCTGGGTTC
seq             GCCTACCTTCGGGCCAGCGACGATCTGCCTAAGATGTTCATCGAGTCCGACCCTGGGTTC
                ************************************************************

ori             TTTTCCAACGCTATTGTCGAGGGAGCTAAGAAGTTCCCTAACACCGAGTTCGTGAAGGTG
seq             TTTTCCAACGCTATTGTCGAGGGAGCTAAGAAGTTCCCTAACACCGAGTTCGTGAAGGTG
                ************************************************************

ori             AAGGGCCTCCACTTCAGCCAGGAGGACGCTCCAGATGAAATGGGTAAGTACATCAAGAGC
seq             AAGGGCCTCCACTTCAGCCAGGAGGACGCTCCAGATGAAATGGGTAAGTACATCAAGAGC
                ************************************************************
```

```
ori             TTCGTGGAGCGCGTGCTGAAGAACGAGCAGGCTAGCGGTTCTGTTGTTACATTCCAGAAT
seq             TTCGTGGAGCGCGTGCTGAAGAACGAGCAGGCTAGCGGTTCTGAT----CACTAGTGAAT
                **************************************** *    ** *   ****
```

Figure 4: Desired sequence of SacI.5' UTR + Luc.NheI (ori) in pHPeV1Luc sequenced A) using a forward primer and B) a reverse primer (Table 2.4). C) Shows the reverse complement sequence of B) which gives the sequence read in the 5'-3' orientation hence giving the complete sequence.

pHPeV1Luc Sequenced with 5UTR Reverse primer

```
NNNNNNNNNNNNCNTCGANAGTGGGGACCAGGCATAGGGTTGACCAAATTGTCTTAGGTGTTGGTTACATATGGGGTTTT
TATTGACCAGCCACCCCAAGGCTGGGAAAGGGACTCTCCTAGAGAAGCTCAGCAAGTCGGGGTATAAGGCCCGACGGCCA
AGCTCTCTAGGAGACCCCTTTCAAAGAGCTCCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTA
CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAA
TAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTATTTTC
TCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGC
CAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGT
GACCGTCTCCGGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATAC
GCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGA
ACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATA
ATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTG
TTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTG
GNTCTCAACAGCGGTANNATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCANGATGAGCACTTTTNAAGTTCTGCTA
TGTGGCGNNGNATTATCCCGTATTGANNNCNGNNAGANCANTNGNNNNNCATACNNNATTNTCANANGANNNNNNNNTAC
NNNNCNGTCNNNNNAAAANNNTNTNNGNNNGNNNNNNNGNNNNNNNNNTGCNNNNNTNCNNNNNNNNNNNNANNNNNNNN
GCNNTNNTNNNNNNNNNATCGNNNNNNNNNNNNNACNNNNNNNNNNNNNNNNNNNANNG
```

pHPeV1Luc Sequenced with 5UTR Reverse primer (reverse complemented to show forward sequence)

```
CNNTNNNNNNNNNNNNNNNNNGTNNNNNNNNNNNNNCGATNNNNNNNNNANNANNGCNNNNNNNNTNNNNNNNNNNNGNANNNNNGCANNNNNNN
NNCNNNNNNNCNNNCNNANANNNTTTTNNNNNGACNGNNNNGTANNNNNNNNTCNTNTGANAATNNNGTATGNNNNNCNANTGNTCTNNCNGNNNTCA
ATACGGGATAATNCNNCGCCACATAGCAGAACTTNAAAAGTGCTCATCNTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATNNTACCGCTGTTG
AGANCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAA
TGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCA
TGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATT
ATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCT
CCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTA
ACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCAT
TCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATT
AAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTGGAGCTCTTTG
AAAGGGGTCTCCTAGAGAGCTTGGCCGTCGGGCCTTATACCCCGACTTGCTGAGCTTCTCTAGGAGAGTCCCTTTCCCAGCCTTGGGGTGGCTGGTCA
ATAAAAACCCCATATGTAACCAACACCTAAGACAATTTGGTCAACCCTATGCCTGGTCCCCACTNTCGANGNNNNNNNNNNN
```

Sequence aligned with the original:

```
seq             TTTGAAAGGGGTCTCCTAGAGAGCTTGGCCGTCGGGCCTTATACCCCGACTTGCTGAGCT
ori             TTTGAAAGGGGTCTCCTAGAGAGCTTGGCCGTCGGGCCTTATACCCCGACTTGCTGAGCT
                ************************************************************

seq             TCTCTAGGAGAGTCCCTTTCCCAGCCTTGGGGTGGCTGGTCAATAAAAACCCCATATGTA
ori             TCTCTAGGAGAGTCCCTTTCCCAGCCTTGGGGTGGCTGGTCAATAAAAACCCCATATGTA
                ************************************************************

seq             ACCAACACCTAAGACAATTTGGTCAACCCTATGCCTGGTCCCCACTNT-CGA--------
ori             ACCAACACCTAAGACAATTTGGTCAACCCTATGCCTGGTCCCCACTATTCGAAGGCAACT
                ******************************************* * ***
```

Figure 5: Desired sequence (ori) of pHPeV1Luc sequenced A) using a reverse primer (Table 2.4). B) Shows the reverse complement sequence of A) which gives the sequence read in the 5'-3' orientation hence giving the complete sequence. The green highlight shows the start of the 5' UTR.

A. pHPeV1LucLoop*cre*Mut

```
NNNNNNNNNNCNNNNCGACCCTGGGTTCTTTTCCACGCTATTGTCGAGGGAGCTAAGAAGTTCCCTAACACCGAGTTCGT
GAAGGTGAAGGGCCTCCACTTCAGCCAGGAGGACGCTCCAGATGAAATGGGTAAGTACATCAAGAGCTTCGTGGAGCGCG
TGCTGAAGAACGAGCAGGCTAGCGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGGAAACCACACAGGCT
GACTTATGTATCCCCTATGTTGCTGACACTTTCTATGTCAAGACCGATTCGTCAGACTTAGGGCAGTTAAAAGTCTATGT
CTGGACTGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGTACTTAAGACTATTTCCAAATTTGAATGTTT
TTGTTAACAGTTATTCATACTTTAGGGGTTCATTAGTTTTAAGATTGAGCGTCTACGCCAGTACTTTCAATAGAGGTCGA
TTGAGGATGGGCTTCTTCCCCAATGCCACAACGGACTCAACTTCCACCTTGGACAATGCTATATACACAATATGTGATAT
AGGTAGTGACAATAGTTTTGAAATTACAATCCCTTACTCTTTTTCCACTTGGATGAGGAAAACAAATGGCCACCCAATTG
GATTGTTTCAAATTGAAGTCTTAAACAGGCTCACATACAACAGTTCTAGTCCATCGGAGGTTTATTGTATAGTCCAAGGT
AAAATGGGACAAGATGCCAGGTTCTTCTGCCCAACTGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGGA
TTTAACCGACCCTCTTTGTATTGAAGATGACACAGAAAATTGCAAACAAACAATGTCTCCAAATGAACTAGGACTCACTT
CAGCCCAAGATGATGGCCCACTTGGTCAAGAAAAGCCAAATTATTTTCTCAATTTTAGGTCGATGAATGTGGACATTTTT
ACTGTATCACATACTAAAGTAGATAACCTATTTGGGCGGGCATGGNTTTTTNTGGAGCATACTTTCACCNATGNNNACNA
TGNNAGTGCCATTGGAATTTNCAAAANNAGGTCATGGGTCTNATCACTNNNNTTGNTNATTTTNCTNNNACTNANATNCA
NNNNNNNNCTANNGNNNNGGNNNTNCTGANGNNNGCNCNNNNCNNNNNNNNTANNNATGNNTCNANNNNNNNNNNNTCN
NNNNNNNNNNANNNNNNCCNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNANNNANTTN
```

Sequence aligned with the original:

```
ori             ------------------------------------------------------------CAGGCT
seq             GATGAAATGGGTAAGTACATCAAGAGCTTCGTGGAGCGCGTGCTGAAGAACGAGCAGGCT
                                                                            ******

ori             AGCGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGGAAACCACACAGGCT
seq             AGCGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGGAAACCACACAGGCT
                ************************************************************

ori             GACTTATGTATCCCCTATGTTGCTGACACAAACTATGTCAAGACCGATTCGTCAGACTTA
seq             GACTTATGTATCCCCTATGTTGCTGACACTTTCTATGTCAAGACCGATTCGTCAGACTTA
                ******************************   ****************************

ori             GGGCAGTTAAAAGTCTATGTCTGGACTGGTTCTGTTGTTACATTCCAGAATTCATGGGGT
seq             GGGCAGTTAAAAGTCTATGTCTGGACTGGTTCTGTTGTTACATTCCAGAATTCATGGGGT
                ************************************************************

ori             TCACAGATGTACTTAAGACTAT---------------------------------------
seq             TCACAGATGTACTTAAGACTATTTCCAAATTTGAATGTTTTTGTTAACAGTTATTCATAC
                **********************
```

B. pHPeV1LucRScreMut

```
NNNNNNNNNTCNNNNCGACCCTGGGTTCTTTTCCACGCTATTGTCGAGGGAGCTAAGAAGTTCCCTAACACCGAGTTCGT
GAAGGTGAAGGGCCTCCACTTCAGCCAGGAGGACGCTCCAGATGAAATGGGTAAGTACATCAAGAGCTTCGTGGAGCGCG
TGCTGAAGAACGAGCAGGCTAGCGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGGAAACCACACAGGCT
GACTTATGTATCCCCTATGTTGCTGACACAAACTATGTCAAGACCGATTCATCCGACTTAGGGCAGTTAAAAGTCTATGT
CTGGACTGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGTACTTAAGACTATTTCCAAATTTGAATGTTT
TTGTTAACAGTTATTCATACTTTAGGGGTTCATTAGTTTTAAGATTGAGCGTCTACGCCAGTACTTTCAATAGAGGTCGA
TTGAGGATGGGCTTCTTCCCCAATGCCACAACGGACTCAACTTCCACCTTGGACAATGCTATATACACAATATGTGATAT
AGGTAGTGACAATAGTTTTGAAATTACAATCCCTTACTCTTTTTCCACTTGGATGAGGAAAACAAATGGCCACCCAATTG
GATTGTTTCAAATTGAAGTCTTAAACAGGCTCACATACAACAGTTCTAGTCCATCGGANGTTTATTGTATAGTCCAAGGT
AAAATGGGACAAGATGCCAGGTTCTTCTGCCCAACTGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGGA
TTTAACCGACCCTCTTTGTATTGAAGATGACACAGAAAATTGCAAACAAACAATGTCTCCAAATGAACTAGGACTCACTT
CAGCCCAAGATGATGGCCCACTTGGTCAAGAAAAGCCAAATTATTTTCTCAATTTTAGGTCGATGAATGNGGACATTTTT
ACTGTATCACATACTAAAGTAGATAANCTATTTGGGCGGGCATGNNTTTTTATGGAGCATACTTTCACCNATGAGGGACN
ATGGNNANNGCCATTGGAATTTNCAAAANNNNCATGGGNNNNTCNCTNNNTNNCTNATTTTACTGNNACTNANNNCANNT
NNNNNNANNGNNAGGGGNNTNNNGANNNNCNNNNNNNNNNNNNNCNANNNANGNNTCNNNNNAATTTTNNNNNNTCNNNNN
```

Sequence aligned with the original:

```
ori             ----------------------------------------------------------CAGGCT
seq             GATGAAATGGGTAAGTACATCAAGAGCTTCGTGGAGCGCGTGCTGAAGAACGAGCAGGCT
                                                                        ******

ori             AGCGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGGAAACCACACAGGCT
seq             AGCGGTTCTGTTGTTACATTCCAGAATTCATGGGGTTCACAGATGGAAACCACACAGGCT
                ************************************************************

ori             GACTTATGTATCCCCTATGTTGCTGACACAAACTATGTCAAGACCGATTCGTCAGACTTA
seq             GACTTATGTATCCCCTATGTTGCTGACACAAACTATGTCAAGACCGATTCATCGGACTTA
                ************************************************** ** ******

ori             GGGCAGTTAAAAGTCTATGTCTGGACTGGTTCTGTTGTTACATTCCAGAATTCATGGGGT
seq             GGGCAGTTAAAAGTCTATGTCTGGACTGGTTCTGTTGTTACATTCCAGAATTCATGGGGT
                ************************************************************

ori             TCACAGATGTACTTAAGACTAT-------------------------------------
seq             TCACAGATGTACTTAAGACTATTTCCAAATTTGAATGTTTTTGTTAACAGTTATTCATAC
                **********************
```

Figure 6: Desired sequence (ori) of A) pHPeV1LucLoop*cre*Mut and B) pHPeV1LucRS*cre*Mut in pHPeV1Luc. The red highlight shows the restriction enzymes NheI and AflII respectively while the pink highlight shows the mutations in the loop (A) and the right stem (B) in the cre.

Appendix C

Production of the 21 mutations and YGDD mutations in pHPeV1Luc

pHPeV1(21mut)Luc sequenced through Mph1103I and to middle of the 3D^{pol}-encoding region of HPeV1

```
NNNNNNNNNNNNTNANNACAAGACTATCAGTTATACTGTTAAATCTTGCAAAGGAATGTGTGGTGGCCTACTTATTTCAA
AAGTAGAAGGTAACTTCAAAATCCTGGGTATGCATATTGCTGGTAATGGTGAAATGGGAGTAGCTATACCCTTTAATTTT
CTTAAAAATGACATGTCTGATCAAGGCATTGTTACTGAGATCACTCCAATCCAGCCCATGTATATAAACACTAAAACTCA
GATCCACAAGAGCCCTGTCTATGGTGCTGTTGAGGTGAAAATGGGTCCAGCTGTCCTAAGTAAGTCAGACACCAGGTTGG
AGGAACCTGTAGAATGTCTTATTAAAAAATCGGCTTCCAAGTACAGAGTTAACAAATTCCAGGTGAACAATGAACTCTGG
CAAGGTGTCAAAGCATGTGTTAAATCTAAATTCAGAGAAATCTTTGGAATGAATGGTATTGTTGACATGAAAACTGCCAT
TTTAGGAACTTCTCATGTAAATTCTATGGATTTGAGCACTTCAGCTGGTTATAGTTTTGTTAAATCTGGCTACAAAAAGA
AAGATCTAATTTGCTTGGAACCATTCTCAGTAGCCCCATTGTTAGAGAGGCTTGTACAGGACAAATTCCACAACTTACTA
AAAGGGAATCAAATAACTACAACCTTCAACACTTGCCTTAAAGATGAGTTAAGAAAGCTAGATAAGATTGCCTCANGAAA
AACTAGGTGTATAGAGGCTTGTGAAGTTGATTATTGCATTGTTTACAGAATGATCATGATGGAAATTTATGACAAGATTT
ATCAAACTCCATGTTACTACTCANGACTAGCTGTTGGAATCAACCCATACAAAGATTGGCACTTCATGATTAATGCTCTC
AATGATTACAATTATGAAATGGACTATTCTCAATATGATGGCTCCCTTANTTCAATGTTGCTATGGGAGGCTGNGGNAGT
TCTAGCTTACTGTCATGATTCACCTGATCTCNNGATGCAACTGCACAAANCCANTCATNNACTCTGANNATNNTNNTTTT
CNNCGAGANATGNNNGNNNANGGNNGNNTGNCCATCNNNNNTNCANGTACTACNNNNNNNNNNNNNNGNNATTNNANNAN
NNNNATTTNNNNNANNNNNNNCANNNNNNNNTNNNNNNANNNNNNNNNNNNNNNNNNNNNNTNNNNNNNCNNNNNNNNN
```

```
ori             GAGTGCACCATTCAGGTGGCATCAAGACAAGGGAAGGAACTGAAAGCACAAAGACTATCA
21              ------------------------------NNNNNNNNNNNNTNANNACAA-GACTATCA
                                                *  **** ********

ori             GTTATACTGTTAAATCTTGCAAAGGAATGTGTGGTGGCCTACTTATTTCAAAAGTAGAAG
21              GTTATACTGTTAAATCTTGCAAAGGAATGTGTGGTGGCCTACTTATTTCAAAAGTAGAAG
                ************************************************************

ori             GTAACTTCAAAATCCTGGGTATGCATATTGCTGGTAATGGTGAAATGGGAGTAGCTATAC
21              GTAACTTCAAAATCCTGGGTATGCATATTGCTGGTAATGGTGAAATGGGAGTAGCTATAC
                ************************************************************

ori             CCTTTAATTTTCTTAAAAATGACATGTCTGATCAAGGCATTGTTACTGAGATCACTCCAA
21              CCTTTAATTTTCTTAAAAATGACATGTCTGATCAAGGCATTGTTACTGAGATCACTCCAA
                ************************************************************

ori             TCCAGCCCATGTATATAAACACTAAAACTCAGATCCACAAGAGCCCTGTCTATGGTGCTG
21              TCCAGCCCATGTATATAAACACTAAAACTCAGATCCACAAGAGCCCTGTCTATGGTGCTG
                ************************************************************

ori             TTGAGGTGAAAATGGGTCCAGCTGTCCTAAGTAAGTCAGACACCAGGTTGGAGGAACCTG
21              TTGAGGTGAAAATGGGTCCAGCTGTCCTAAGTAAGTCAGACACCAGGTTGGAGGAACCTG
                ************************************************************

ori             TAGAATGTCTTATTAAAAAATCGGCTTCCAAGTACAGAGTTAACAAATTCCAGGTGAACA
21              TAGAATGTCTTATTAAAAAATCGGCTTCCAAGTACAGAGTTAACAAATTCCAGGTGAACA
                ************************************************************

ori             ATGAACTCTGGCAAGGTGTCAAAGCATGTGTTAAATCTAAATTCAGAGAAATCTTTGGAA
21              ATGAACTCTGGCAAGGTGTCAAAGCATGTGTTAAATCTAAATTCAGAGAAATCTTTGGAA
                ************************************************************

ori             TGAATGGTATTGTTGACATGAAAACTGCCATTTTAGGAACTTCTCATGTAAATTCTATGG
21              TGAATGGTATTGTTGACATGAAAACTGCCATTTTAGGAACTTCTCATGTAAATTCTATGG
                ************************************************************

ori             ATTTGAGCACTTCAGCTGGTTATAGTTTTGTTAAATCTGGCTACAAAAAGAAAGATCTAA
21              ATTTGAGCACTTCAGCTGGTTATAGTTTTGTTAAATCTGGCTACAAAAAGAAAGATCTAA
                ************************************************************
```

```
ori             TTTGCTTGGAACCATTCTCAGTAGCCCCATTGTTAGAGAGGCTTGTACAGGACAAATTCC
21              TTTGCTTGGAACCATTCTCAGTAGCCCCATTGTTAGAGAGGCTTGTACAGGACAAATTCC
                ************************************************************

ori             ACAACTTACTAAAAGGGAATCAAATAACTACAACCTTCAACACTTGCCTTAAAGATGAGT
21              ACAACTTACTAAAAGGGAATCAAATAACTACAACCTTCAACACTTGCCTTAAAGATGAGT
                ************************************************************

ori             TAAGAAAGCTAGATAAGATTGCCTCAGGAAAAACTAGGTGTATAGAGGCTTGTGAAGTTG
21              TAAGAAAGCTAGATAAGATTGCCTCAGGAAAAACTAGGTGTATAGAGGCTTGTGAAGTTG
                ************************************************************

ori             ATTATTGCATTGTTTACAGAATGATCATGATGGAAATTTATGACAAGATTTATCAAACTC
21              ATTATTGCATTGTTTACAGAATGATCATGATGGAAATTTATGACAAGATTTATCAAACTC
                ************************************************************

ori             CATGTTACTACTCAGGACTAGCTGTTGGAATCAACCCATACAAAGATTGGCACTTCATGA
21              CATGTTACTACTCANGACTAGCTGTTGGAATCAACCCATACAAAGATTGGCACTTCATGA
                ************** ********************************************

ori             TTAATGCTCTCAATGATTACAATTATGAAATGGACTATTCTCAATATGATGGCTCCCTTA
21              TTAATGCTCTCAATGATTACAATTATGAAATGGACTATTCTCAATATGATGGCTCCCTTA
                ************************************************************

ori             GTTCAATGTTGCTATGGGAGGCTGTGGAAGTTCTAGCTTACTGTCATGATTCACCTGATC
21              NTTCAATGTTGCTATGGGAGGCTGNGGNAGTTCTAGCTTACTGTCATGATTCACCTGATC
                 ********************** ** ********************************

ori             TCGTGATGCAACTGCACAAACC-AGTCATTGACTCTGACCATGTTGTTTT-CAACGAGAG
21              TCNNGATGCAACTGCACAAANCCANTCATNNACTCTGANNATNNTNNTTTTCNNCGAGAN
                **  **************** * * ****  *******  **  *  *** *  *****

ori             ATGGTTGATTCATGGTGGTATGCCATCAGGATCTCCATGTACTACTGTGTTAAATTCACT
21              ATGNNNGNNN-ANGGNNGNNTGNCCATCNNNNNTNCANGTACTACNNNNNNNNNNNNN--
                ***   *    * **  *  ** *         * ** *******

ori             ATGCAATTTGATGATGTGTATTTACACTACAAATTTAATCAGTCCTGGAATTGATTGTCT
21              -NGNNATTNNANNANNNNNNATTTNNNNNANNNNNNNCANNNNNNNNTNNNNNNANNNNNN
                  *  ***  *  *     ****     *        *              *
```

pHPeV1(21mut)Luc sequenced through the 3D^{pol}-encoding region of HPeV1

```
CNNNNNNNNNNNGGTGTATAGAGGCTTGTGAGTTGATTATTGCATTGTTTACAGAATGATCATGATGGAAATTTATGACA
AGATTTATCAAACTCCATGTTACTACTCAGGACTAGCTGTTGGAATCAACCCATACAAAGATTGGCACTTCATGATTAAT
GCTCTCAATGATTACAATTATGAAATGGACTATTCTCAATATGATGGCTCCCTTAGTTCAATGTTGCTATGGGAGGCTGT
GGAAGTTCTAGCTTACTGTCATGATTCACCTGATCTCGTGATGCAACTGCACAAACCAGTCATTGACTCTGACCATGTTG
TTTTCAACGAGAGATGGTTGATTCATGGTGGTATGCCATCAGGATCTCCATGTACTACTGTGTTAAATTCACTATGCAAT
TTGATGATGTGTATTTACACTACAAATTTAATCAGTCCTGGAATTGATTGTCTACCAATCGTATATGGCGATGATGTTAT
TCTTTCATTGGACAAAGAAATTGAACCAGAGAAACTGCAAAGTATCATGGCAGATTCATTTGGAGCCGAAGTGACTGGGT
CACGCAAAGATGAACCTCCTTCACTTAAACCAAGAATGGAGGTTGAATTCTTAAAGCGGAAACCTGGTTATTTCCCAGAA
AGCACTTTCATTGTAGGAAAGTTAGACACAGAAAATATGATACAACATCTTATGTGGATGAAAAACTTTTCAACCTTTAA
ACAGCAACTGCAAAGTTATTTAATGGGAACTATGCCTCCATGGAAAAGACACTTATCAACACTACATTAAGATCTTGGAA
CCATATCTACAGGAATGGAATATCACTGTGGATGATTATGATGTGGTTATAACTAAGTTGATGCCCATGGTGTTTGATTA
AGATTAATGTTTTGTTTTTCTTTTGCTATGGACTATGTGGNTAGAACACTAATTAATTNNNCACTAACTTGGNNCATTNG
NCATGANNNNNNNNNNAANNANAAAANNNANNANCNCGTCTCGAGTNNNNTCANTCCNNNNNNNTGNNNNGNANCATNNC
NNNNTCNNNNNNNCNNCTTTGTNCNNNNNNNAGNNNNNNNNTGNNNANCNNNGNNNNNCNNNNNNNNNAANNNNNNCNNN
NANNCNNNNNNNCNNNNNANCNNNNNNNNNCCNGNNNNNANGNNNGNNNN
```

```
ori             TAAGAAAGCTAGATAAGATTGCCTCAGGAAAAACTAGGTGTATAGAGGCTTGTGAAGTTG
seq             ------------------------CNNNNNNNNNNNNGGTGTATAGAGGCTTGTGA-GTTG
                                        *           ******************* ****

ori             ATTATTGCATTGTTTACAGAATGATCATGATGGAAATTTATGACAAGATTTATCAAACTC
seq             ATTATTGCATTGTTTACAGAATGATCATGATGGAAATTTATGACAAGATTTATCAAACTC
                ************************************************************

ori             CATGTTACTACTCAGGACTAGCTGTTGGAATCAACCCATACAAAGATTGGCACTTCATGA
seq             CATGTTACTACTCAGGACTAGCTGTTGGAATCAACCCATACAAAGATTGGCACTTCATGA
                ************************************************************

ori             TTAATGCTCTCAATGATTACAATTATGAAATGGACTATTCTCAATATGATGGCTCCCTTA
seq             TTAATGCTCTCAATGATTACAATTATGAAATGGACTATTCTCAATATGATGGCTCCCTTA
                ************************************************************

ori             GTTCAATGTTGCTATGGGAGGCTGTGGAAGTTCTAGCTTACTGTCATGATTCACCTGATC
seq             GTTCAATGTTGCTATGGGAGGCTGTGGAAGTTCTAGCTTACTGTCATGATTCACCTGATC
                ************************************************************

ori             TCGTGATGCAACTGCACAAACCAGTCATTGACTCTGACCATGTTGTTTTCAACGAGAGAT
seq             TCGTGATGCAACTGCACAAACCAGTCATTGACTCTGACCATGTTGTTTTCAACGAGAGAT
                ************************************************************

ori             GGTTGATTCATGGTGGTATGCCATCAGGATCTCCATGTACTACTGTGTTAAATTCACTAT
seq             GGTTGATTCATGGTGGTATGCCATCAGGATCTCCATGTACTACTGTGTTAAATTCACTAT
                ************************************************************

ori             GCAATTTGATGATGTGTATTTACACTACAAATTTAATCAGTCCTGGAATTGATTGTCTAC
seq             GCAATTTGATGATGTGTATTTACACTACAAATTTAATCAGTCCTGGAATTGATTGTCTAC
                ************************************************************

ori             CAATCGTATATGGCGATGATGTTATTCTTTCATTGGACAAAGAAATTGAACCAGAGAAAC
seq             CAATCGTATATGGCGATGATGTTATTCTTTCATTGGACAAAGAAATTGAACCAGAGAAAC
                ************************************************************

ori             TGCAAAGTATCATGGCAGATTCATTTGGAGCCGAAGTGACTGGGTCACGCAAAGATGAAC
seq             TGCAAAGTATCATGGCAGATTCATTTGGAGCCGAAGTGACTGGGTCACGCAAAGATGAAC
                ************************************************************

ori             CTCCTTCACTTAAACCAAGAATGGAGGTTGAATTCTTAAAGCGGAAACCTGGTTATTTCC
seq             CTCCTTCACTTAAACCAAGAATGGAGGTTGAATTCTTAAAGCGGAAACCTGGTTATTTCC
                ************************************************************

ori             CAGAGTCTACTTTTATAGTAGGTAAATTAGACACTGAAAATATGATACAACATTTAATGT
seq             CAGAAAGCACTTTCATTGTAGGAAAGTTAGACACAGAAAATATGATACAACATCTTATGT
                ****    ***** ** ***** ** ******** ***************** * ****

ori             GGATGAAAAACTTTAGCACATTTAAGCAGCAGCTCCAATCCTATTTAATGGGAACTATGC
seq             GGATGAAAAACTTTTCAACCTTTAAACAGCAACTGCAAAGTTATTTAATGGGAACTATGC
                ************   ** ***** ***** ** ***   ******************

ori             CTCCATGGAAAAGACACTTATCAACACTACATTAAGATCTTGGAACCATATCTACAGGAA
seq             CTCCATGGAAAAGACACTTATCAACACTACATTAAGATCTTGGAACCATATCTACAGGAA
                ************************************************************

ori             TGGAATATCACTGTGGATGATTATGATGTGGTTATAACTAAGTTGATGCCCATGGTGTTT
seq             TGGAATATCACTGTGGATGATTATGATGTGGTTATAACTAAGTTGATGCCCATGGTGTTT
                ************************************************************
```

```
ori             GATTAAGATTAATGTTTTGTTTTTCTTTTGCTATGGACTATGTGGTTAGAACACTAATTA
seq             GATTAAGATTAATGTTTTGTTTTTCTTTTGCTATGGACTATGTGGNTAGAACACTAATTA
                **************************************** **************

ori             ATTAGGACACTAACTTGGAACATTGGACATGACAAAAAAAAAAAAAAAAAAAAAAAAA-
seq             ATTNNN-CACTAACTTGGNNCATTNGNCATGANNNNNNNNNNAANNANAAAANNNANNAN
                ***    ***********  **** * *****          **  * ****   *  *
```

Figure 7: Desired sequence (ori) of pHPeV1(21mut)Luc through Mph1103I (ATGCAT, highlighted in red) in the $3D^{pol}$-encoding region of HPeV1. The blue highlight shows the continuation of the correct $3D^{pol}$ sequence while the pink highlight shows the correct mutations in the sequence.

pHPeV1(YGAA)Luc sequenced through Mph1103I and to middle of the 3D^{pol}-encoding region of HPeV1

```
NNNNNNNNNNNNNNTNANGCACAAGACTATCAGTTATACTGTTAAATCTTGCAAAGGAATGTGTGGTGGCCTACTTATTT
CAAAAGTAGAAGGTAACTTCAAAATCCTGGGTATGCATATTGCTGGTAATGGTGAAATGGGAGTAGCTATACCCTTTAAT
TTTCTTAAAAATGACATGTCTGATCAAGGCATTGTTACTGAGATCACTCCAATCCAGCCCATGTATATAAACACTAAAAC
TCAGATCCACAAGAGCCCTGTCTATGGTGCTGTTGAGGTGAAAATGGGTCCAGCTGTCCTAAGTAAGTCAGACACCAGGT
TGGAGGAACCTGTAGAATGTCTTATTAAAAAATCGGTTTCCAAGTACAGAGTTAACAAATTCCAGGTGAACAATGAACTC
TGGCAAGGTGTCAAAGCATGTGTTAAATCTAAATTCAGAGAAATCTTTGGAATGAATGGTATTGTTGACATGAAAACTGC
CATTTTAGGAACTTCTCATGTAAATTCTATGGATTTGAGCACTTCAGCTGGTTATAGTTTTGTTAAATCTGGCTACAAAA
AGAAAGATCTAATTTGCTTGGAACCATTCTCAGTAGCCCCATTGTTAGAGAGGCTTGTACAGGACAAATTCCACAACTTA
CTAAAAGGGAATCAAATAACTACAACCTTCAACACTTGCCTTAAAGATGAGTTAAGAAAGCTAGATAAGATTGCCTCANG
AAAAACTAGGTGTATAGAGGCTTGTGAAGTTGATTATTGCATTGTTTACAGAATGATCATGATGGAAATTTATGACAAGA
TTTATCAAACTCCATGTTACTACTCAGNACTAGCTGTTGGAATCAACCCATACAAAGATTGGCACTTCATGATTAATGCT
CTCAATGATTACAATTATGAAATGGACTATTCTCAATATGATGGCTCCCNTAGTTCAATGNTGCTATGGGAGGCTGTGGN
AGTTCTAGCTTACTGTCATGATTCACCTGATCTCGTGATGCANTGCANAAANCAGTCATTGACTCTGANNATNNTGNTTT
CANGANAGATGGNNNANTCATNNNNNTGCATCNNGNTNNNCNNNNACTACNNNNTAAANNNCTATGCANTTGATGANGNG
NATTTNNNCNNNCNAATTTNNCANNCNNNNNTGNNNNNNCTNNNNNCNNNNNNNNNNNNNNNNTTCNNNNNNNAANAAT
NNNNCNNNNNAAAACNNNNNNNN
```

```
seq             GTGTGGTGGCCTACTTATTTCAAAAGTAGAAGGTAACTTCAAAATCCTGGGTATGCATAT
ori             ------------------------------------------------------GTATGCATAT
                                                                       **********

seq             TGCTGGTAATGGTGAAATGGGAGTAGCTATACCCTTTAATTTTCTTAAAAATGACATGTC
ori             TGCTGGTAATGGTGAAATGGGAGTAGCTATACCCTTTAATTTTCTTAAAAATGACATGTC
                ************************************************************

seq             TGATCAAGGCATTGTTACTGAGATCACTCCAATCCAGCCCATGTATATAAACACTAAAAC
ori             TGATCAAGGCATTGTTACTGAGATCACTCCAATCCAGCCCATGTATATAAACACTAAAAC
                ************************************************************

seq             TCAGATCCACAAGAGCCCTGTCTATGGTGCTGTTGAGGTGAAAATGGGTCCAGCTGTCCT
ori             TCAGATCCACAAGAGCCCTGTCTATGGTGCTGTTGAGGTGAAAATGGGTCCAGCTGTCCT
                ************************************************************

seq             AAGTAAGTCAGACACCAGGTTGGAGGAACCTGTAGAATGTCTTATTAAAAAATCGGTTTC
ori             AAGTAAGTCAGACACCAGGTTGGAGGAACCTGTAGAATGTCTTATTAAAAAATCGGCTTC
                ************************************************************ ***

seq             CAAGTACAGAGTTAACAAATTCCAGGTGAACAATGAACTCTGGCAAGGTGTCAAAGCATG
ori             CAAGTACAGAGTTAACAAATTCCAGGTGAACAATGAACTCTGGCAAGGTGTCAAAGCATG
                ************************************************************

seq             TGTTAAATCTAAATTCAGAGAAATCTTTGGAATGAATGGTATTGTTGACATGAAAACTGC
ori             TGTTAAATCTAAATTCAGAGAAATCTTTGGAATGAATGGTATTGTTGACATGAAAACTGC
                ************************************************************

seq             CATTTTAGGAACTTCTCATGTAAATTCTATGGATTTGAGCACTTCAGCTGGTTATAGTTT
ori             CATTTTAGGAACTTCTCATGTAAATTCTATGGATTTGAGCACTTCAGCTGGTTATAGTTT
                ************************************************************

seq             TGTTAAATCTGGCTACAAAAAGAAAGATCTAATTTGCTTGGAACCATTCTCAGTAGCCCC
ori             TGTTAAATCTGGCTACAAAAAGAAAGATCTAATTTGCTTGGAACCATTCTCAGTAGCCCC
                ************************************************************

seq             ATTGTTAGAGAGGCTTGTACAGGACAAATTCCACAACTTACTAAAAGGGAATCAAATAAC
ori             ATTGTTAGAGAGGCTTGTACAGGACAAATTCCACAACTTACTAAAAGGGAATCAAATAAC
                ************************************************************

seq             TACAACCTTCAACACTTGCCTTAAAGATGAGTTAAGAAAGCTAGATAAGATTGCCTCANG
ori             TACAACCTTCAACACTTGCCTTAAA----------------------------------
                *************************
```

pHPeV1(YGAA)Luc sequenced through the middle of 3D^{pol}-encoding region of HPeV1

```
NNNNNNNNAANNNGGTGTATAGAGGCTTGTGAAGTTGATTATTGCATTGTTTACAGAATGATCATGATGGAAATTTATGA
CAAGATTTATCAAACTCCATGTTACTACTCAGGACTAGCTGTTGGAATCAACCCATACAAAGATTGGCACTTCATGATTA
ATGCTCTCAATGATTACAATTATGAAATGGACTATTCTCAATATGATGGCTCCCTTAGTTCAATGTTGCTATGGGAGGCT
GTGGAAGTTCTAGCTTACTGTCATGATTCACCTGATCTCGTGATGCAACTGCACAAACCAGTCATTGACTCTGACCATGT
TGTTTTCAACGAGAGATGGTTGATTCATGGTGGTATGCCATCAGGATCTCCATGTACTACTGTATTAAATTCACTATGCA
ATTTGATGATGTGTATTTACACTACAAATTTAATCAGTCCTGGAATTGATTGTCTACCAATCGTATATGGCGCTGCTGTT
ATTCTTTCATTGGACAAAGAAATTGAACCAGAGAAACTGCAAAGTATCATGGCAGATTCATTTGGAGCCGAAGTGACTGG
GTCACGCAAAGATGAACCTCCTTCACTTAAACCAAGAATGGAGGTTGAATTCTTAAAGCGGAAACCTGGTTATTTCCCAG
AGTCTACTTTTATAGTAGGTAAATTAGACACTGAAAATATGATACAACATTTAATGTGGATGAAAAACTTTAGCACATTT
AAGCAGCAGCTCCNATCCTATTTAATGGAACTATGCCTCCATGGAAAAGACACTTATCAACACTACATTAAGATCTTGGA
ACCATATCTACAGGAATGGAATATCACTGTGGATGATTATGATGTGGTTATAACTAAGTTGATGCCCATGGTGTTTGATT
AAGATTAATGTTTTGTTTTTCTTTTGCTATGGACTATGTGGGTTAGAACACTAATTNATTAGGACACNNNNTGNNNATTG
GNCATGANNNNNNNNNNNNNNNNNNNAANNNGTNNTCGAGTCNNNTCAATCCCGNNGNCATGGCNGCNGGANCATGCNAC
GTNNNNCNNGNNCNGNNNNNTNCCTTTANNGANGNNNTTCNNNCTGNGTATCANGNCANANCNGNTTNNNNNNNNNNNT
ANCNNCNNCNNNNNNNNNNNNNNNNNNNNNNAANNNNNANNGNNNNNGGNNNNNNNGNNNGNNTANNNNNNNNNNNNNNNN
NNNNNNNNNNN
```

```
ori             GAGTTAAGAAAGCTAGATAAGATTGCCTCAGGAAAAACTAGGTGTATAGAGGCTTGTGAA
seq             ---------------------------NNNNNNNNAANNNGGTGTATAGAGGCTTGTGAA
                                            **   ********************

ori             GTTGATTATTGCATTGTTTACAGAATGATCATGATGGAAATTTATGACAAGATTTATCAA
seq             GTTGATTATTGCATTGTTTACAGAATGATCATGATGGAAATTTATGACAAGATTTATCAA
                ************************************************************

ori             ACTCCATGTTACTACTCAGGACTAGCTGTTGGAATCAACCCATACAAAGATTGGCACTTC
seq             ACTCCATGTTACTACTCAGGACTAGCTGTTGGAATCAACCCATACAAAGATTGGCACTTC
                ************************************************************

ori             ATGATTAATGCTCTCAATGATTACAATTATGAAATGGACTATTCTCAATATGATGGCTCC
seq             ATGATTAATGCTCTCAATGATTACAATTATGAAATGGACTATTCTCAATATGATGGCTCC
                ************************************************************

ori             CTTAGTTCAATGTTGCTATGGGAGGCTGTGGAAGTTCTAGCTTACTGTCATGATTCACCT
seq             CTTAGTTCAATGTTGCTATGGGAGGCTGTGGAAGTTCTAGCTTACTGTCATGATTCACCT
                ************************************************************

ori             GATCTCGTGATGCAACTGCACAAACCAGTCATTGACTCTGACCATGTTGTTTTCAACGAG
seq             GATCTCGTGATGCAACTGCACAAACCAGTCATTGACTCTGACCATGTTGTTTTCAACGAG
                ************************************************************

ori             AGATGGTTGATTCATGGTGGTATGCCATCAGGATCTCCATGTACTACTGTGTTAAATTCA
seq             AGATGGTTGATTCATGGTGGTATGCCATCAGGATCTCCATGTACTACTGTATTAAATTCA
                ************************************************* *********

ori             CTATGCAATTTGATGATGTGTATTTACACTACAAATTTAATCAGTCCTGGAATTGATTGT
seq             CTATGCAATTTGATGATGTGTATTTACACTACAAATTTAATCAGTCCTGGAATTGATTGT
                ************************************************************

ori             CTACCAATCGTATATGGCGATGATGTTATTCTTTCATTGGACAAAGAAATTGAACCAGAG
seq             CTACCAATCGTATATGGCGCTGCTGTTATTCTTTCATTGGACAAAGAAATTGAACCAGAG
                ******************  **  ************************************

ori             AAACTGCAAAGTATCATGGCAGATTCATTTGGAGCCGAAGTGACTGGGTCACGCAAAGAT
seq             AAACTGCAAAGTATCATGGCAGATTCATTTGGAGCCGAAGTGACTGGGTCACGCAAAGAT
                ************************************************************

ori             GAACCTCCTTCACTTAAACCAAGAATGGAGGTTGAATTCTTAAAGCGGAAACCTGGTTAT
seq             GAACCTCCTTCACTTAAACCAAGAATGGAGGTTGAATTCTTAAAGCGGAAACCTGGTTAT
                ************************************************************

ori             TTCCCAGAGTCTACTTTTATAGTAGGTAAATTAGACACTGAAAATATGATACAACATTTA
```

```
seq             TTCCCAGAGTCTACTTTTATAGTAGGTAAATTAGACACTGAAAATATGATACAACATTTA
                ************************************************************

ori             ATGTGGATGAAAAACTTTAGCACATTTAAGCAGCAGCTCCAATCCTATTTAATGGAACTA
seq             ATGTGGATGAAAAACTTTAGCACATTTAAGCAGCAGCTCCNATCCTATTTAATGGAACTA
                ************************************ ******************

ori             TGCCTCCATGGAAAAGACACTTATCAACACTACATTAAGATCTTGGAACCATATCTACAG
seq             TGCCTCCATGGAAAAGACACTTATCAACACTACATTAAGATCTTGGAACCATATCTACAG
                ************************************************************

ori             GAATGGAATATCACTGTGGATGATTATGATGTGGTTATAACTAAGTTGATGCCCATGGTG
seq             GAATGGAATATCACTGTGGATGATTATGATGTGGTTATAACTAAGTTGATGCCCATGGTG
                ************************************************************

ori             TTTGATTAAGATTAATGTTTTGTTTTTCTTTTGCTATGGACTATGTGG-TTAGAACACTA
seq             TTTGATTAAGATTAATGTTTTGTTTTTCTTTTGCTATGGACTATGTGGGTTAGAACACTA
                ********************************************* ***********

ori             ATTAATTAGGACACTAACTTGGAACATTGGACATGACAAAAAAAAAAAAAAAAAAAAAA
seq             ATTNATTAGGACACNNNNT--GNNNATTGGNCATGANNNNNNNNNNNNNNNNNNNAANNN
                *** **********    *  *   ***** *****                   **

ori             AAA-------------------------------------------------------
```

Figure 8: Desired pHPeV1(YGAA)Luc sequence (ori) through Mph1103I (ATGCAT, highlighted in red) in the 3D^{pol}-encoding region of HPeV1. The blue highlight shows the continuation of the correct 3D^{pol} sequence while the pink highlight shows the correct mutations in the sequence. The yellow highlight shows a mutation in the sequence which should not affect the overall result since the YGAA mutation should prevent the replication of the RNA.