

Virus-host interactions in the early stages of the Equine Hepacivirus (EqHV) lifecycle

Joseph Robert Lattimer

**Submitted in accordance with the requirements for the degree of
Doctor of Philosophy**

The University of Leeds

Faculty of Biological Sciences

School of Molecular and Cellular Biology

September 2017

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement

© 2017 The University of Leeds and Joseph Robert Lattimer

The right of Joseph Robert Lattimer to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgements

I would firstly like to thank my supervisors Prof. Mark Harris and Prof. Nicola Stonehouse for their continued guidance and support throughout this project.

I would also like to thank Dr. Hazel Stewart, Dr. Cheryl Walter, and Dr. Zsofia Igloi for helping me learn the ropes, and for their scientific input at various stages throughout this project.

Thank you to the past and present members of Garstang 8.61, with a special mention to Andrew, Chris, Emma, and Lauren, in part for their helpful advice and discussion, but mostly for keeping me sane through four long years! A special thank you also goes out to my parents for supporting me through thick and thin, often having more enthusiasm than I did, and especially for always asking about my project even if they had no idea what I was talking about. A big thank you also goes to Penny for seeing me through the final hurdle, and not complaining when I had to abandon her to go to work.

Finally, I would like to thank the Biotechnology and Biological Sciences Research Council (BBSRC) for funding this project.

Abstract

EqHV is the most closely related virus to HCV, and is proposed to have diverged from HCV within the last 1000 years. Studying viruses which share a close evolutionary relationship, yet differ greatly in their ability to cause disease, provides a unique opportunity to facilitate comparative analysis of both replication strategies and pathogenic mechanisms. It has previously been demonstrated that, like HCV, the EqHV 5'UTR functioned as an IRES. Here we set out to undertake a more detailed structural and functional analysis of the 5'UTR.

A mutational analysis of the structural features of the EqHV 5'UTR was undertaken to investigate their roles in its function as an IRES. A complementary study utilising 2' hydroxyl acylation and analysis by primer extension (SHAPE), to experimentally confirm the predicted structure of the EqHV 5'UTR in order to reveal how this related to function, was also conducted. This combined approach revealed that miR122 mediated enhancement of translation was dependent on a seed site located directly upstream of stem-loop (SL) II and that SLI was not required for EqHV IRES mediated translation. SLIII was absolutely essential, as was SLIIId and the conserved GGG motif. An essential role for SLIIIb was demonstrated for translation from an EqHV sub-genomic replicon. SHAPE experiments revealed that the conserved GGG motif was essential for EqHV 5'UTR:40S ribosomal subunit interactions, and that a GTC sequence in the apical loop of SLIIIb mediated an interaction with eIF3.

This study also set out to establish an EqHV sub-genomic replicon capable of replicating *in vitro* in mammalian cells; however, this could not be achieved. When taken with published data it is possible to conclude that current EqHV isolates are not able to replicate in currently available cell culture systems and it is likely that new isolates, or cell lines, will be required to achieve *in vitro* replication.

Table of contents

Acknowledgements.....	ii
Abstract.....	iii
Abbreviations.....	xii
Chapter 1 Introduction	1
1.1 Overview	2
1.2 Translation initiation.....	3
1.2.1 Cap-dependent initiation.....	3
1.2.2 Viral IRES	6
1.2.3 Hepaciviral IRES.....	8
1.3 RNA structure and determination	11
1.3.1 The folding of RNA into secondary and tertiary structures.....	11
1.3.2 Viral RNA structure	12
1.3.3 RNA structure prediction software.....	13
1.3.4 Selective 2'-hydroxyl acylation analysed by primer extension (SHAPE)	13
1.4 Hepatitis C virus	16
1.4.1 Prevalence, transmission and pathogenicity	16
1.4.2 Organisation of the HCV genome	16
1.4.3 The HCV viral particle and entry	17
1.4.4 Polyprotein expression and processing, and replication	18
1.4.5 Assembly and release	19
1.4.6 The HCV sub genomic replicon	20
1.5 Equine hepacivirus.....	21
1.5.1 Discovery and initial characterisation of EqHV.....	21
1.5.2 Prevalence, host range, and classification, of EqHV	23
1.5.3 Transmission, tissue tropism, and pathology, of EqHV	24
1.5.4 Molecular biology of EqHV – the story so far	26
1.6 Other newly discovered hepaciviruses	27
1.6.1 Genomic organisation of the newly discovered hepaciviruses	27
1.6.2 Tissue tropism, pathology, and prevalence, of the newly discovered hepaciviruses	29
1.6.3 Update to the taxonomy of the hepaciviruses	31
1.6.4 The newly discovered hepaciviruses as a model system to study HCV	32
1.7 Project aims.....	34
Chapter 2 Materials and Methods.....	35

2.1 Materials:	36
2.1.1 Mammalian cell lines	36
2.1.2 Bacterial strains.....	36
2.1.3 Virus sequences	36
2.1.4 SGR constructs	36
2.1.5 Bicistronic expression constructs.....	37
2.1.6 miR122 reporter constructs.....	37
2.1.7 N-Methylisatoic anhydride (NMIA).....	37
2.1.8 Purified eIF3 and 40S ribosomal subunit	37
2.1.9 Primers	37
2.1.10 Tissue culture microscope	37
2.2 Nucleic acid manipulation.....	38
2.2.1 Bacterial transformation.....	38
2.2.2 Preparation of DNA from bacterial culture.....	38
2.2.3 Nucleic acid quantification.....	38
2.2.4 Q5 site-directed mutagenesis.....	38
2.2.5 Restriction digests.....	39
2.2.6 Agarose gel electrophoresis.....	39
2.2.7 Gel extraction of DNA	39
2.2.8 Ligation.....	39
2.2.9 Verification of ligation products	39
2.2.10 Linearisation of DNA	39
2.2.11 RNA manipulation and DEPC treatment	40
2.2.12 <i>In vitro</i> transcription of RNA	40
2.2.13 Lithium chloride precipitation of RNA	40
2.2.14 RNA gel electrophoresis.....	40
2.3 Tissue culture	41
2.3.1 Mammalian cell lines	41
2.3.2 Freezing cell stocks.....	41
2.3.3 Retroviral transduction	41
2.3.4 Maintenance of mammalian cells expressing miR122, SEC14L2, or the PIV5 V protein	41
2.3.5 RNA electroporation	42
2.3.6 Nucleic acid transfection.....	42

2.3.7 Firefly luciferase assays.....	42
2.3.8 Dual Luciferase assay	42
2.3.9 Selection of stable cell lines	43
2.4 <i>In vitro</i> methods.....	43
2.4.1 Selective 2'hydroxyl acylation and analysis by primer extension (SHAPE).....	43
2.4.2 Gel Based SHAPE	44
2.4.3 SHAPE footprinting.....	44
2.4.4 SHAPE data analysis	44
2.4.5 Pull downs.....	44
2.4.6 TMT mass spectrometry	45
Chapter 3 A structural and functional analysis of the EqHV 5'UTR	46
3.1 Introduction	47
3.2 Results.....	51
3.2.1 Predicting the structure of the EqHV 5'UTR and structural mutants	51
3.2.2 The execution of SHAPE experiments.....	57
3.2.3 SHAPE analysis of the WT 5'UTR.....	61
3.2.4 SHAPE analysis of the Δ SLI 5'UTR.....	67
3.2.5 SHAPE analysis of the Δ SLI+II 5'UTR.....	70
3.2.6 SHAPE analysis of the Δ SLIII 5'UTR.....	73
3.2.7 SHAPE analysis of the Δ SLIIIb 5'UTR.....	76
3.2.8 SHAPE analysis of the Δ SLIIId 5'UTR.....	79
3.2.9 Investigating the activity of EqHV 5'UTR mutants in a bicistronic construct.....	82
3.2.10 Investigating the activity of EqHV 5'UTR mutants in an SGR.....	88
3.3 Discussion.....	94
3.3.1 Predicting the structure of the EqHV 5'UTR, and structural mutants	94
3.3.2 Experimental determination of the EqHV 5'UTR RNA secondary structure	94
3.3.3 The structure of EqHV SLI, SLIA, and SLII	95
3.3.4 The structure of EqHV SLIII	96
3.3.5 The structure of the EqHV pseudoknot	97
3.3.6 EqHV 5'UTR Δ SLI and Δ SLI+II structural deletions have no overall impact on 5'UTR secondary structure	98
3.3.7 The effect of SLI+II on EqHV 5'UTR IRES function	100
3.3.8 Structural deletions in EqHV 5'UTR SLIII have no overall impact on 5'UTR secondary structure.....	102

3.3.9 The effect of SLIII on EqHV 5'UTR IRES function	104
3.3.10 The role of miR122 in translation from the EqHV 5'UTR IRES – a tale of two constructs.....	106
Chapter 4 EqHV 5'UTR interactions with the host cell translational machinery	109
4.1 Introduction	110
4.2 Results.....	112
4.2.1 Sequence specific mutations in EqHV SLIIIb and SLIIId disrupt translation.....	112
4.2.2 SHAPE footprinting analysis	114
4.2.3 The EqHV 5'UTR interacts with eIF3 and the 40S ribosomal subunit	116
4.2.4 Mutations in SLIIIb and SLIIId can disrupt EqHV 5'UTR interaction with eIF3 and the 40S ribosomal subunit	121
4.2.5 The EqHV 5'UTR interacts with host cell translational machinery necessary for HCV translation.....	126
4.3 Discussion.....	130
4.3.1 eIF3 interacts with the apical loop of EqHV 5'UTR SLIIIb during IRES mediated translation.....	130
4.3.2 40S ribosomal subunit interacts with ³⁰⁹ GGG ₃₁₁ during IRES mediated translation	131
4.3.3 The EqHV 5'UTR interacts with all of the factors required for HCV translation initiation.....	132
4.3.4 A predicted model of EqHV translation initiation.....	133
Chapter 5 Rational modification of an EqHV sub-genomic replicon to enhance replicative capacity	134
5.1 Introduction	135
5.2 Results.....	137
5.2.1 First generation EqHV, and chimeric, SGRs.....	137
5.2.2 Investigating alternative cell lines for SGR EqHV replication.....	145
5.2.3 Investigating stable replication of SGR EqHV.....	149
5.2.4 A second generation, consensus sequence, EqHV SGR	153
5.2.5 Replication assays of SGR NZPI and NZCI.....	156
5.3 Discussion:	160
5.3.1 Construction of an EqHV SGR from the EF369_11J isolate.....	160
5.3.2 pSGR NZPI and NZCI	161
5.3.3 Cell lines for the assay of EqHV SGR replication	163
5.3.4 Alternative factors to enable EqHV replication in cell culture.....	164

5.3.5 Alternative resistance genes to study stable replication in antibiotic resistant cell lines	165
5.3.6 Currently available EqHV SGRs are unable to replicate in cell culture	166
5.3.7 Roadblocks to establishing a cell culture system to study EqHV replication.....	167
Chapter 6 Conclusions and future perspectives	169
Chapter 7 References.....	174
Chapter 8 Appendices.....	192

List of Figures

Figure 1.1: Canonical and HCV mediated translation initiation.....	5
Figure 1.2: A Schematic representations of the predicted EqHV, and HCV, 5'UTRs	9
Figure 1.3: Hepaciviral genomic organisation.....	17
Figure 1.4: Phylogenetic analysis of the hepaciviruses.....	32
Figure 3.1: A Schematic representations of the predicted EqHV, and HCV, 5'UTRs	53
Figure 3.2: A schematic representation of the structural deletion mutations	56
Figure 3.3: Schematic representation of SHAPE	58
Figure 3.4: Determination of SHAPE primer extension using gel based SHAPE	59
Figure 3.5: Optimisation of NMIA concentrations for high-throughput SHAPE	60
Figure 3.6: The secondary structure of the EqHV 5'UTR	64
Figure 3.7: Experimental structure of the WT 5'UTR in SGR	65
Figure 3.8: Experimental structure of the EqHV 5'UTR in the bicistronic construct.....	66
Figure 3.9: Experimental structure of EqHV Δ SLI 5'UTR in the SGR.....	68
Figure 3.10: Experimental structure of EqHV Δ SLI 5'UTR in the bicistronic construct	69
Figure 3.11: Experimental structure of EqHV Δ SLI+II 5'UTR in the SGR.....	71
Figure 3.12: Experimental structure of EqHV Δ SLI+II 5'UTR in the bicistronic construct	72
Figure 3.13: Experimental structure of EqHV Δ SLIII 5'UTR in the SGR	74
Figure 3.14: Experimental structure of EqHV Δ SLIII 5'UTR in the bicistronic construct	75
Figure 3.15: Experimental structure of EqHV Δ SLIIIb 5'UTR in the SGR	77
Figure 3.16: Experimental structure of EqHV Δ SLIIIb 5'UTR in the bicistronic construct	78
Figure 3.17: Experimental structure of EqHV Δ SLIIId 5'UTR in the SGR	80
Figure 3.18: Experimental structure of EqHV Δ SLIIId 5'UTR in the bicistronic construct	81
Figure 3.19: EqHV mutant IRES activity in an RLuc/FLuc bicistronic construct	84
Figure 3.20: Assay of miR122 expression in FHK and 293T cells	85
Figure 3.21: The effect of miR122 on EqHV IRES activity in a bicistronic construct.....	87
Figure 3.22: EqHV mutant IRES activity in an SGR	90
Figure 3.23: The effect of miR122 on EqHV IRES activity in an SGR	92
Figure 3.24: The effect of miR122 on NZCI 124	93
Figure 4.1: Translation activity of SLIIIb and SLIIId apical loop mutations.....	113
Figure 4.2: A schematic representation of the SHAPE experimental procedure.....	115

Figure 4.3: Shape footprinting reveals an interaction between the apical loop of EqHV SLIIIb and eIF3.....	118
Figure 4.4: Shape footprinting reveals an interaction between the apical loop of EqHV SLIIId and the 40S ribosomal subunit	120
Figure 4.5: A mutation in the apical loop of EqHV SLIIIb prevents interaction with eIF3.....	122
Figure 4.6: A mutation in the apical loop of EqHV SLIIId prevents interaction with the 40S ribosomal subunit	124
Figure 5.1: HCV S2204 is conserved in EqHV at a.a position 2180	138
Figure 5.2: Transient replication of SGR H-E-H in human hepatoma cells	139
Figure 5.3: Transient replication of SGR H-E-E in human hepatoma cells	141
Figure 5.4: Transient replication of SGR E-E-H in human hepatoma cells	142
Figure 5.5: Transient replication of SGR EqHV in human hepatoma cells	144
Figure 5.6: Transient replication of SGR EqHV in FHK and 293T (+/- 122) cells.....	146
Figure 5.7: Transient replication of SGR EqHV in cell lines expressing SEC14L2 or the PIV5 V protein.....	148
Figure 5.8: A schematic representation of SGR NZPI and SGR NZCI.....	154
Figure 5.9: Transient replication of NZCI in human hepatoma cells.....	157
Figure 5.10: Transient replication of NZCI in equine cells	158
Figure 5.11: Transient replication of NZCI in human kidney cells.....	159
Figure 8.1: Mfold predicted structure of the EqHV 5'UTR.....	193
Figure 8.2: Mfold predicted structure of the EqHV Δ SLI 5'UTR	194
Figure 8.3: Mfold predicted structure of the EqHV Δ SLI+II 5'UTR	195
Figure 8.4: Mfold predicted structure of the EqHV Δ SLIII 5'UTR	196
Figure 8.5: Mfold predicted structure of the EqHV Δ SLIIIb 5'UTR	197
Figure 8.6: Mfold predicted structure of the EqHV Δ SLIIId 5'UTR	198
Figure 8.7: Mfold predicted structure of the EqHV GUC 5'UTR.....	199
Figure 8.8: Mfold predicted structure of the EqHV AGU 5'UTR	200
Figure 8.9: SHAPE primer sequences	201
Figure 8.10: Comparison of SHAPE reactivity data between SGR NZCI and pGL3 NPHV	202
Figure 8.11 A representative SGR JFH Replication assay and RNA gel	203
Figure 8.12: Alignment of the EqHV 3'UTR (accession number JX948116.1 [denoted SMKL]) and the HCV 3'UTR (accession number AF011753.1)	204

Figure 8.13: Alignment of the EqHV NZPI 3'UTR (accession number KP325401) and the HCV 3'UTR (accession number AF011753.1)	204
Figure 8.14: Alignment of the "original" EqHV 3'UTR (accession number JX948116.1 [denoted SMKL]) and the consensus EqHV NZPI 3'UTR (accession number KP325401).....	205

List of Tables

Table 4.1: SHAPE reactivity values for the apical loop of WT SLIIIb +/- eIF3	118
Table 4.2: SHAPE reactivity values for the apical loop of WT SLIIId +/- 40S ribosomal subunit	120
Table 4.3: SHAPE reactivity values for the apical loop of GUC SLIIIb +eIF3 in comparison to WT without eIF3	123
Table 4.4: SHAPE reactivity values for the apical loop of AGU SLIIIb +40S ribosomal subunit in comparison to WT no 40S.....	125
Table 4.5: eIF interacting partners of the EqHV 5'UTR.....	127
Table 4.6: 40S ribosomal interacting partners of the EqHV 5'UTR.....	128
Table 4.7: 60S ribosomal interacting partners of the EqHV 5'UTR.....	129
Table 5.1: pSGR EqHV, and HCV chimeric, constructs used in this study	150
Table 5.2: Antibiotic resistance profile of mammalian cells used in this study.....	151
Table 5.3: Stable selection of EqHV, and chimeric, SGRs in a range of cell lines.....	152
Table 5.4: A summary of pSGR NZPI/NZCI constructs	155

Abbreviations

Amino acid	a.a
ATCC	American Type Culture Collection
Apo E	Apolipoprotein E
ATP	Adenosine 5' triphosphate
BHV	Bat hepacivirus
BovHepV	Bovine hepacivirus
cDNA	Complementary DNA
CHV	Canine hepacivirus
CLDN1	Claudin 1
Con1b	Consensus 1b
CrPV	Cricket paralysis virus
Cryo EM	Cryo-electron microscopy
CSFV	Classical swine fever virus
DAA	Direct acting antiviral
ddGTP	2', 3' Dideoxy Guanosine triphosphate
DEPC	Diethyl pyrocarbonate
dH2O	Deionised water
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	2'Deoxyribonucleic acid
DTT	Dithiothreitol
E1	Envelope protein 1
E2	Envelope protein 2
EDTA	Ethylenediaminetetraacetic acid
EFLC	Equine foetal liver culture
eIF	Eukaryotic initiation factor
EM	Electron microscopy
EMCV	Encephalomyocarditis virus
EPgV	Equine pegivirus
EqHV	Equine hepacivirus
ER	Endoplasmic reticulum

eRF	Eukaryotic release factor
FBS	Foetal bovine serum
FDR	False discovery rate
FHK	Foetal horse kidney
FLuc	Firefly luciferase
FMDV	Foot and mouth disease virus
GAP	GTPase-activating protein
GHV	Guereza hepacivirus
GORS	Genome-scale ordered RNA structure
GTP	Guanosine 5' triphosphate
h.p.e	Hours post electroporation
h.p.t	Hours post transfection
HAV	Hepatitis A virus
HCV	Hepatitis C virus
HFLC	Human foetal liver cultures
HIV-1	Human immunodeficiency virus 1
IGR	Intergenic region
Ile	Isoleucine
Interferon	IFN
IP	Immunoprecipitation
IRES	Internal ribosome entry site
IVT	In vitro transcription
JFH	Japanese fulminant hepatitis
Kb	kilobases
KLD	Kinase/ligase/DNAse
LB	Lysogeny broth
LDLR	Low-density-lipoprotein receptor
LIPS	Luciferase-based immunoprecipitation system
LNA	Locked nucleic acid
m7G	7-methylguanylate
MAVS	Mitochondrial antiviral signalling protein
Met	Methionine

miR122	microRNA 122
mRNA	Messenger RNA
N/S	Not significant
NANBH	Non-A non-B hepatitis
NMIA	N-methyl isatoic anhydride
NPC1L1	Niemann–Pick C1-like 1
NPHV	Non-primate hepacivirus
NPT	Neomycin phosphotransferase
NrHV	Norway rat hepacivirus
OCLN	Occludin
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEI	Polyethylenimine
Phe	Phenylalanine
PIC	Pre-initiation complex
PLB	Passive lysis buffer
PV	Poliovirus
RdRP	RNA dependent RNA polymerase
RHV	Rodent hepaciviruses
RLuc	Renilla luciferase
RNA	Ribonucleic acid
RT	Reverse transcriptase
SCID	Severe combined immunodeficiency disease
SDM	Site-directed mutagenesis
SGR	Sub-genomic replicon
SHAPE	Selective 2'-hydroxyl acylation analysed by primer extension
SL	Stem loop
SRB1	Scavenger receptor class B member 1
SSIV	Superscript IV
ssRNA	Single stranded RNA

TC	Ternary complex
TFR1	Transferrin receptor 1
TLR	Toll-IL-1 receptor
TMT	Tandem mass tagged
TRIF	Toll-IL-1 receptor domain-containing adaptor inducing interferon-beta
tRNA	Transfer RNA
UTR	Untranslated region
WLSV	Wenling shark virus
WT	Wild type

Chapter 1 Introduction

1.1 Overview

The recent discovery of a number of animal hepaciviruses has provided an interesting and alternative means to investigate the molecular biology of the human pathogen hepatitis C virus (HCV). Of these newly discovered hepaciviruses, equine hepacivirus (EqHV) is phylogenetically the most closely related to HCV. An animal model of HCV has long been sought in the field; as such, investigating the molecular biology of the newly discovered hepaciviruses, and establishing replicative clones, is essential to further our understanding of HCV. This thesis describes an investigation into the 5' untranslated region (UTR) of EqHV; its RNA secondary structure, its function as an internal ribosome entry site (IRES), and its interactions with host cell machinery, as well as the rational construction and assay of an EqHV sub-genomic replicon (SGR).

To this end, the canonical pathway of eukaryotic translation initiation will be introduced and viral IRES will be described; specifically focussing on the type IV, HCV-like, IRES and the current state of knowledge of the EqHV IRES. Following this, the importance of viral RNA structure, and how RNA secondary structure is experimentally determined, will be introduced.

EqHV is closely related to HCV, and is predicted to have an identical genomic organisation. The EqHV proteins are also predicted to carry out identical functions to their HCV counterparts. As such the molecular biology of HCV, and the importance of the SGR, will be described. This will be followed by the initial discovery and characterisation of EqHV, its prevalence, transmission, and pathology, and the current state of knowledge regarding its molecular biology. This introduction will conclude with an overview of the other newly discovered hepaciviruses and their potential usefulness as a model system with which to investigate HCV.

1.2 Translation initiation

1.2.1 Cap-dependent initiation

Eukaryotic cap-dependent translation initiation minimally requires the 40S and 60S ribosomal subunits and eukaryotic initiation factors (eIFs) 1, 1A, 2, 2b, 3, 4A, 4B, 4E, 4F, 4G, 4H, 5, and 5B (Jackson et al., 2010). It is a two-step process comprising the formation of a 48S initiation complex, with Met-tRNA bound to the start codon in the 40S P-site, and the joining of a 60S ribosomal subunit; resulting in the formation of an elongation competent 80S ribosome.

Eukaryotic translation requires the recycling of post-termination complexes (post-TCs) to make ribosomal subunits available for a new round of translation initiation. Post-TCs consist of an 80S ribosome bound to mRNA, a deacylated tRNA in the P-site, and eukaryotic release factors (eRFs) 1 and 3. Dissociation of the post-TC is mediated by eukaryotic initiation factors (eIFs) 1, 1A, 3, and 3j and absolutely requires eIF3. eIF3j and mRNA bind to 40S subunits with negative cooperativity (Unbehaun et al., 2004; Fraser et al., 2007), therefore binding of eIF3j to the 40S subunit promotes dissociation from the mRNA. However, P-site codon-anticodon interaction protects the 40S-mRNA complex from eIF3j mediated dissociation, therefore eIF1 mediated release of the deacylated tRNA is required before 40S subunit dissociation can be achieved (Pisarev et al., 2007).

The 40S ribosomal subunit, in complex with eIF1, 1A, 3, and 3j, and the ternary complex (TC) eIF2-GTP-Met-tRNA, constitutes the 43S complex. eIF3, and eIF3j, are known to be positioned on the solvent facing side and mRNA binding channel (Siridechadilok, 2005; Fraser et al., 2007), respectively. Binding of eIF1 and eIF1A induces a conformational shift in the 40S subunit resulting in the opening of the mRNA entry channel. Both factors accelerate TC binding to 40S, with eIF1A stabilising this interaction (Passmore et al., 2007). The structured 5' untranslated regions (UTRs) of cellular mRNAs require unwinding in order to facilitate loading of the 43S complex. This is carried out by eIF4F, and eIF4B or eIF4H (homologous to a fragment of eIF4B). eIF4F consists of eIF4A, eIF4E, and eIF4G, and functions to recruit eIF4E, polyA binding protein (PABP), and eIF3. eIF4E binds the cap and eIF4G induces structural changes in eIF4E which increase its affinity for the cap (von der Haar et al., 2004). The activity of the DEAD-box helicase eIF1A is enhanced by eIF4G and eIF4B/H with eIF4G promoting the "closed", active, confirmation of the two eIF1A domains (Andersen, 2006; Rogers et al., 2001). It has been suggested eIF4B/H mediated stimulation of eIF4A helicase activity may be achieved through eIF4B/H binding single

stranded mRNA upstream of eIF1A to prevent it from re-annealing; promoting processive unidirectional translocation of eIF4A (Marintchev et al., 2009). Loading of the 43S complex onto the unwound mRNA therefore proceeds through eIF4G (in complex with eIF4E bound to the cap) interacting with eIF3 in the 43S complex (Lefebvre et al., 2006).

Loading of the 43S complex onto the mRNA is followed by ribosomal scanning of the downstream sequence to identify the start codon. eIF1 and eIF1A are both required for efficient scanning, indicating that the “open” 40S conformation induced by attachment of these factors is crucial in maintaining the scanning competent confirmation of 40S (Pestova and Kolupaeva, 2002; Passmore et al., 2007). Scanning of structured 5'UTRs requires ATP, eIF1A, eIF4G, and eIF4B. However, 43S complexes are able to scan unstructured 5'UTRs in their absence (Pestova and Kolupaeva, 2002). Two alternative modes of action have been postulated for the function of eIF4A, eIF4B, and eIF4G during scanning: either they drive the movement of mRNA by helicase mediated ratcheting, with the 40S subunit unwinding the RNA structure as it moves forward; or where they precede 40S, unwinding the RNA structure prior to it entering the mRNA binding channel (Siridechadilok, 2005; Marintchev et al., 2009).

Initiation codon recognition is highly specific for AUG triplets, usually the first downstream of the cap in the correct sequence context; GCC(A/G)CCAUGG, where -3 is a purine and +4 is a guanine (Kozak, 1991). eIF1 is essential to this process and functions to prevent Met-tRNA_i base pairing with non-AUG codons, AUG triplets which exist in a sub-optimal context, and those which exist <8 nucleotides from the 5' end of the mRNA; eIF1 also functions to dissociate complexes which have formed at one of these incorrect sites (T. V Pestova, Borukhov, et al., 1998; Pestova and Kolupaeva, 2002; Pisarev et al., 2006). Codon-anticodon base pairing is inhibited in the “open” conformation that promotes ribosomal scanning, therefore the 40S subunit must undergo conformational change to achieve this. Base pairing enhances the interaction between the 40S subunit and eIF1A, displacing eIF1 from the P-site and “closing” the ribosomal conformation. Furthermore eIF1 prevents eIF2 release from the TC by repressing hydrolysis of GTP bound eIF2 mediated by the GTPase-activating protein (GAP) eIF5 (Unbehaun et al., 2004; Lomakin et al., 2003; Maag et al., 2006; Cheung et al., 2007).

The final stage in translation initiation is the displacement of 48S associated eIFs, and joining of the 60S subunit to form an elongation competent 80S ribosome. This process is mediated by the ribosome dependent GTPase eIF5B which, in the presence of 60S, causes complete dissociation

of eIF2-GDP (Pisarev et al., 2006). An interaction between the C-terminal domains of eIF5B and eIF1A (Marintchev et al., 2003) is necessary for efficient subunit joining and promotes eIF5B-GTP hydrolysis; resulting in dissociation of the eIF5B-eIF1A complex and formation of the elongation competent 80S ribosome. Figure 1.1A is a model of the pathway of canonical eukaryotic translational initiation (reviewed in Jackson et al., 2010).

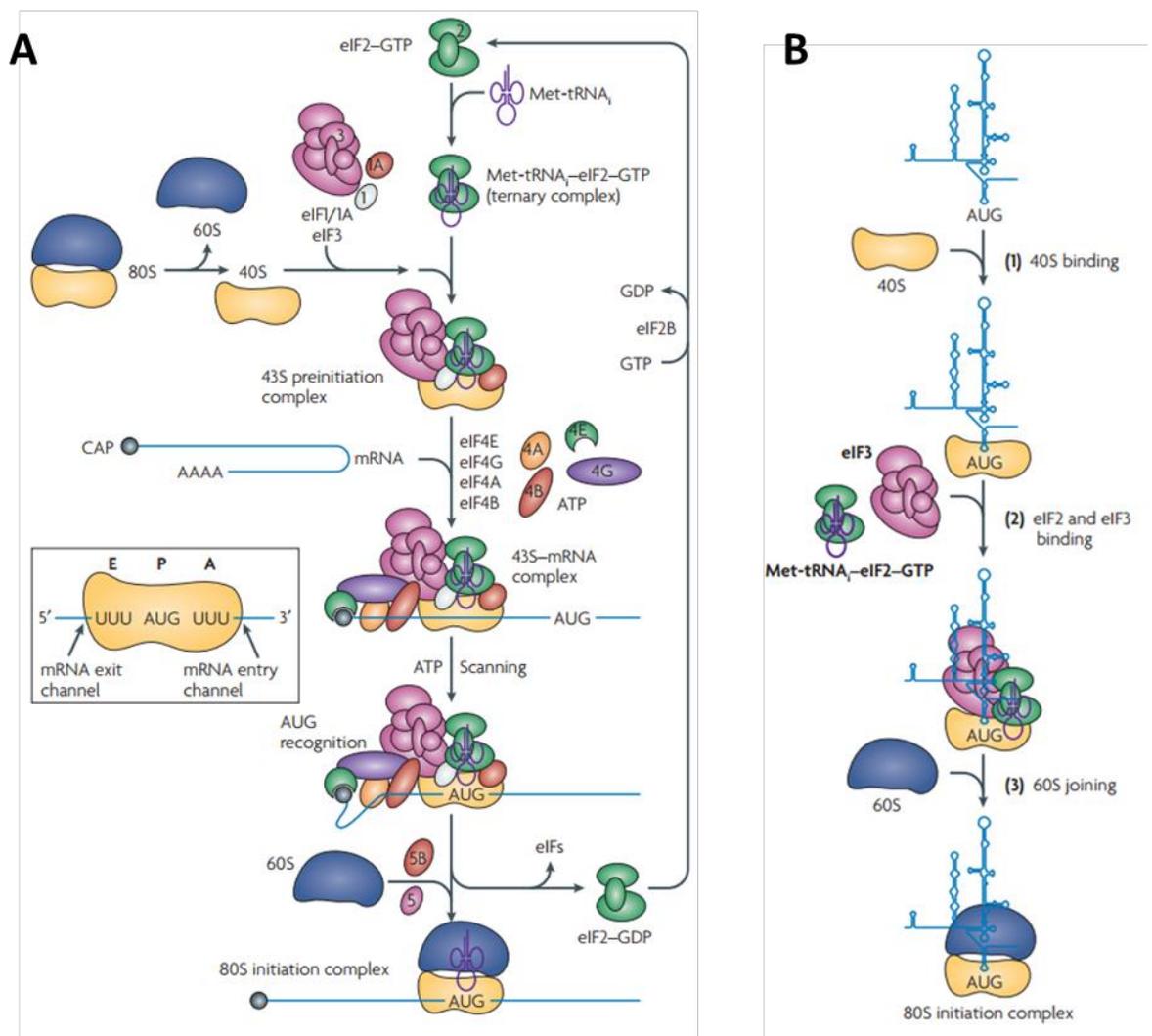


Figure 1.1: Canonical and HCV mediated translation initiation

A model of A) The canonical eukaryotic cap-dependent and B) HCV IRES-dependent translation initiation pathway (adapted from Fraser and Doudna, 2007).

1.2.2 Viral IRES

As obligate intracellular parasites, viruses rely on the host cell machinery to translate their genomes. To avoid the complex, and tightly regulated, canonical initiation pathway some viruses utilise internal ribosome entry sites (IRES). Some viruses require a subset of eIFs to mediate cap-independent initiation, others are able to directly recruit the 40S subunit, and some do not require any eIFs.

Viral IRES were first reported for poliovirus (PV) and encephalomyocarditis virus (EMCV) in 1988 (Jang et al., 1988; Pelletier and Sonenberg, 1988). Since then many other viral IRES have been reported and there is growing evidence for cellular mRNAs containing IRES, which may be involved in tumour survival (Martínez-Salas et al., 2012; Walters and Thompson, 2016). Viral IRES have since been classified into six types depending upon their requirement for host cell factors: picornavirus type I-V IRES (type IV IRES are also known as HCV like IRES), and intergenic region IRES (IGR IRES).

Type I IRES are exclusively found in enteroviruses and are epitomised by the PV IRES. This archetypal type I IRES is ~450 nucleotides in length and consists of domains II-VI of the PV 5'UTR. PV domain I forms a cloverleaf structure which is essential for replication but does not function in IRES mediated translation (Andino et al., 1990; Andino et al., 1993). During PV infection the 2A protease cleaves eIF4G to induce host cell shut off of eIF4E mediated cap-dependent translation (Gradi et al., 1998), therefore cap-independent translation is required to express the viral proteins. The minimal set of eIFs required for PV IRES mediated translation are eIF2, eIF3, eIF4A and the central domain of eIF4G; translation is strongly stimulated by eIF1A and eIF4B (Sweeney et al., 2014). The PV IRES directly interacts with the C-terminal cleavage product of eIF4G, which interacts with eIF4A and eIF3 (Gradi et al., 1998; Pestova et al., 1996; Pilipenko et al., 1992; de Breyne et al., 2009). Recruitment of eIF4A induces conformational changes around the 3' border of the IRES which has been suggested to enable 43S loading onto the domain VI AUG triplet (Kolupaeva et al., 2003; de Breyne et al., 2009). This AUG, however, is in poor context and translation is not initiated from this site. Therefore ribosome scanning proceeds through the ~160 nucleotide spacer region to initiate translation at the authentic AUG (Lozano and Martínez-Salas, 2015).

Type II IRES are also ~450 nucleotides in length and are found in many of the *Picornaviridae*. The most well-known, and well-studied are those of EMCV and foot and mouth disease virus (FMDV).

The first domain of these viruses is also involved in replication and is not required for IRES function; however these viral 5'UTRs are formed of four domains (I-IV) rather than the six domains of enteroviruses (Lozano and Martínez-Salas, 2015; Gao et al., 2016). Both EMCV and FMDV infections induce host cell shut off and utilise an IRES to circumvent inactivation of eIF4G by cleavage or phosphorylation (Belsham et al., 2000; Lee et al., 2017; Gingras et al., 1996). A role for EMCV 2A in sequestering eIF4E to inhibit cap-dependent translation has also been reported (Aminev et al., 2003). The type II IRES also harbour two in frame AUG start codons which, conversely to type I IRES, may both be used to initiate translation.

The only type III IRES is found in hepatitis A virus (HAV) IRES, which is both structurally and functionally distinct from type I and II IRES. Strikingly it has been previously demonstrated that this IRES requires intact eIF4G to function, is dependent on eIF4E, and can be inhibited by competition with m⁷G cap (Ali et al., 2001; Brown et al., 1994; Borman and Kean, 1997). However recent studies have shown that this may not be that case and that cleavage of eIF4G by the FMDV L protease in fact strongly stimulates HAV IRES mediated translation, and that translation proceeds when eIF2 α has been inactivated by phosphorylation. This suggests that in the presence of the FMDV L protease translation from this IRES proceeds in a similar manner to the type I and II IRES (Redondo et al., 2012).

Type IV IRES, or HCV like IRES, were originally discovered in HCV and Classical swine fever virus (CSFV) but have since been identified in a number of picornavirus genera (reviewed in Asnani et al., 2015). Whilst the HCV IRES contains four domains type IV IRES typically only contain three. These IRES are shorter than those of type I-III, but like with type I-II IRES domain one is not necessary for translation and is involved in genome replication (Khawaja et al., 2015). The structure and function of the HCV IRES is discussed in 1.2.3.

Type V IRES have only recently been classified and were described for the kobuvirus and proposed salivirus and paraturdivirus genera of the picornaviridae (Sweeney et al., 2012). They are of similar length (~410 nucleotides) to those of type I and II IRES and exhibit three major domains which are conserved throughout this class of IRES. Translation initiation proceeds through specific interaction with domain K, and requires the canonical factors eIF2, eIF3, eIF4A, and eIF4Gm.

Finally the IGR IRES, found in the *Discistroviridae* family, and exemplified by the cricket paralysis virus (CrPV) IRES, are a unique class of viral IRES which require no eIFs. Translation is initiated from a CUU, rather than AUG, triplet (Wilson et al., 2000). The CrPV IGR IRES is ~200 nucleotides in length and folds into three pseudoknots (Hertz and Thompson, 2011). It is located 6024 nucleotides downstream from the 5' end of the genome and is situated between two open reading frames (ORFs) (Sasaki and Nakashima, 1999). The lack of requirement for canonical eIFs or initiator tRNA is unique to this class of IRES (Thompson et al., 2001).

1.2.3 Hepaciviral IRES

Type IV, or HCV-like, IRES are found in members of both the *Picornaviridae* and *Flaviviridae*, with the archetypal IRES being that of HCV. The 5'UTR of equine hepatitis virus (EqHV), also known as canine hepatitis virus and non-primate hepatitis virus (CHV; NPHV) has also recently been described to function as an IRES and constitutes another type IV IRES from a hepatitis virus species (Figure 1.1A and C). The HCV 5'UTR is ~330 nucleotides long and comprises four major structural domains I-IV, which are comprised of SLI, SLII, SLIIa,b,c,d,e,f, a pseudoknot, and SLIV (Figure 1.2B). Domain I consists of a short stem loop which has been demonstrated, similarly to the picornaviral IRES, to function in replication; playing no role in translation. Domains II-IV constitute the IRES, however domains III and IV have been demonstrated to exhibit IRES activity in the absence of SLII (Ray and Das, 2004; Kalliampakou et al., 2002). The HCV IRES directly recruits the ribosomal 40S subunit and requires a minimal subset of canonical factors eIF3, eIF5, eIF5B, and the eIF2-GTP-Met-tRNA ternary complex (TC). There is no requirement for scanning in the HCV IRES, with SLIIId and the pseudoknot mediating loading of the ribosome directly onto the AUG start codon. The initial 42 nucleotides of the coding region also play a role in efficient HCV IRES mediated translation (Honda et al., 1996).

HCV IRES mediated translation initiation begins with direct recruitment of the 40S ribosomal subunit (Pestova et al., 1998) through an interaction between a GGG motif in the apical loop of SLIIId, and ¹¹¹⁶CCC₁₁₁₈ of the 40S subunit (Matsuda and Mauro, 2014; Malygin et al., 2013). Mutation of this motif has been demonstrated to both reduce the affinity of the IRES for 40S subunit, and to severely impair translation (Kieft et al., 2001; Jubin et al., 2000). Following formation of this binary complex, eIF2 is recruited to the 40S subunit, via an interaction with eIF3 which has been suggested to be dependent on IRES RNA-eIF3-eIF2 interactions, to correctly position Met-tRNA in the ribosomal peptidyl-tRNA site (P site) (Ji, 2004).

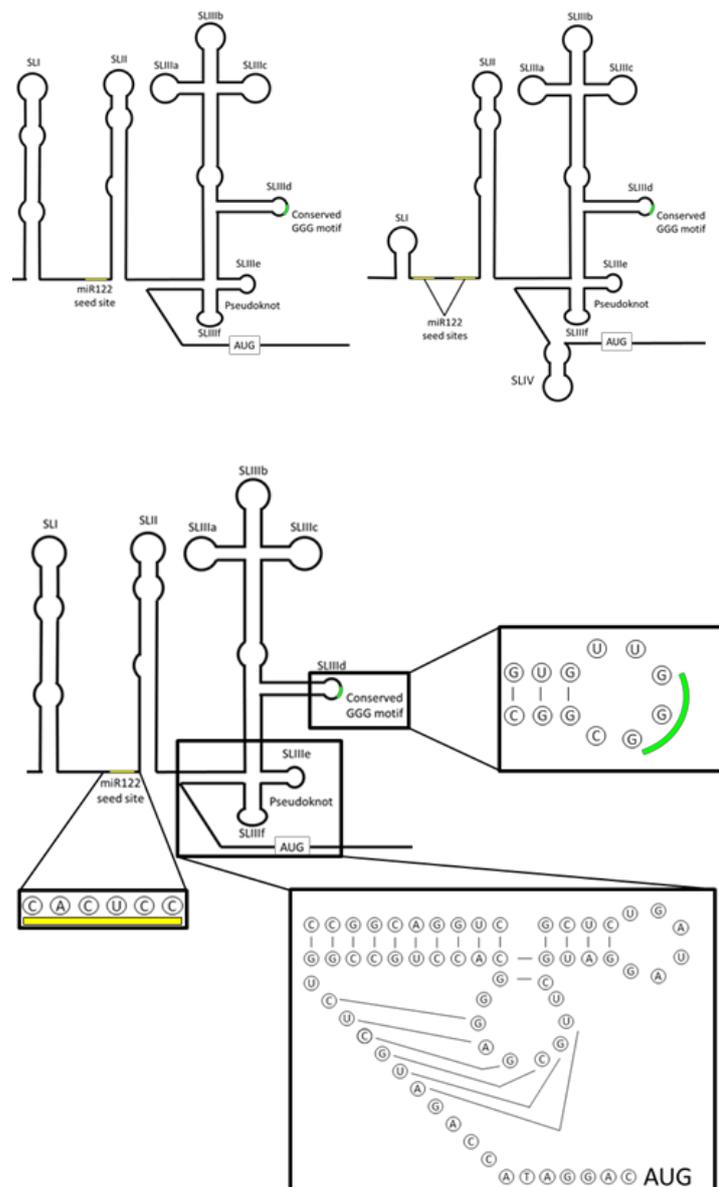


Figure 1.2: A Schematic representations of the predicted EHV, and HCV, 5'UTRs

A schematic representation of the secondary structure of the EHV 5'UTR. Secondary structure was predicted using EHV sequence NZPI (GenBank accession no. KP325401) in Mfold and a schematic representation of this was modelled. B) A schematic representation of the secondary structure of the HCV 5'UTR. miR122 seed sites are highlighted in yellow and the conserved GGG motif is highlighted in green, the start codon AUG is denoted in text. C) The predicted secondary structure of the EHV 5'UTR as in Figure 3.1A, with enlarged sections demonstrating the sequence of the conserved miR122 site between SLI and SLII, the conserved GGG motif in the apical loop of SLIIId, and the pseudoknot. The EHV pseudoknot was manually modelled based upon the Mfold prediction, the sequence of the EHV 5'UTR in this region, previous predictions of the EHV 5'UTR secondary structure, and the known structure of the HCV 5'UTR.

SLIIIb of the HCV IRES has been demonstrated to specifically interact with eIF3 via its apical loop, and a mismatched loop located in its stem (Sun et al., 2013; Sizova et al., 1998), and mutations in this region inhibit IRES activity (Kieft et al., 2001). Recent studies using cryo-electron microscopy (cryo-EM) have demonstrated that the HCV IRES interacts with the ribosome binding face of eIF3, causing displacement of the 40S subunit upon binding. This functions to promote the accumulation of free 40S, thus favouring HCV translation (Hashem et al., 2013).

Following ribosomal recruitment SLII contacts the ribosome to mediate a multitude of effects. This interaction causes a conformational change in the ribosome which closes the mRNA binding cleft (Spahn et al., 2001). SLII has also been implicated in the stable placement of mRNA in the decoding groove (Filbin and Kieft, 2011), release of eIF2 (Locker et al., 2007), the dissociation of eIF3j (Fraser et al., 2009), progression of 48S complexes to 80S ribosomes (Otto and Puglisi, 2004), and manipulates the ribosome to mediate a switch between translation initiation and elongation, promoting the first translocation event (Filbin et al., 2013). SLII is not involved in positioning the AUG start codon in the P-site (Locker et al., 2007). Correct positioning of the AUG in the mRNA binding cleft is mediated by the pseudoknot and mutations which destabilise the pseudoknot, or alter the distance of the AUG start from the pseudoknot, significantly impair translation (Berry et al., 2010). The final step in translation initiation from the HCV IRES is 60S subunit joining to form an elongation competent 80S ribosome correctly positioned on the AUG start codon. This proceeds in the same manner as canonical, cap-dependent, initiation and is mediated by the ribosome dependent GTPase eIF5B (Lee et al., 2002; Pestova et al., 1998; Yamamoto et al., 2014).

The only other hepaciviral IRES that has been described to date is that of EqHV (Stewart et al., 2013), however there is only limited information available regarding the structure and function of this type IV IRES (Kapoor et al., 2011; Burbelo et al., 2012; Stewart et al., 2013; Scheel et al., 2015). The EqHV 5'UTR demonstrates ~66% sequence identity with its HCV counterpart and a minimum free energy prediction of the 5'UTR showed a similar secondary structure organisation. SLII, SLIII, and the pseudoknot were predicted to exhibit secondary structure almost identical to that of HCV (Figures 3.1A and 4.2). Localised sequences of high conservation were also observed including the SLIIId GGG motif, and the pseudoknot. This analysis also revealed two major differences: a large and extended SLI, and a lack of SLIV (Burbelo et al., 2012;

Kapoor et al., 2011; Stewart et al., 2013). Functional analysis of the EqHV 5'UTR demonstrated that IRES activity was not affected by a deletion of SLI (Stewart et al., 2013). Furthermore, like HCV (Bradrick et al., 2006; Song et al., 2006; Jangra et al., 2010), EqHV IRES activity was enhanced in the presence of the 3'UTR, and the liver specific microRNA 122 (miR122); depletion of miR122 with locked nucleic acid abrogated this effect (Scheel et al., 2015). Given the structural and sequence conservation exhibited between these two, closely related, viral IRES it would be tempting to speculate that the mechanism of translation initiation (Figure 1.1B) is conserved between these two viruses, however this is yet to be determined.

1.3 RNA structure and determination

1.3.1 The folding of RNA into secondary and tertiary structures

RNA has an exceptional tendency towards folding into stable structures. As such, almost all RNAs, even the very shortest, form well defined and highly stable secondary structures. Secondary structure is the formation of helices and loops within an RNA and is largely driven by two factors: base pairing through hydrogen bonds, and base stacking. Base pairing occurs between AU and GC bases in the canonical Watson-Crick base pairing formation. However wobble (GU) base pairs are also frequently observed in native RNA structure. Triplet bases, the formation of base pairs between three bases, pseudoknots, kissing interactions, and other long range RNA-RNA interactions also contribute to the formation of RNA secondary structure. RNA secondary structure is inextricably linked to function. RNA stem loops have been extensively demonstrated to exhibit important functions, and whether a base, or motif, is in a single, or double, stranded conformation can significantly affect the ability of an RNA to carry out its function i.e. by disrupting RNA-protein interactions. RNA secondary structure is commonly represented in 2D (an example of which can be seen in Figure 1.2) and used to guide investigations into the link between RNA structure and function. To this end a suite of RNA secondary structure prediction tools, and programs, have been developed to predict how the secondary structures of RNA (discussed in 1.3.3).

RNA tertiary structure is somewhat more complex than secondary structure and constitutes the 3D conformation. The most common tertiary structure is the double helix, as described by Watson and Crick (Watson and Crick, 1953), but also include major and minor groove triplexes, quadruplexes, and coaxial stacking. However, the study of the 3D conformation of RNA is not straightforward and requires techniques such as x-ray crystallography and NMR.

The formation of these structures is driven by the same forces that dictate the formation of structure in a range of other biological molecules: hydrogen bonding, van der Waals forces, and the concealment of hydrophobic regions. The folding of RNA into secondary and tertiary structures, driven by these forces, ultimately results in energetically favourable, and stable conformations.

1.3.2 Viral RNA structure

Viruses are highly constrained with regards to the length, and therefore coding capacity, of their genomes. They must therefore employ novel methods to fully utilise the limited genetic material at their disposal in order to achieve replication, assembly, and release of fully infectious virus particles. There are a number of ways in which viruses achieve this: multifunctional proteins such as the HCV NS5A carry out numerous roles and thus reduce the need for individually encoded proteins with specific functions (Ross-Thriepland and Harris, 2015; Macdonald and Harris, 2004). Some viruses, such as the arenaviruses, employ ambisense genomes to increase their coding capacity (Auperin et al., 1984). Viruses are also able to “hijack” the cellular machinery to carry out viral functions; one such example being the translational machinery, as described in 1.2.2. RNA secondary structure is also an extremely versatile tool at the disposal of viruses to facilitate stages of the viral lifecycle without “wasting” precious coding capacity.

Viruses employ RNA structure throughout their lifecycles and the structure of viral RNA can be as important as its sequence. As described in 1.2.2 viral IRES substitute for the m⁷G cap and canonical eIFs; their structure is paramount to their function. RNA structure in the HCV genome is highly conserved and essential to the viral lifecycle (Fricke et al., 2015). It has also been demonstrated to function in packaging of the viral genome (Stewart et al., 2016). Human immunodeficiency virus 1 (HIV-1) also utilises RNA structure to package the viral genome (Clever et al., 1995). Genome-scale ordered RNA structure (GORS) have also been identified in positive sense RNA viruses and have been demonstrated to function in viral immune evasion; preventing the induction of interferon- β mediated through RIG-I and PKR activation (Witteveldt et al., 2013; Simmonds et al., 2004). In addition to the important, and necessary, functions that RNA structure plays throughout the viral lifecycle, the conservation of these structures further illustrates how important viral RNA structure truly is.

1.3.3 RNA structure prediction software

As discussed in 1.3.1 and 1.3.2 RNA has an exceptional tendency to fold into stable structures which play an essential role in viral lifecycles. Therefore understanding viral RNA structure is extremely important when investigating its function. However, experimental determination of these structures is a difficult, and lengthy, process; especially in cases where there is not a current model to work from. It is therefore often not feasible to experimentally determine RNA structure before undertaking an investigation into its function. Consequently, the prediction of RNA structure, using prediction software, is widely used to guide investigations into RNA structure and function. Furthermore, this software can also be utilised in combination with experimentally acquired data, such as SHAPE data, to refine structure prediction to accurately model RNA secondary structure (Deigan et al., 2009).

Two commonly used suites of prediction software are Mfold (Zuker, 2003) and RNAstructure (Reuter and Mathews, 2010), both of which are based on the Zuker algorithm for RNA structure prediction (Zuker, 1989) and utilise nearest neighbour free energy parameters. Furthermore, RNA folds in an extremely complex energy landscape which can produce myriad suboptimal structures and the algorithms employed by these suites of software predict a lowest free energy structure, and a range of structures within a given increment of free energy. The parameters of Watson-Crick helical regions have been well defined experimentally by investigating the thermodynamic properties of RNA sequences (Freier et al., 1986; Zhu and Wartell, 1997). The free energy algorithms, shaped by knowledge of RNA folding obtained experimentally, have been constantly refined and updated. This continual improvement, driven by ever increasing knowledge of RNA biology, has resulted in secondary structure prediction software which can efficiently predict RNA folding. These widely used programs are an essential tool in investigating RNA structure and link it to its function.

1.3.4 Selective 2'-hydroxyl acylation analysed by primer extension (SHAPE)

Numerous enzymes, and chemical reagents, have long been known to specifically modify or cleave nucleotides depending upon their single, or double, stranded conformation. These reagents have traditionally been used in combination to determine the secondary structure of RNA (Ziehler and Engelke, 2001). However, these methods of RNA structure probing have serious limitations. Many reagents lack specificity or exhibit preferential reactivity, and are base specific. Due to the nature of their action parallel experiments utilising different probes are

necessary to investigate the confirmation of every nucleotide. Furthermore, these techniques rely on technically challenging, and laborious, gel electrophoresis and densitometry to analyse the data.

Recent advances in reagents to probe nucleotide structure have yielded a set of chemical reagents known as SHAPE reagents, which are now widely in use for the probing of RNA secondary structure (McGinnis et al., 2012; Deigan et al., 2009). These reagents represent a significant improvement over the traditional RNA structure probing reagents as they simultaneously probe every nucleotide; reactivity is independent of nucleotide type (Wilkinson et al., 2009). Additionally, the advent of high throughput SHAPE and analysis by capillary electrophoresis (Wilkinson et al., 2008) has increased both the read length, and the number of samples which can be consecutively analysed. The inclusion of SHAPE data as a pseudo free energy constraint increases secondary structure prediction accuracy; *E. coli* 16S RNA was predicted to ~95% accuracy compared with the accepted model, up from 72% for traditional chemical probing methods. Inclusion of SHAPE reactivity data routinely results in base pair prediction accuracies >90% (Rice et al., 2014; Deigan et al., 2009; Hajdin et al., 2013).

SHAPE reagents preferentially react with the 2'-hydroxyl group of flexible (unpaired) nucleotides, forming bulky ester adducts with reactive bases (Figure 4.1). The reactivity of the 2'-ribose position is strongly modulated by the adjacent 3'-phosphodiester due to conformational constraints. SHAPE reagents have been demonstrated to be exquisitely sensitive to local nucleotide flexibility because of this (Merino et al., 2005). Following RNA labelling with a SHAPE reagent reverse transcriptase (RT) is employed to reverse transcribe the RNA into a cDNA copy; using a labelled primer to facilitate detection. The adduct formed by the SHAPE reagent causes termination of transcription at any reactive base. This results in different length cDNA transcripts which can then be separated and detected (Figure 4.1).

In order to identify the specific reactivity of each base, and produce a SHAPE reactivity profile for the RNA under investigation, it is necessary to carry out a dimethyl sulfoxide (DMSO) negative control, and a sequencing ladder, to determine background reactivity and correctly align the reactivity profile with the RNA sequence. There are a number of available programs with which to analyse SHAPE data such as CAFA, FAST, HiTRACE, SHAPE-CE, ShapeFinder, and QuSHAPE (Mitra et al., 2008; Vasa et al., 2008; Yoon et al., 2011; Pang et al., 2011; Aviran et al., 2011; Karabiber et al., 2013). Of these available programs QuSHAPE appears to be the most

attractive as it produces high quality data sets with a low level of user input. Processing of the data obtained from the capillary electrophoresis, and alignment to the sequencing ladder, results in normalised reactivities which are assigned to each base and can be used to influence RNA structure prediction. When used as a “soft” pseudo-free energy constraint, which guide rather than restrain folding (Lorenz et al., 2015), SHAPE reactivity data has been demonstrated to markedly increase base pairing prediction for known structures. This method significantly improves RNA structure predictions (Lorenz et al., 2015) and resulted in near-perfect predictions of RNA secondary structure when used in the RNAstructure software (Deigan et al., 2009; Mathews et al., 2004).

1.4 Hepatitis C virus

1.4.1 Prevalence, transmission and pathogenicity

HCV was first reported in 1989 as the major causative agent of non-A non-B hepatitis (NANBH) (Choo et al., 1989; Choo et al., 1990). It is a blood borne pathogen which is largely transmitted through contaminated needles used by intravenous drug users. However prior to widespread screening the majority of newly acquired infections were due to contaminated blood used for transfusion (Lauer and Walker, 2001). This route of transmission has now all but been eliminated in developed countries. Perinatal transmission is infrequent and is often associated with HIV-1 coinfection; sexual transmission of the virus is extremely inefficient (Thomas et al., 1998; Wyld et al., 1997; Ohto et al., 1994).

HCV is an extremely diverse virus and can be classified into seven distinct genotypes which vary by 30-35% at the nucleotide level. These are further classified into subtypes which vary by 20-25% at the nucleotide level. HCV genotypes also demonstrate uneven geographical distribution, and prevalence rates of the virus as a whole vary greatly between countries. This is most striking in Egypt which displays a much higher rate of infection, and prevalence of genotype 4 infections (Frank et al., 2000).

HCV is a global health concern with 71 million people estimated to be infected (WHO, 2017). Infection usually proceeds through an asymptomatic initial infection which will progress to chronic infection in ~75-85% of patients, with the remaining 15-25% spontaneously clearing the infection (Chen and Morgan, 2006). Without treatment chronic HCV infections will progress to liver cirrhosis and liver failure. Chronically infected patients with liver cirrhosis also have an increased chance of developing hepatocellular carcinoma.

1.4.2 Organisation of the HCV genome

HCV is a ~9.6 kb single stranded positive sense RNA virus. It contains one large open reading frame of ~9 kb which is translated into a large polyprotein which is cleaved by viral and host cell proteases to yield 10 individual proteins. The structural proteins: core, E1, E2, and p7, and the non-structural (NS) proteins: NS2, NS3, NS4A, NS4B, NS5A, and NS5B. The ORF is flanked by a 341 nucleotide 5'UTR, which contains the viral IRES, and a 3'UTR which contains conserved structures involved in the synthesis of negative sense genomes (Figure 1.2A).

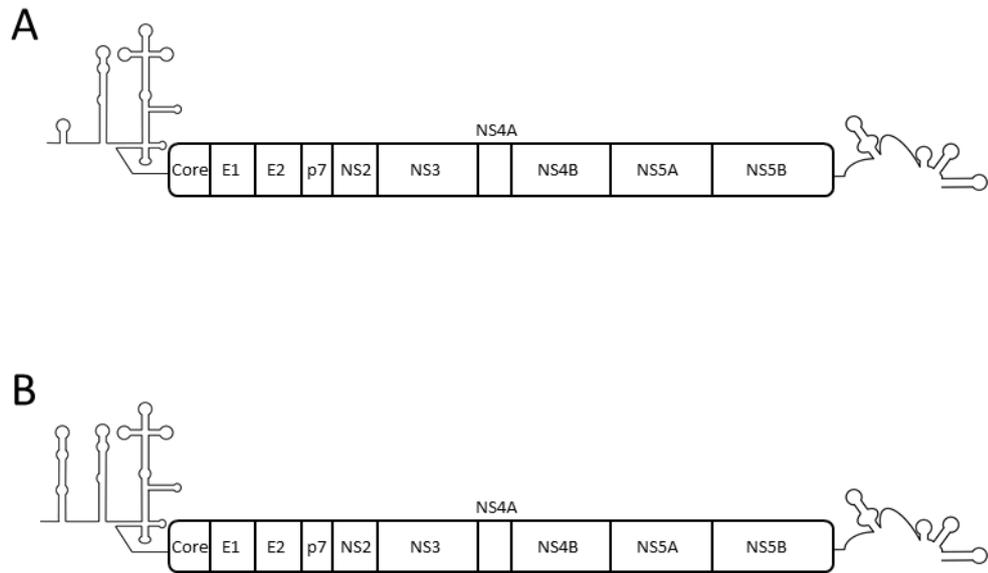


Figure 1.3: Hepaciviral genomic organisation

Genomic organisation of A) HCV and B) EqHV, displaying the 5' and 3' untranslated regions and the individual proteins expressed following cleavage of the polyprotein.

1.4.3 The HCV viral particle and entry

The HCV genomic RNA is encapsidated in a nucleocapsid particle formed of the core protein and enveloped in host derived membranes. It is 40-80 nm in diameter, pleomorphic, and E1 and E2 are exposed upon the surface (Gastaminza et al., 2010; Catanese et al., 2013). HCV viral particles exhibit extremely low buoyant density due to their association with serum lipoproteins and are also known as lipoviral particles; this has been postulated to play a role in immune evasion (Bartenschlager et al., 2011).

HCV entry begins with interactions between particle associated apolipoprotein E (apoE) and proceeds through clathrin mediated endocytosis to fusion of the viral and endosomal membranes and genome release into the cytoplasm. However, whilst this may seem a familiar story the mechanism of HCV entry is somewhat more complex.

Initial viral attachment to the cell membrane is mediated by particle associated apoE interactions with low-density-lipoprotein receptor (LDLR) and with glycosaminoglycans (GAGs)

present on the cell surface (Monazahian et al., 1999; Agnello et al., 1999; Germi et al., 2002). Five cell surface receptors are required for efficient entry of the HCV particle into the cell: CD81, scavenger receptor class B member 1 (SRB1), claudin 1 (CLDN1), occludin (OCLN) and Niemann–Pick C1-like 1 (NPC1L1) (Pileri et al., 1998; Scarselli et al., 2002; Evans et al., 2007; Ploss et al., 2009; Sainz et al., 2012). Transferrin receptor 1 (TFR1) has also been demonstrated to play an important role after CD81 binding, although the exact role of this protein in HCV entry is yet to be fully elucidated (Martin and Uprichard, 2013). Following particle attachment E2 binds CD81 and this complex is trafficked to CLDN1 tight junctions, where clathrin mediated endocytosis is induced (Farquhar et al., 2012). Following endocytosis, acidification of the endosome, and membrane fusion between the viral and endosomal membranes, the HCV genome is released into the cytoplasm and viral replication may begin (reviewed in Lindenbach and Rice, 2013).

1.4.4 Polyprotein expression and processing, and replication

Upon release into the cytoplasm the positive sense RNA genome of HCV can be directly translated. Translation is initiated by direct recruitment of the 40S ribosomal subunit by the HCV IRES and progresses to an 80S elongation competent ribosome positioned on the AUG start codon, requiring only a subset of canonical eIFs (described in 1.2.3). Translation takes place at the rough endoplasmic reticulum (ER) and the polyprotein is co- and post-translationally cleaved by cellular and viral proteases. Signal peptidase liberates core at the core-E1 junction and further cleavages at the E1-E2, E2-p7, and p7-NS2 junctions liberate all of the structural proteins (Hijikata et al., 1993; McLauchlan, 2000; Griffin et al., 2005). The NS2 auto protease cleaves at the NS2-NS3 junction to liberate NS2, and the remaining NS proteins are liberated by NS3-4A mediated cleavage (Bartenschlager et al., 1994).

HCV replication, like other positive stranded RNA viruses, occurs in host cell derived membrane complexes (Mackenzie, 2005; Salonen et al., 2005). During HCV replication NS4B induces the formation of single, double and multi-membrane vesicles which has been termed the membranous web (Egger et al., 2002). This is the site of replication complexes, formed of the NS proteins 3-5B which are sufficient for HCV replication (Lohmann et al., 1999).

In addition to its role in cleaving the HCV polyprotein, NS3-4A also plays a role in immune evasion by cleaving cellular mitochondrial antiviral signalling protein (MAVS) and toll-IL-1 receptor (TLR) domain-containing adaptor inducing interferon-beta (TRIF) to disrupt the RIG-I and TLR-signalling pathways (Meylan et al., 2005; Li et al., 2005). The C-terminal domain of NS3 has also

been demonstrated to possess helicase activity, driving unidirectional 3'-5' unwinding of HCV genomic RNA (Kim et al., 1995; Tai et al., 1996). NS4A contributes to correct folding of the NS3 protease and also acts as a membrane anchor (Bartenschlager et al., 1995).

NS5A is a multifunctional phosphoprotein which carries out a range of functions during the HCV lifecycle (reviewed in Ross-Thriepland and Harris, 2015) and the well-known culture adaptive mutation S2204I is located in this viral protein (Blight et al., 2000). NS5A is anchored to the membrane by its N-terminal helix and is an essential component of the replication complex; interacting with both RNA and proteins to exhibit myriad functions.

NS5B is the viral RNA dependent RNA polymerase (RdRP) carrying out synthesis of both the positive sense genome and negative sense replicative intermediate, and contains the characteristic GDD active site; which is frequently mutated to GND/GNN to create a replication deficient control. Like NS5A it is also membrane anchored but this occurs in the C-terminal domain of NS5B. Anchoring has been shown to be essential to *in vivo* function (Moradpour et al., 2004).

The HCV 5' and 3' UTRs have also been demonstrated to function in replication. SLI of the HCV 5'UTR is essential for replication and the proximal 120 nucleotides are sufficient; however the full 5'UTR enhances replication (Luo et al., 2003; Friebe et al., 2001). Likewise the 150 nucleotide 3' portion of the 3'UTR has also been demonstrated to be essential for replication, with the 5' portion shown to enhance replication (Yi and Lemon, 2003).

1.4.5 Assembly and release

The mechanism of viral assembly and release is not well understood. Particles assemble by budding into the ER and p7, NS2, NS3/4A, and NS5A are all involved in the recruitment of core to the site of viral assembly (Counihan et al., 2011). Particles are transited through the secretory pathway resulting in release from the cell by exocytosis (Gastaminza et al., 2008). During secretion the HCV particle acquires low density lipoproteins and the glycoproteins undergo post translational modification to contain complex glycans (Vieyres et al., 2010). There is also a role for p7 in protecting the particles from low pH during their maturation and secretion (Bentham et al., 2013; Wozniak et al., 2010; reviewed in Lindenbach and Rice, 2013).

1.4.6 The HCV sub genomic replicon

For 10 years following the first identification of HCV, replication in cell culture systems could not be achieved. Therefore investigations into the molecular biology of HCV was limited to functions which could be reproduced *in vitro* such as the enzymatic activity of NS3 and 5B, and the function of the 5'UTR IRES. The advent of a replicative clone of HCV in cell culture represented a major step forward in HCV research. It has been instrumental in understanding the molecular biology of the virus and the development of direct acting antivirals (DAAs). *In vitro* replication was first achieved through the construction of a consensus clone sub genomic replicon (SGR) Con1b (Lohmann et al., 1999). The HCV SGR consists of the 5'UTR which drives expression of a reporter gene, the EMCV IRES which drives expression of the NS3-5B coding region, and the 3'UTR (Figure 5.1A).

The first cell line in which HCV replication was achieved was the human hepatoma cell line Huh7 (Nakabayashi et al., 1982). In the intervening years Huh7, and the Huh7 derived Huh7.5, cells have remained the cell line of choice for much of the work in the field. Huh7.5 cells were created by curing Huh7 cells which stably harboured an SGR with interferon (IFN) and are defective in the innate immune pathway RIG-I due to a dominant negative T55I mutation which impairs RIG-I signalling (Blight et al., 2002).

Whilst Huh7 and Huh7.5 cells are commonly the cell lines of choice to study HCV, a number of liver, and non-liver, derived cell lines have been shown to support HCV replication: stomach-derived FU97, liver derived Hep3B and Hepg2 cells, uterine Hec1B, mouse embryonic fibroblasts (MEFs), human embryonic kidney 293T, and cervical carcinoma HeLa cells (Shiokawa et al., 2014; Kambara et al., 2012; Fukuhara et al., 2012; Lin et al., 2010; Kato et al., 2005) have all been demonstrated to support HCV SGR replication, with a common dependence on the expression of miR122, a liver specific microRNA known to be essential to the HCV lifecycle (Jopling et al., 2005).

Whilst there are a large number of cell lines that have been demonstrated to support HCV replication the pool of HCV isolates that have been able to replicate in mammalian cell culture is somewhat limited. Only the JFH-1 isolate is able to efficiently replicate without the need for culture adaptive mutations. However there has been progress made regarding the diversity of viral genotypes that are able to replicate in mammalian cell culture with replication capable constructs for genotypes 1a, 1b, 2a, 3a, 4a, 5a, and 6a. However many of these isolates require

cell culture adaptive mutations, such as the well-known S2204I, to be introduced before replication can be established (Saeed et al., 2012; Kinge et al., 2014; Yu et al., 2014; Lohmann et al., 1999; Kato et al., 2003; Blight et al., 2003; reviewed in Lohmann and Bartenschlager, 2014).

Whilst a wide variety of genotypes are now available to study in cell culture, the isolation and study of patient derived sequences is still extremely challenging, and patient derived viruses do not readily replicate in cultured cells. However the recent discovery of pro-viral properties of the human gene SEC14L2 has led to pan genotypic replication of patient derived samples in cell culture. It has been suggested that this pro-viral gene promotes HCV infection by enhancing vitamin E-mediated protection against lipid peroxidation (Saeed et al., 2015).

The advent of the HCV SGR represented a huge leap forward in HCV research and has facilitated a litany of discoveries that have culminated in the development of pan-genotypic direct acting antivirals (DAAs). These new therapies have ushered in a new era of IFN free HCV therapy which have achieved up to 100% cure rates (reviewed in Zhang et al., 2016).

1.5 Equine hepacivirus

1.5.1 Discovery and initial characterisation of EqHV

The first novel hepacivirus to be identified was canine hepacivirus (CHV) in 2011 (Kapoor et al., 2011). Using unbiased high throughput screening respiratory samples of dogs, associated with respiratory illness, were screened for the presence of viral RNA. This returned a predicted amino acid sequence of viruses similar to flaviviruses. Further amplification revealed a highly convergent (99.2%) virus, closely related to HCV, which was detected in the nasal swabs of 9 dogs; with viral copies $>1 \times 10^7$. Further investigation of healthy dogs yielded no further detection of CHV, however screening of dogs that had died of unexplained gastrointestinal illness showed low levels ($<1 \times 10^3$) of CHV RNA that could only be identified from liver samples, which were also demonstrated to contain viral RNA by in situ hybridization.

In a follow up study conducted by the same group a luciferase-based immunoprecipitation system (LIPS) seroscreening approach was utilised to screen 80 dogs, 81 deer, 84 cows, 103 horses, and 14 rabbits for the presence of CHV NS3 reactive IgG (Burbelo et al., 2012). Surprisingly no dog samples were seroreactive to CHV NS3, however 35% of the horse samples were, and genetically diverse hepaciviral RNA was detected in 8 of these samples. Due to the unresolved nature of the true host of the virus originally named CHV, and the genetic diversity

of the newly isolated viral sequences, the authors putatively named them non-primate hepaciviruses (NPHVs).

In both of the aforementioned studies the viral genomes, excluding the full CHV 3'UTR, were sequenced and characterised based upon sequence conservation, predicted amino acid (a.a) sequence, and RNA secondary structure. Kapoor *et al.* reported the CHV genome to be at least 9,195 nucleotides in length with a 366 nucleotide 5'UTR and a 2,942 a.a polyprotein. Whilst they were unable to sequence the CHV 3'UTR the authors did note the possibility of a poly(A) stretch at the 3' end of the genome; an unusual finding unique amongst hepaciviruses (Kapoor *et al.*, 2011). poly(A) tails of variable lengths were also reported for the NPHVs (Burbelo *et al.*, 2012). However a separate study by Scheel *et al.* reported a full length NPHV 3'UTR of 328 nucleotides, including a short poly(A) tract, a variable region, a poly(U/C) tract, a conserved intermediate region, a long poly(U) tract, and a conserved 3'X region; which resembled the HCV 3'UTR in its downstream sequence and structure (Scheel *et al.*, 2015). The polyprotein cleavage map for CHV was predicted by alignment with HCV sequences and revealed conserved signalase, NS2/3, and NS3/4A cleavage motifs which would yield 10 mature proteins from the CHV polyprotein with conserved function with HCV: core, E1, E2, p7, NS2, NS3, NS4A, NS4B, NS5A, NS5B. Maximum a.a identity was observed between CHV and HCV NS3 and NS5B (55-65%) and the lowest identity was between E1, NS2, and NS5A, mimicking that which is seen between HCV genotypes. This study also demonstrated that CHV is more similar at the a.a level to HCV than it is to GBV-B, but varied more from HCV than HCV genotypes do from each other (Kapoor *et al.*, 2011). Burbelo *et al.* noted that NPHV isolates showed a moderate, 14%, sequence divergence from each other over the length of the genome, lower than the mean diversity observed between HCV subtypes (Burbelo *et al.*, 2012).

Initial prediction of the secondary structure of the CHV 5'UTR was conducted by Kapoor *et al.* and refined by Burbelo *et al.* who identified a further 17 nucleotides in the NPHV 5'UTRs and hypothesised that the CHV 5'UTR sequence was incomplete (Kapoor *et al.*, 2011; Burbelo *et al.*, 2012). This analysis revealed that the CHV 5'UTR resembled a type IV, HCV-like, IRES and contained four domains (SLI, SLII, SLIIIa-f, and a pseudoknot; Figure 3.1A and B). SLII, SLIII, and the pseudoknot highly resembled the HCV 5'UTR structure with sequence conservation most prevalent in SLIIIc, e, and f. Differences in the 5'UTR structures were most notable in the 5' region with a large and extended SLI in CHV which is absent in HCV. This also revealed the presence of an intermediate SL (herein termed SLIA) with sequence homology to the unpaired HCV miR122

seed site one. The CHV 5'UTR was also not predicted to contain SLIV, which has been demonstrated in HCV to play a role in correct ribosomal positioning (Kapoor et al., 2011; Berry et al., 2010). Using the multiple sequences from NPHVs Burbelo *et al.* further verified and refined the NPHV 5'UTR secondary structure. This revealed an extended SLI in comparison to CHV and further confirmed the conservation of sequences known to be important in ribosome binding and translation initiation from the HCV IRES; most notably the GGG motif in SLIIId and the pseudoknot region (Burbelo et al., 2012). Both studies also reported RNA secondary structure in the coding region of CHV/NPHVs concluding that the CHV/NPHV genomes contain GORS, indicative of viral persistence (Kapoor et al., 2011; Burbelo et al., 2012; Simmonds et al., 2004).

1.5.2 Prevalence, host range, and classification, of EqHV

Due to the unresolved host range of CHV/NPHVs (herein termed NPHVs) a number of studies set out to assess the host range and prevalence of these newly discovered viruses. These studies largely focused on dogs and horses as host species. Attempts to isolate NPHVs from dogs proved largely unsuccessful (Bexfield et al., 2014; Lyons et al., 2012; Burbelo et al., 2012; Van Der Laan et al., 2014) However one report isolated NPHVs from both tracheal and liver samples of UK dogs (El-Attar et al., 2015). It has therefore been suggested that this may represent a recent cross species transmission event, or even a false transmission event (Hartlage et al., 2016). Given the prevalence of NPHV sequences in horses, and commercial horse serum (Postel et al., 2016), whether dogs represent a natural host for NPHVs warrants further investigation.

Identification of NPHVs from dogs remains a rarity; however, studies that have focused on isolating NPHVs from equine species have had a consistent level of success, indicating that equines represent the natural host of NPHVs. For this reason newly identified viruses, which would previously have been denoted as CHV/NPHV, are frequently being referred to as equine hepaciviruses (EqHVs). This nomenclature will be used henceforth to describe hepaciviral isolates from both equines and dogs. EqHV has been isolated from equines worldwide with reports from the UK, US, Japan, China, Germany, Korea, France, Brazil, Italy, and Hungary (Lyons et al., 2012; Burbelo et al., 2012; Matsuu et al., 2015; Lu et al., 2016; Drexler et al., 2013; Pronost et al., 2016; Figueiredo et al., 2015; Reuter et al., 2014; Kim et al., 2017; Gabriella Elia et al., 2017; Pfaender et al., 2015). However prevalence of EqHV does appear to vary between regions, with 34-43% seroprevalence reported for tested horses and up to 14% of seropositive horses also testing positive for viral RNA (Lyons et al., 2014; Burbelo et al., 2012; Matsuu et al., 2015).

EqHV appears to exhibit a much higher seroprevalence than HCV with a lower prevalence of active infection. This is suggestive that EqHV either has a more efficient route of transmission than HCV, or it may be that the numbers of infected horses are artificially high due to infection from contaminated sera (Postel et al., 2016).

The EqHV viral isolates sequenced to date appear to demonstrate less genetic diversity than the closely related HCV; as such EqHV has not yet been classified into viral genotypes as are established for HCV. Whilst the circulation of two viral subtypes has been suggested (Pronost et al., 2016) viral sequences did not diverge >20% as would be expected for an HCV subtype and it is likely that more viral isolates will be required to definitively classify EqHV.

1.5.3 Transmission, tissue tropism, and pathology, of EqHV

The natural route of transmission of EqHV is yet to be fully elucidated. One study has demonstrated vertical transmission from an EqHV positive mare to foal; transmission was associated with high viremia at birth (Gather et al., 2016). This study also noted a greater susceptibility to EqHV infection following a decrease in EqHV NS3 reactive antibodies in foals, indicating a protective effect of antibodies transferred from mare to foal. Furthermore the authors also observed potential horizontal transmission of EqHV, and the infection of a foal which received a plasma transfusion with EqHV positive plasma. Whilst vertical transmission occurs for EqHV it does not appear to represent the main route of transmission; arthropod borne infection and contaminated medical products may therefore still represent major routes of EqHV infection (Burbelo et al., 2012; Postel et al., 2016).

Whilst the route of transmission has not yet been fully elucidated, EqHV appears to display distinct liver tropism. EqHV RNA is most concentrated within the liver and the replicative intermediate negative strand genome has been exclusively identified in liver tissue (Pfaender et al., 2015; Scheel et al., 2015; Ramsay et al., 2015). These findings are further supported by evidence for increased liver enzyme levels, liver damage, and membrane rearrangements, in the livers of infected horses (Lyons et al., 2012; Scheel et al., 2015; Pfaender et al., 2015; Pfaender et al., 2017). However reports on the pathology of EqHV infection are contradictory and the true effect of EqHV infection is still to be determined. Whilst many studies have reported only mild increases in liver enzyme levels, that remain within the reference range, some have demonstrated significant pathology upon infection with EqHV including naturally infected horses with hepatitis (Reuter et al., 2014) and chronic wasting (Elia et al., 2017). One study

reported that horses exposed to EqHV via blood transfusions developed fulminant hepatitis, however co-infection with equine pegivirus (EPgV) was also present (Ramsay et al., 2015). This study also demonstrated liver damage in normal adult horses experimentally infected with EqHV; inoculation of SCID horses resulted in only mild liver disease. Chronic infection of horses has also been demonstrated with EqHV RNA detectable for >1 year following infection. EqHV also appears to be cleared much more efficiently than HCV in humans; the majority of EqHV infections are cleared in association with the onset of antibodies reactive to NS3 (Gather et al., 2016; Elia et al., 2017; Scheel et al., 2015).

The true transmission and pathology of EqHV is yet to be fully elucidated, however it appears likely that EqHV is a blood borne disease that has the potential to cause hepatitis in infected horses. EqHV is also able to establish chronic infection in horses however, in contrast to HCV, viral clearance appears to occur in the majority of infections. EqHV demonstrates both marked similarities and differences with HCV with regards to the progress of infection and the disease which it causes. Whilst many questions still remain to be answered, it appears as if EqHV may be an intriguing model with which to study HCV.

1.5.4 Molecular biology of EqHV – the story so far

Little is known about the molecular biology of the EqHV RNA structure and proteins, how they function, and how they interact with the host cell. Progress has largely been impeded by a lack of an *in vitro* system with which to study the viral lifecycle but progress is being made. The EqHV 5'UTR has been extensively mapped based on sequence and phylogenetic conservation (Burbelo et al., 2012). Stewart *et al.* demonstrated that the EqHV 5'UTR functions as an IRES, and that SLI is not required to initiate translation (Stewart et al., 2013). A separate study also demonstrated that EqHV IRES activity is enhanced by miR122, and the EqHV 3'UTR (Scheel et al., 2015). However the *in vitro* secondary structure of the 5'UTR, or the reported GORS (Kapoor et al., 2011), are yet to be experimentally confirmed. The core protein of EqHV has been suggested to be cleaved by signal peptidase and localises to lipid droplets, as is seen for HCV. Analogous mutation of key Ile and Phe residues inhibits core cleavage (Tanaka et al., 2014). The EqHV NS3/4A serine protease has also been demonstrated to cleave human MAVS and TRIF, like its HCV counterpart, and also shows specificity for both the HCV and EqHV NS5A/5B cleavage site. Interestingly this study also reported that two different HCV NS3/4A protease inhibitors were unable to inhibit the activity of the EqHV NS3/4A protein (Parera et al., 2012). Furthermore, EqHV infection has been demonstrated to induce membrane rearrangements in infected cells reminiscent of those observed during HCV infection. Whilst it has not been experimentally confirmed the similarity of the two viruses would suggest a role for NS4B in mediating these changes (Pfaender et al., 2017).

Although research into the molecular biology of EqHV is still evidently in the early stages, similarities between EqHV and HCV are beginning to be elucidated. The predicted function of EqHV proteins is highly likely to be accurate, and the predicted activity of the 5'UTR as an IRES has already been confirmed. Furthermore, miR122 also plays a role in EqHV translation, and it is likely due to the seed site located in the EqHV IRES, although this is as yet unconfirmed. This would agree well with the conservation of this site observed between EqHV and HCV, and the liver tropism of EqHV. However, the inability of HCV inhibitors to inhibit the activity of the EqHV NS3/4A protease suggests that key differences between these viruses are there to be discovered. Whilst much research into the molecular biology of EqHV is likely to confirm predictions regarding conserved functions with HCV, finding differences could be key to understanding the differential pathology of these viruses.

1.6 Other newly discovered hepaciviruses

Since the first isolation of HCV as the causative agent of NANBH there has been an ongoing search to identify animal models to study this clinically important virus. Early investigations focused on the identification of primate hepaciviruses but proved fruitless; no new hepaciviruses were isolated for over 20 years. However the use of deep sequencing in a range of hosts has revealed a broad range of animal hepacivirus. This recent boom in hepacivirus discovery has expanded this genus from 2 to 14 species in just 5 years; providing an ever growing pool of isolates with the potential to become the long awaited HCV animal model.

1.6.1 Genomic organisation of the newly discovered hepaciviruses

The first novel hepacivirus, since the discovery of HCV, to be isolated was equine hepacivirus (EqHV) in 2011 (Kapoor et al., 2011). This discovery prompted investigations to identify further hepacivirus species in a range of hosts by next generation sequencing. This approach led to the identification of hepaciviral RNA in bats, bovines, equines, rodents, a new-world primate, and the graceful cat shark (Quan et al., 2012; Baechlein et al., 2015; Corman et al., 2015; Burbelo et al., 2012; Lyons et al., 2012; Lauck et al., 2013; Drexler et al., 2013; Kapoor et al., 2013; Shi et al., 2015).

Isolation of the bat hepacivirus (BHV) 5'UTR yielded sequences of varying length with the shortest being 156 nucleotides and the longest 445. It is therefore likely that this 5'UTR represents an elongated sequence (the EqHV 5'UTR is 385 and HCV ~330 nucleotides in length) in comparison to other hepaciviral 5'UTRs. However secondary structure prediction shows that the BHV 5'UTR contains a conserved HCV-like IRES element SLIII which harbours a conserved GGG motif in the apical loop of SLIIId (Quan et al., 2012). The bovine hepacivirus (BovHepV) 5'UTR is 294 nucleotides in length and is characteristic of an HCV-like IRES. SLIIIa,b,c,d,e are all conserved, with a GGG motif in the apical loop of SLIIId, and a miR122 seed site located directly upstream of SLII. However the BovHepV 5'UTR does not appear to contain a homologous SLI despite repeated attempts to isolate this sequence (Corman et al., 2015; Baechlein et al., 2015). Hepacivirus genomes isolated from bank voles demonstrated marked variability in their 5'UTRs which displayed characteristics of both hepaci- and pegi- virus IRES elements. One isolate (SAR-46) displayed the typical HCV-like IRES SLIII structure with the conserved GGG motif presented in the apical loop of SLIIId (Drexler et al., 2013). The 5'UTR of hepaciviral isolates from deer mice also displayed the HCV-like IRES SLIII with SLIIIa-f and the pseudoknot all conserved. However,

the 5' sequence of these UTRs could not be matched with any known sequences in genbank. All of the HCV-like IRES elements did however contain conserved miR122 seed sites (Kapoor et al., 2013; Drexler et al., 2013). The 5'UTRs of both of the Guereza hepacivirus (GHV) isolates displayed limited homology to the other hepacivirus UTRs in their 5' sequences, returning no matches to sequences in the genbank database. However the 99 nucleotides directly upstream of the coding region displayed analogous structural features related to SLIIId-f of HCV-like IRES with the characteristic conserved GGG motif presented in the apical loop of SLIIId. A conserved miR122 seed site was also identified within the 5' sequence of both GHV isolates (Lauck et al., 2013). The 5'UTR of wenling shark virus (WLSV) was not characterised but was shown to be only 132 nucleotides in length indicating that the full 5'UTR may not have been sequenced (Shi et al., 2015). The 5'UTRs of the Norway rat hepaciviruses (NrHVs) were not characterised (Firth et al., 2014).

Predictions based upon the sequences of new hepaciviral isolates have all demonstrated a genomic organisation which mimics that of HCV and EqHV, one large polyprotein encoding 10 proteins: core, E1, E2, p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B. Sequence alignment and homology were used to identify the protein boundaries of non-structural proteins predicted to be cleaved by NS3/4A. Signal peptidase cleavage sites have been identified in the structural proteins which would yield structural and non-structural proteins of similar size to HCV, with predicted identical functions (Quan et al., 2012; Baechlein et al., 2015; Kapoor et al., 2013; Lauck et al., 2013; Shi et al., 2015). Interestingly, whilst most of the newly discovered hepaciviruses exhibit ORFs shorter than that of HCV (typically ~2700 a.a compared with ~3000 a.a.) the GHV ORF was predicted to encode a polyprotein ~3300 a.a in length. This is due to an unusually long and disordered NS5A protein which is over twice as long as known NS5A proteins.

The new hepacivirus isolates, like HCV, exhibited high levels of genetic diversity within the E proteins and NS5A; these proteins exhibited the highest levels of sequence divergence, in some cases up to 85% (Kapoor et al., 2013). NS3 and NS5B exhibited the highest levels of conservation, up to 75.4% (Baechlein et al., 2015). Furthermore many of the viral isolates displayed sequence diversity from one and other reminiscent of HCV subtypes (~20%). However between species viral sequences exhibited expectedly higher levels of sequence divergence of >40% and up to 70% in the case of WLSV (Shi et al., 2015).

All of the hepaciviruses are also predicted to have a 3'UTR. Obtaining the full sequences of hepaciviral 3'UTRs is difficult and thus the majority of the newly discovered viruses have not yet been fully characterised in this region. Some bovine isolates have been sequenced and predicted to form three SLs which resemble the HCV 3'UTR, however they did not contain a poly(U) stretch which may indicate that the sequences are incomplete (Corman et al., 2015). Furthermore one rodent isolate demonstrated incomplete structural similarity with the HCV X-tail with the terminal stem loop conserved but not the preceding structures. These data indicate that the 3'UTR structure of these related viruses may be conserved, or at least that some structural features are maintained. The observed conservation throughout the majority of the genome would suggest that this is the case. However, more 3'UTR sequences from more isolates will need to be determined to confirm this.

1.6.2 Tissue tropism, pathology, and prevalence, of the newly discovered hepaciviruses

Like HCV and EqHV the newly discovered hepaciviruses exhibit liver specific tropism. The highest levels of genomic RNA were isolated from sera and liver tissue in all of the species investigated, with minimal to no viral RNA isolated from other tissues. Higher levels of viral RNA were also detected in liver lymph nodes. The presence of conserved miR122 seed sites in the 5'UTRs provides further evidence that these viruses exhibit liver tropism.

BHV was isolated from two species of African bats with 0.6% prevalence. The authors did note that they were unable to assess the true prevalence, host range, or geographical distribution of the BHVs identified in this study, but that it can be said that they are highly genetically diverse and are circulating worldwide. Despite this no obvious signs of pathological disease were observed in any infected bats even though high viral titres (up to 3×10^8) were detected (Quan et al., 2012).

Viral prevalence appeared to be higher in bovines, 3.2% of German cattle tested positive for BovHepV infection. This study also demonstrated viral persistence, with 5 cows testing positive over a one year period following initial detection of viral RNA; however some animals spontaneously cleared virus within this time frame indicating both chronic and acute infection of BovHepV. Whilst chronic infection was detected, no RNA positive animal exhibited significant differences in liver enzymes levels compared to uninfected samples. Liver degeneration or inflammation indicative of viral infection could not be detected in a post mortem analysis of the liver of an infected animal. However, viral loads of 2.92×10^5 were detected in the sampled liver

compared with 8.94×10^4 in the serum. Viral loads $< 4.15 \times 10^3$ were observed in other tissues and no virus could be isolated from milk, faeces, urine, or nasal discharge; in agreement with a liver specific, blood borne virus (Baechlein *et al.*, 2015).

In another investigation into hepaciviral infection in bovines (Corman *et al.*, 2015), this time in African cattle, a higher prevalence of 8.5% was detected in 106 animals, with similar viral loads of 1.6×10^5 . No tissue specific samples were available to confirm liver tropism however the data presented by Baechlein *et al.* and the presence of a conserved miR122 site in the 5'UTR would suggest this to be the case. Whilst Corman *et al.* did not present findings regarding the pathology of BovHepVs in African cattle they did present evidence of GORs within viral isolates; suggesting that they exhibit persistence.

Rodent hepaciviruses exhibited a lower level of prevalence than that of bovines (0.8-1.9%) which was more similar to the prevalence of hepaciviruses in bats. Mean RNA concentrations were highest in the liver and RNA could be detected by in-situ hybridisation; whereas no staining was observed in samples negative for hepacivirus RNA by PCR, or in spleen, kidney, heart and lung tissues. Histopathological signs of liver inflammation were also reported as well as an observation that antibodies did not co-occur with RNA in most animals, indicating that bank voles may be able to spontaneously clear virus (Drexler *et al.*, 2013). Another report of rodent hepacivirus in Norway rats observed that hepaciviral RNA was most consistently detected in liver tissue and the negative strand replicative intermediate was exclusively identified in liver tissue. Kapoor *et al.* also noted the existence of a conserved miR122 site in the 5'UTR of RHV indicating liver tropism (Kapoor *et al.*, 2013). Lauck *et al.* suggested a putative liver tropism for GHV based upon a similar observation (Lauck *et al.*, 2013). WLSV tropism is presumed to be specific to the liver as this virus was only isolated from fish liver tissue (Shi *et al.*, 2015).

Beyond the information regarding the viral sequences and tissue tropism obtained during the isolation and sequencing of this newly discovered set of viruses little is known about them. Excluding EqHV, no work has been undertaken to investigate the roles of the viral proteins and RNA structures that are predicted to function throughout the hepaciviral lifecycle.

1.6.3 Update to the taxonomy of the hepaciviruses

The naming of newly discovered hepaciviruses has so far used the animal host to denote new viral species i.e. bovine-, bat-, and rodent-, hepacivirus. However this has caused some confusion for EqHV regarding the nomenclature for this virus (CHV, NPHV, EqHV) and no true consensus exists in the field. Due to the number of new hepaciviral species being discovered, and the number of new isolates, an update to the taxonomy of the hepacivirus genus was proposed (Smith et al., 2016). This paper proposes hepaciviral species be denoted by a letter, rather than host species, and should encompass 14 species: hepacivirus A-N based upon the date of publication of the first complete sequence. Due to historical nomenclature for HCV and GBV-B these two viruses would be renamed hepacivirus C and hepacivirus B respectively, as opposed to A and B based on publication date. However CHV/NPHV/EqHV would be renamed hepacivirus A. Interestingly the analysis conducted in this proposal suggested that WLSV is distinct from the hepaciviruses, and also the closely related pegiviruses. Whilst it most closely grouped with the hepaciviruses the authors did not go as far as to provide a hepacivirus classification. They suggested that until further information about the biology and molecular biology of WSLV was obtained it remain as an unclassified virus in the *Flaviviridae*. This study confirmed that EqHV (hepacivirus A) is the most closely related hepacivirus, discovered to date, to HCV (Figure 1.3).

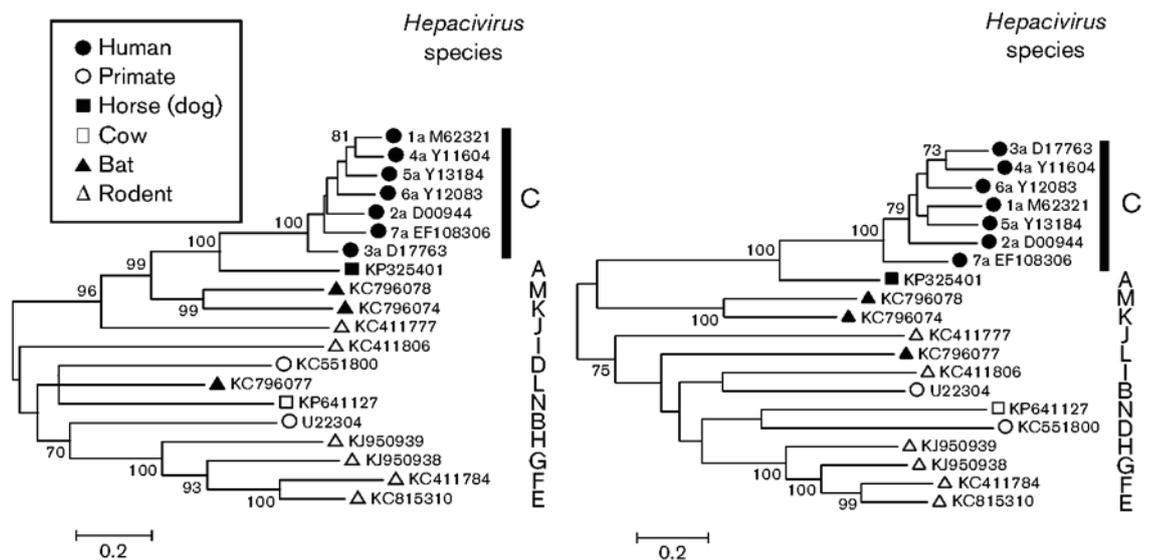


Figure 1.4: Phylogenetic analysis of the hepaciviruses

Phylogenetic trees using A) positions 1123–1566 and B) 2536–2959 of the virus polyprotein of all of the currently identified hepaciviruses indicating their species of origin and proposed new hepacivirus A-N classification.

1.6.4 The newly discovered hepaciviruses as a model system to study HCV

Given the close genetic relationship of EqHV and HCV, and apparent similarities between infections with regards to chronicity, it has been cited as a potential model virus with which to study HCV (Ramsay et al., 2015; Scheel et al., 2015). Horses are too large to be considered an ideal animal for laboratory testing, however one study has demonstrated immune protection from reinfection with EqHV in horses (Pfaender et al., 2017). Bovines are also too large to be considered a viable laboratory animal and given the ethical implications of testing on primates GHV is also not an optimal candidate. However a number of rodent hepaciviruses have been discovered which would appear to provide an ideal model with which to study the full hepaciviral lifecycle in an immunocompetent host.

In a recent study Billerbeck *et al.* used a rodent hepacivirus isolated from Norway rats (NrHV-1/Hepacivirus G) to infect laboratory mice, achieving acute and chronic infection in which the virus also appeared to adapt to its new host after passaging through it. Acute infection did not cause disease and chronically infected mice showed minimal to mild liver inflammation

(Billerbeck et al., 2017). This study also characterised the immune response to NrHV infection reporting a broadly directed virus-specific T cell response during acute resolving infection, and demonstrated that CD8+ T cell depletion rendered mice unable to clear the infection. NK cells were also activated upon NrHV infection, as were T_{regs} and this was suggested to contribute to the establishment of chronic infections. Overall this study has clearly demonstrated that NrHV infection in laboratory mice provides an immunocompetent model system with which to study hepaciviral infection and immunity, and may be able to contribute to the ongoing search for an HCV vaccine.

Whilst the NrHV mouse model represents a major breakthrough in hepaciviral research NrHV/Hepacivirus G is phylogenetically distant from HCV compared with the other newly described hepaciviruses, and another rodent hepacivirus (hepacivirus I/RHV) is more closely related to HCV (Figure 1.3). Therefore infection models using other hepaciviruses may yet provide more informative systems.

Furthermore, the advent of a replicative, immunocompetent, model should not diminish research into the other hepaciviruses, especially those which have not yet been demonstrated to replicate *in vitro* such as EqHV. Increasing understanding of the molecular biology of the most closely related virus to HCV still provides an invaluable tool with which to investigate this human pathogen. Differences identified between these two viruses may be instrumental in shaping future work on HCV, or in the mouse model. Furthermore, understanding the molecular biology of the genus as a whole may provide insights into this class of viruses that cannot be obtained from the study of one or two species in isolation. The scientific community must still investigate this novel group of viruses, and search for new ones, to increase understanding of the biology of hepaciviruses, HCV, and positive sense RNA viruses as a whole.

1.7 Project aims

EqHV is the most closely related virus to HCV, however it does not match the persistence or pathology exhibited by HCV. The molecular biology of this virus is poorly understood and has not been extensively studied. EqHV research is currently being held back by the lack of a cell culture system with which to study its replication, and as such is limited to functions which can be studied independently of replication. This study therefore set out to further investigate the molecular biology of this newly discovered model virus by determining the secondary structure of the EqHV 5'UTR IRES and assigning functional attributes to structural elements, and to establish an EqHV SGR capable of replicating in mammalian cell culture.

Positive strand RNA viruses are able to directly translate their genome upon release into the cytoplasm and this represents one of the first viral-host interactions. EqHV does not exhibit the same pathology as HCV, therefore identifying differences between these two closely related viruses may begin to provide insights into the mechanisms which underlie them. By investigating the structure of the EqHV 5'UTR using SHAPE, and how this relates to its function and interactions with host cell factors, we hoped to observe differences in translation initiation between EqHV and HCV, which may begin to explain differences in pathology.

A replicative model for HCV transformed the field and has been instrumental in the advent of new, highly successful, DAA therapies. EqHV research is currently limited by a lack of a replicative model. As such an EqHV SGR able to efficiently replicate in cell culture would represent a leap forward not only in the field of EqHV research but also in the wider hepacivirus field. Understanding how EqHV replicates could provide evidence for highly conserved hepaciviral mechanisms, or key differences, either of which would provide insights into the HCV lifecycle and replication strategies of the ever expanding hepacivirus genus. We therefore attempted to construct a replicative EqHV SGR using two different EqHV sequences, and assayed novel cell lines expressing known HCV factors.

Chapter 2 Materials and Methods

2.1 Materials:

2.1.1 Mammalian cell lines

Huh7 cells are a human hepatoma cell line known to support replication of HCV (Nakabayashi et al., 1982; Lohmann et al., 1999).

Huh7.5 cells are a Huh7 derived line created through the clearing of Huh7 cells stably harbouring an HCV SGR; supporting higher levels of HCV replication due to a dominant negative RIG-I mutation (Blight et al., 2002).

FHK cells are a foetal horse kidney cell line known to support the replication of a number of equine viruses and were kindly provided by Dr Hiroshi Sentsui (Nihon University, Japan) (Maeda et al., 2007).

293T cells are a human embryonic kidney cell line which have been demonstrated to support HCV replication when engineered to express miR122 (Fukuhara et al., 2012).

2.1.2 Bacterial strains

Escherichia coli (*E. coli*) DH5 α : Genotype; F- Φ 80lacZ Δ M15 Δ (lacZYA-argF) U169 recA1 endA1 hsdR17 (rk-, mk-) phoA supE44 λ -thi-1 gyrA96 relA1 were purchased from Life Technologies

Chemically competent bacteria were produced using the Z-competent kit (Zymo Research).

2.1.3 Virus sequences

Two EqHV viral sequences were used in this study

EF369_11J; GenBank accession no. JX948116.1.

NZPI; GenBank accession no. KP325401.

2.1.4 SGR constructs

pSGR NZPI was kindly provided for use by C. Rice (Rockefeller University).

pSGR H-E-H was constructed at the University of Leeds by Cheryl Walter and Hazel Stewart by adapting the pSGR JFH as described (Targett-Adams and McLauchlan, 2005) using the EqHV isolate EF369_11J; GenBank accession no. JX948116.1.

2.1.5 Bicistronic expression constructs

pGL3 RLuc/FLuc expression constructs were kindly provided for use by Kensuka Hirasawa (Memorial University, Newfoundland, Canada).

2.1.6 miR122 reporter constructs

pGL3 1225 and pGL3 MCS were kindly provided by Catherine Jopling (University of Nottingham).

2.1.7 N-Methylisatoic anhydride (NMIA)

NMIA was purchased from Sigma-Aldrich and resuspended in dimethyl sulfoxide (DMSO [Sigma]) to a concentration of 100 mM.

2.1.8 Purified eIF3 and 40S ribosomal subunit

Purified eIF3 and 40S ribosomal subunit was kindly provided for use by Dr. Nicolas Locker (University of Surrey).

2.1.9 Primers

Oligonucleotide primers were purchased from Integrated DNA technologies (IDT), resuspended in deionised water (dH₂O) to 100 μM, and stored at -20°C.

Fluorescent primers were also ordered from IDT as PAGE purified and 5'6FAM, or 5'HEX, fluorescently labelled.

2.1.10 Tissue culture microscope

Cultured cells were visualised with an OPTIKA XDS-2 microscope.

2.2 Nucleic acid manipulation

2.2.1 Bacterial transformation

Z-competent *E. coli* DH5 α were transformed by addition of 10-50 ng of DNA to 50 μ l of competent bacteria, on ice for 5 minutes, before spreading onto Luria Bertani (LB) agar plates with 100 μ g/ml tetracycline or ampicillin. Plates were incubated at 37°C, or 30°C for bacteria transformed with DNA coding for the EqHV NS5A.

2.2.2 Preparation of DNA from bacterial culture

Single bacterial colonies were picked and grown in lysogeny broth (10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract [LB]) in a rotary shaking incubator at 200 rpm in the presence of antibiotic. Constructs coding for the EqHV NS3-5B region were grown for 24 hours at 30°C, all others for 16 hours at 37°C. Bacterial cells were harvested by centrifugation at 4,000 \times *g* for 15 min at 4°C. Plasmid DNA was purified by alkaline lysis using GeneJET Plasmid Miniprep kit (Thermo Scientific) according to the manufacturer's protocol. For long term storage of plasmids, glycerol stocks of bacterial cultures were prepared from 1 ml cultures re-suspended in 30% glycerol, 70% LB (v/v) and stored at -80°C.

2.2.3 Nucleic acid quantification

DNA and RNA was quantified using the absorbance at 260 nm on a NanoDrop 1000.

2.2.4 Q5 site-directed mutagenesis

All mutagenesis, including large deletions, was carried out using the NEB Q5 SDM kit according to the manufacturer's instructions. Q5 SDM primers were designed in the NEBasechanger online tool. Primer pairs contained the desired mutation at the 5' (forward primer) and 3' (reverse primer) ends, or annealed to exclude the sequence to be removed. An initial denaturation step, 98°C for 30 seconds, was followed by 25 cycles of 98°C for 20 seconds, 20 seconds at the annealing temperature recommended by the NEBasechanger tool, 72°C for 30 seconds/kb of amplified DNA, and a final extension at 72°C for 2 minutes. PCR products were treated with the kinase/ligase/DNAse (KLD) mix for 5 minutes at room temperature and transformed into 50 μ l Z-competent *E. coli* DH5 α . All of the constructs described of this study are archived in the Harris group glycerol store (Garstang 8.61, available from Prof. M Harris).

2.2.5 Restriction digests

Restriction digests were carried out using enzymes purchased from NEB according to the manufacturer's instructions. A minimum of one enzyme unit:1 µg of DNA was used for a minimum of one hour.

2.2.6 Agarose gel electrophoresis

DNA gels consisted of 1% (w/v) agarose in TAE buffer (40 mM Tris-HCl, 0.11% (v/v) acetic acid, 1 mM ethylenediaminetetraacetic acid [EDTA]) with 1:10,000 SYBR Safe DNA Gel Stain (Invitrogen). To separate fragments samples were loaded in DNA loading buffer (1% (w/v) sucrose, 0.0125% (w/v) bromophenol blue in dH₂O) and run in TAE buffer at 8 V/cm and compared to Bioline Hyperladder I markers.

2.2.7 Gel extraction of DNA

DNA bands were visualised using blue light with an orange filter and bands of interest were excised. DNA was purified using the QIAquick Gel Extraction kit (QIAGEN).

2.2.8 Ligation

DNA fragments were ligated at a 1:5 molar ratio (vector:insert) by mixing and incubation on ice for 5 minutes and made up to a final volume of 20 µl with dH₂O, T4 DNA Ligase buffer, and 10,000 units of T4 DNA ligase (NEB). Ligations were incubated at room temperature for 1 hour and 5 µl was transformed into Z-competent *E. coli* DH5α.

2.2.9 Verification of ligation products

Single colonies were picked and DNA prepared as described in 2.2.2, sequenced by Beckman Coulter Genomics, and aligned to the WT construct using DNA Dynamo to confirm the construct sequence.

2.2.10 Linearisation of DNA

To linearise pSGR constructs 10 µg of DNA was digested overnight with either XbaI (pSGR EqHV) or BspI (pSGR NZPI), or AscI to linearise directly following the 5'UTR for use in pull-downs. Linearised DNA was purified by phenol:chloroform:isoamylalcohol (25 : 24 : 1 [pH8.0]) extraction by addition of an equal volume to that of the digestion. The mixture was vortexed for 1 minute and centrifuged at 17,000 x *g* for 2 minutes. The upper aqueous phase was extracted and mixed

with an equal volume chloroform, vortexed, centrifuged, and the upper aqueous phase again collected. DNA was precipitated in 2.5 sample volumes of 100% (v/v) ethanol with 0.1 sample volumes of 3M NaAc at -20°C for > 1 hour. DNA was pelleted by centrifugation at 20,000 x *g*, 4°C, for 30 minutes and washed once in 500 µl 70% ethanol. The DNA pellet was air dried and resuspended in nuclease free H₂O.

2.2.11 RNA manipulation and DEPC treatment

When handling RNA all surfaces were pre-treated with RNase AWAY (Molecular BioProducts). Phosphate buffered saline used in RNA manipulation was treated with 0.1% (v/v) diethyl pyrocarbonate (DEPC) and incubated with shaking at 37°C for 16 hours, excess DEPC was removed by autoclaving. Nuclease free water for use in RNA manipulation was purchased from ThermoFisher.

2.2.12 *In vitro* transcription of RNA

Linearised DNA (2 µg) was used as a template in the T7 RiboMAX™ Large Scale RNA Production System (Promega). Reactions were incubated at 37 °C for 60 min before degradation of template DNA using 2 units of DNase for 30 min at 37 °C.

2.2.13 Lithium chloride precipitation of RNA

IVT RNA was precipitated using 7.5 M Lithium Chloride, 50 mM EDTA, pH 8.0 (ThermoFisher) at a final [LiCl] of 2.5 M and incubation at -20°C for >1 hour. RNA was pelleted by centrifugation at 20,000 x *g* for 30 minutes, washed with 70% ethanol and resuspended in nuclease free H₂O.

2.2.14 RNA gel electrophoresis

IVT RNA was analysed by denaturing agarose gel electrophoresis (1% (w/v) agarose, 1X MOPS (40 mM 3-(N-morpholino)-propanesulfonic acid [pH 7.0], 10 mM sodium acetate, 1 mM EDTA), 4.7% (v/v) formaldehyde), supplemented with 1:10,000 SYBR Safe DNA Gel Stain (Invitrogen). RNA samples were diluted in RNA loading buffer (47.5% (v/v) formamide, 9 mM EDTA, 0.0125% (w/v) SDS, xylene cyanol and bromophenol blue), denatured at 65°C for 10 minutes and loaded onto gels. Gels were run for 1 hour at 70 V and visualised under blue light illumination, compared with a ssRNA ladder (NEB).

2.3 Tissue culture

2.3.1 Mammalian cell lines

All mammalian cells used in this study were cultured in Dulbecco's modified Eagle's medium (DMEM; Sigma), 10% (v/v) foetal bovine serum (FBS), 100 IU/ml penicillin, 100 µg/ml streptomycin and 1% (v/v) non-essential amino acids (Lonza) in a humidified incubator at 37°C with 5% CO₂. Cells were passaged at 70-80% confluency by washing with PBS and incubation with 0.5 mg/ml trypsin-EDTA (Sigma). Trypsin was inactivated by addition of excess complete DMEM and seeded into culture flasks, or plates, for further passage or experimentation.

2.3.2 Freezing cell stocks

Cultured cells were frozen for long term storage at -80°C in full media supplemented with 10% DMSO, 40% FBS (v/v).

2.3.3 Retroviral transduction

Pseudotyped virus containing the gag/pol HIV capsid and VSVg glycoproteins, with the gene of interest packaged via the psi packaging signal, were created by transfection of 4 µg of plasmids coding for the gag/pol, glycoprotein, and psi flanked gene of interest, into 4x10⁵ 293T cells in a 10 cm² dish using PEI (as described in 2.3.6). Supernatant was collected 3 days post transfection, clarified at 1200 x *g* for 5 minutes and filtered through a 0.45 µm filter, to collect pseudotyped particles.

Mammalian cells were transduced by addition of 6 ml of the collected supernatant, and 4 ml full DMEM, to 4x10⁵ cells in a 10 cm² dish. After 24 hours of incubation media was replaced with 3:7 ml supernatant:full DMEM and this was repeated every 24 hours up to 72 hours. The supernatant was removed and replaced with full media supplemented with 5 µg/ml at 72 hours. This was conducted for both miR122 and SEC14L2 in Huh7, FHK, and 293T cells.

2.3.4 Maintenance of mammalian cells expressing miR122, SEC14L2, or the PIV5 V protein

For cells transduced to express miR122 or SEC14L2 media was supplemented with 5 µg/ml puromycin every second passage to ensure maintenance of miR122 expression. Cells expressing the PIV5 V protein were passaged in media supplemented with 500 µg/ml neomycin.

2.3.5 RNA electroporation

For either translation or replication assays cells were harvested by centrifugation following trypsin treatment and washed twice with ice cold PBS. Cells were counted and a final suspension of 5×10^6 created in ice cold DEPC treated PBS. 400 μ l of cell suspension was mixed with 2 μ g RNA in a chilled electroporation cuvette (Geneflow) and cells were electroporated at 950 μ F, 260 V for 25 msec (BioRad Gene Pulser). Cells were immediately recovered in complete media, seeded into culture plates and incubated at 37°C, 5% CO₂ until lysed for assay.

2.3.6 Nucleic acid transfection

DNA and RNA were both transfected using polyethylenimine (PEI) into 4×10^5 cells. Nucleic acid (2 μ g) was diluted in 100 μ l optimem (Sigma) and mixed with 10 μ l 1 mg/ml PEI and incubated for 10 minutes at room temperature. During this time cells were washed twice with PBS. Following incubation 600 μ l complete media was added to the transfection mix and this was immediately added to the washed cells. Transfected cells were incubated for 2 hours at 37°C, 5% CO₂, washed twice with PBS and complete media was added.

2.3.7 Firefly luciferase assays

Plates seeded with cells from electroporations (96 well, 2×10^5 cells) or transfections (6 well, 4×10^5) were harvested for luciferase activity by washing in PBS and lysis in Passive Lysis Buffer (PLB; Promega). For replication assays cells were harvested at 4, 24, 48, and 72 hours post transfection (h.p.t), for translation assays using SGRs cells were harvested 6 h.p.t, and for translation assays using the bicistronic constructs cells were harvested 24 h.p.t. For 96 well plates cells were lysed in 30 μ l PLB and 6 well plates 100 μ l. Samples were stored at -20°C until analysis. For samples obtained from a 6 well plate 30 μ l of sample was transferred to a 96 well plate for assay with 50 μ l of Luciferase Assay Reagent (Promega). Light emission was monitored on a BMG plate reader.

2.3.8 Dual Luciferase assay

Cells were treated as described for 6 well plates in 2.3.7 and assayed using the dual-luciferase reporter assay system (Promega) using 50 μ l of each reagent with light emission monitored on a BMG plate reader.

2.3.9 Selection of stable cell lines

Following transfection of SGR RNA cells were placed under selection with: 5 µg/ml puromycin, 10 µg/ml blasticidin, or 500 µg/ml neomycin, depending upon the resistance gene located in the SGR (Table 5.3) until discreet colonies had formed, or all cells had died.

2.4 *In vitro* methods

2.4.1 Selective 2'hydroxyl acylation and analysis by primer extension (SHAPE)

IVT RNA (12 pmol) was resuspended in 20 µl 0.5x TE buffer, incubated at 95°C for 2 minutes and cooled on ice for 2 minutes. Following this 103 µl ddH₂O, 45 µl 3.3x folding buffer (333 mM HEPES, 20 mM MgCl₂, 330 mM NaCl), and 2 µl RNase inhibitor (RNase OUT), was added and incubated for 30 minutes at 37°C. After incubation the mixtures were evenly split into positive and negative reactions to which 8 µl of 100 mM NMIA (positive [to a final concentration of 10 mM]) or DMSO (negative) was added. Mixtures were incubated for 50 minutes at 37°C and precipitated in 4 µl 5 M NaCl, 2 µl 100 mM EDTA, 1 µl 20mg/ml glycogen, 18 µl ddH₂O, and 350 µl 100% ethanol at -80°C for 30 minutes. RNA was pelleted by centrifugation at 20,000 x *g* for 30 minutes at 4°C, aspirated, dried, and resuspended in 10 µl 0.5x TE buffer.

For both the positive and negative reactions 5 µl of this RNA was incubated with 1 µl 10 µM FAM labelled fluorescent primer, and 6 µl ddH₂O at 85°C for 1 minute, 60°C for 10 minutes, and 30°C for 10 minutes. A master mix of 4 µl superscript IV (SSIV) RT buffer, 1 µl 100 mM DTT, 0.5 µl 100 mM dNTPs, 0.5 µl RNase OUT, 1 µl ddH₂O, and 1 µl SSIV RT, was added to each reaction and incubated for 30 minutes at 55°C.

For the sequencing ladder reaction 6 pmol unfolded IVT RNA in 7.5 µl 0.5x TE buffer, 1 µl 10 mM HEX labelled primer, and 2 µl ddH₂O was incubated at 85°C for 1 minute, 60°C for 10 minutes, and 30°C for 10 minutes and a master mix of 4 µl SSIV RT buffer, 1 µl 100 mM DTT, 0.5 µl 100 mM dNTPs, 0.5 µl RNase OUT, 1 µl ddGTP, and 1 µl SSIV RT, was added before incubation for 30 minutes at 55°C.

All RT extensions were heated at 95°C for 3 minutes, and cooled on ice with 2 µl 2 M HCl for 2 minutes. cDNA was precipitated in 4 µl 3 M NaAc, 4 µl 100 mM EDTA, 1 µl 20 mg/ml glycogen, and 60 µl 100% ethanol for 30 minutes at -80°C, pelleted by centrifugation, aspirated and

resuspended in 40 μ l deionised formamide (Hi-Fi formamide). Samples were pooled with 20 μ l of ladder and stored at -80°C .

2.4.2 Gel Based SHAPE

SHAPE reactions were carried out as described in 2.4.1. However, primers were first labelled with $\gamma\text{ATP-}^{33}\text{P}$ through the addition of 5 μ l H_2O , 2 μ l 10x Poly nucleotide kinase (PNK) buffer, 2 μ l PNK, and 11 μ l $\gamma\text{ATP-}^{33}\text{P}$ to 1 μ l (60 pmol) of unlabelled primer. Reactions were incubated for 30 minutes at 37°C , and 20 minutes at 65°C . Labelled primers were precipitated in 4 μ l 5 M NaCl, 2 μ l 100 mM EDTA, 18 μ l dd H_2O , and 350 μ l 100% ethanol at -80°C for 30 minutes, pelleted by centrifugation at 20,000 $\times g$ for 30 minutes at 4°C , aspirated, dried, and resuspended in 50 μ l 1 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES). For the SHAPE reaction 3 μ l of primer was used. Primer extensions were separated by running on a polyacrylamide sequencing gel (5 ml 10x TBE, 20g Ureare, 9 ml 40% acrylamide, 35 ml H_2O , 400 μ l APS, 10 μ l TEMED) for 3 hours at 40 watts. Gels were dried and exposed to a phosphor screen for one week and imaged using a FujiFilm FLA-500 imager

2.4.3 SHAPE footprinting

SHAPE footprinting was carried out as described in 2.4.1 however following folding of the IVT RNA, and prior to NMIA/DMSO treatment, 300 nM eIF3, or 40S ribosomal subunit, was added and incubated for 20 minutes at 37°C .

2.4.4 SHAPE data analysis

SHAPE fragment analysis was conducted by capillary electrophoresis by DNA sequencing and services, Dundee University. SHAPE data was analysed in the program QuSHAPE (Karabiber et al., 2013) and RNA structure prediction was carried out using the RNAstructure software (Reuter and Mathews, 2010) using the SHAPE reactivity profile as a pseudo free energy constraint. RNA secondary structure was modelled in VarNA.

2.4.5 Pull downs

To pull-down EqHV 5'UTR binding partners from Huh7 cells the Pierce™ Magnetic RNA-Protein Pull-Down Kit was used. IVT EqHV 5'UTR RNA was 3' biotinylated using the Pierce RNA 3' End Desthiobiotinylation Kit according to the manufacturer's instructions. Huh7 cells were lysed in Thermo Scientific Pierce IP Lysis Buffer and total protein concentration determined using a

NanoDrop. Biotinylated EqHV 5'UTR RNA (50 pmol) was bound to Nucleic-Acid Compatible Streptavidin Magnetic Beads and used as bait to pull-down from the Huh7 cell lysate (>2mg/ml). In place of a final elution step beads were resuspended in 50 µl wash buffer and stored at -80°C to be sent for tandem mass tag (TMT) mass spectrometry analysis.

2.4.6 TMT mass spectrometry

TMT mass spectrometry was carried out Kate Heesom (University of Bristol Proteomics Facility) and analysed against a bead only negative control sample.

Chapter 3

A structural and functional analysis of the EqHV 5'UTR

3.1 Introduction

The EqHV 5'UTR is a 385 nucleotide sequence located at the 5' end of the EqHV genome. It has been predicted to comprise of three main structural domains, SLI, SLII, and SLIII (comprising domains SLIIIa,b,c,d,e,f), and a pseudoknot (Figure 3.1A and Figure 3.1C). Previous work on the EqHV 5'UTR demonstrated that, like its HCV counterpart, it functioned as a viral IRES. Initiating translation in a cap-independent manner in mammalian cells in the context of both a bicistronic reporter construct, and an HCV SGR in which the HCV 5'UTR was replaced with that of EqHV (Stewart et al., 2013). This work also demonstrated that a Δ SLI deletion had no significant effect on translation efficiency from the EqHV IRES. In addition to this, work conducted by Scheel *et al.* investigated the effect of the presence of the EqHV 3'UTR, and miR122, on IRES activity. This work demonstrated that IRES activity was significantly enhanced by either the presence of the EqHV 3'UTR, or miR122, in a monocistronic system. Conversely, sequestration of miR122 using a locked nucleic acid (LNA) significantly decreased IRES efficiency (Scheel et al., 2015).

The EqHV 5'UTR shares 66% sequence identity and a predicted high degree of structural similarity with the HCV 5'UTR (Kapoor et al., 2011). Translation of the viral genome is an essential part of the viral lifecycle and represents one of the first virus-host interactions during the establishment of infection. Translation initiation from the HCV 5'UTR IRES is a well-studied, and well understood, process. How the structural elements of this type IV IRES contribute to its function have been well established (reviewed in Fraser and Doudna, 2007). Domains II-IV constitute the IRES with domain I involved in replication, playing no role in IRES function. SLIIIb has been shown to interact with the ribosome binding face of eIF3, SLIIIc with the 40S ribosomal subunit, and the pseudoknot positions the AUG start in the ribosome binding cleft (Berry et al., 2010).

Previous studies were able to demonstrate that the EqHV 5'UTR possessed IRES activity, and identified viral and host cell factors which influenced its activity. However, neither conducted an in depth study regarding how the individual domains, and conserved sequences, of the 5'UTR contribute to IRES activity. This chapter will therefore describe the design, construction, and analysis, of a series of EqHV 5'UTR structural mutants, in comparison to the WT.

RNA secondary structure prediction is a fast and widely used method to investigate viral RNA structure where no experimental data exists, in order to analyse conserved structures and guide

mutational analysis. However, this does not come without its limitations (reviewed in Schroeder, 2009). One such limitation, pertinent to this study in particular, is the inability of many programs to reliably predict pseudoknots. Whilst some programs are designed specifically for this function each has its own limitations which would rendered them unsuitable for the purposes of this study. Therefore, the well-known and widely used program Mfold, a free energy based prediction software, was selected (Zuker, 2003).

The first section of this chapter (3.2.1) describes the predicted structure of the EqHV 5'UTR and how this correlates with that of HCV. The predicted conformation of a number of structural deletions introduced into the EqHV 5'UTR are also described.

As the structure of the 5'UTR, which guided the mutational analysis, was predicted it was important to experimentally determine the structure of the EqHV 5'UTR, and the structural mutants, to provide a more detailed insight into EqHV translation initiation to begin to build a model of EqHV IRES-dependent translation initiation.

There are a number of ways to analyse RNA secondary structure *in vitro*. One such approach is the use of RNAses with different specificities (i.e. single and double stranded). This method utilises parallel analysis of cleavage patterns, resolved on denaturing polyacrylamide gels, to assess the single, or double stranded conformation of RNA (Nilsen, 2013). An alternative approach to RNA secondary structure determination is the use of chemical agents and enzymes and functions by acting on bases sensitive to chemical and enzymatic attack; however this labour intensive method is most powerful in comparison with a model of secondary RNA structure generated from an alternative experimental approach (Ziehler and Engelke, 2001). Whilst both of these approaches provide a wealth of structural information about the RNA structure under investigation, they both have significant limitations. Therefore SHAPE analysis (described in 1.3.4) was chosen for use in this study.

SHAPE analysis utilises chemical probing reagents which selectively acylate the ribose in the backbone of flexible (unpaired) nucleotides within a structure. These chemical adducts prevent reverse transcription (RT) progression at a reactive base resulting in RT extension terminations. By using a labelled primer to conduct the RT step cDNA fragments can be analysed, in conjunction with a sequencing ladder, to determine which bases exhibit high levels of terminations, and are thus likely unpaired. This experimental methodology therefore has a significant advantage over other RNA structure probing methods as every base is simultaneously

reacted with, and therefore its structure interrogated, in one reaction; theoretically allowing the capture of information regarding the base pairing state of every single nucleotide with one experiment. The advantages of SHAPE are further extended by the advent of “high throughput” SHAPE (Wilkinson et al., 2008). This method utilises fluorescent labels to allow the cDNA fragments to be analysed by capillary electrophoresis, excluding the need for laborious and technically challenging sequencing gels. Furthermore, analysis by capillary electrophoresis allows for the consecutive analysis of up to 96 samples. The use of data obtained from SHAPE analysis in the guide of RNA secondary structure prediction has been shown to be extremely accurate. The use of SHAPE data as a “soft” pseudo-free energy constraint increased the prediction accuracy of the 16S *E. coli* RNA structure from 72% for traditional chemical probing methods to ~95%. Inclusion of SHAPE reactivity data routinely results in base pair prediction accuracies >90% (Rice et al., 2014; Deigan et al., 2009; Hajdin et al., 2013); put together these advances in chemical probing reagents, and detection, makes SHAPE an extremely powerful and relatively high throughput method to investigate RNA secondary structure (the current state of SHAPE is reviewed in Kenyon et al., 2014).

Section 3.2.2 of this chapter describes the experimental process of the SHAPE analysis used in this study. Subsequent sections describe the structural analysis of the WT (3.2.3) Δ SLI (3.2.4) Δ SLI+II (3.2.5) Δ SLIII (3.2.6) Δ SLIIIb (3.2.7) and Δ SLIIId (3.2.8) EqHV 5'UTR, as a complementary system to further our understanding of the role which structure plays in the function of the EqHV 5'UTR IRES.

Bicistronic reporter constructs have long been used to investigate the activity of nucleic acid which has been postulated to function as an IRES. The principle of such constructs is two cistrons, each under independent translational control. Translation of the first is driven by the m⁷G cap, with the second potentially driven by the putative IRES under investigation (reviewed in Terenin et al., 2017). By assessing the relative translation of each cistron it is possible to draw conclusions regarding the IRES activity of the sequence being investigated.

Section 3.2.9 of this chapter describes the analysis of WT EqHV 5'UTR IRES function, and structural mutants, in the context of a classical RLuc/FLuc bicistronic construct, and the effect that miR122 has on the activity of the EqHV IRES in this context.

Whilst bicistronic reporter constructs have been widely used for the analysis of viral IRES activity, and have many advantageous properties in this regard, there are also a large number of

disadvantages regarding their use (reviewed in Terenin et al., 2017). It is known that aberrant translation of the second cistron can be driven by cryptic promoter sequences and splicing events resulting in false positive IRES activity, especially in the case of cryptic promoters in HCV. It is difficult to establish comprehensive controls for assays utilising a bicistronic system. Furthermore miR122 has been demonstrated to protect HCV viral RNA from the endonuclease Xrn1 (Thibault et al., 2015), a function which may be ablated by internal positioning of the IRES. Due to the known pitfalls of bicistronic assays in general, and the potential for HCV specific problems to arise in EqHV due to their similarity, further analysis of the EqHV IRES using an alternative system was deemed necessary. Both monocistronic, and SGR, systems have been used to assay EqHV 5'UTR IRES activity (Stewart et al., 2013; Scheel et al., 2015) and were considered as potential candidates for this study. Both constructs position the 5'UTR at the proximal 5' end, as in the viral genome, facilitating a more relevant interaction with miR122. Both systems can also include the 3'UTR, which has been shown to upregulate translation. However, the SGR was chosen as it is the more physiologically relevant of the two.

Section 3.2.10 of this chapter describes the analysis of the aforementioned structural mutants in the context of an EqHV SGR, investigating the differences observed between the activity of the IRES in the bicistronic and SGR constructs. The effect of miR122 on translational activity, and the EqHV sequence that is mediating such effects, is also described.

As there are fundamental differences in the pathology of EqHV and HCV and it has not, as yet, been possible to establish replication of EqHV *in vitro* it was postulated that the inability of EqHV to cause significant disease may be linked to its inability to replicate in mammalian cell culture. Therefore it may be that the predicted structural differences observed between these two viral IRES are affecting their functions and, as translation is one of the earliest interactions between the virus and host cell on the path to establishing replication, the ability of EqHV to establish infection. This could therefore limit its pathogenic potential and ability to replicate in mammalian cell culture. It was therefore essential to further investigate the IRES function of the EqHV 5'UTR as this could shed light on the fundamental differences between these two viruses and provide information on why HCV so efficiently establishes a disease with such a significant pathology.

3.2 Results

3.2.1 Predicting the structure of the EqHV 5'UTR and structural mutants

The secondary structure of the EqHV 5'UTR has previously been described and has been predicted to contain three main structural domains: SLI, SLII, SLII (comprising SLIIa,b,c,d,e,f), and a pseudoknot (Kapoor et al., 2011; Burbelo et al., 2012). The HCV 5'UTR has been well characterised and is known to contain four main structural domains: SLI, SLII, SLII (comprising SLIIa,b,c,d,e,f), SLIV and a pseudoknot. An independent prediction of the EqHV 5'UTR was conducted for this study using the NZPI EqHV 5'UTR sequence (GenBank accession no. KP325401) in the RNA prediction software Mfold (see appendices for Mfold predicted structures). A schematic representation in comparison to the known HCV 5'UTR structure can be seen in Figure 3.1A and Figure 3.1B.

This agreed well with previously published predictions of the secondary structure (Kapoor et al., 2011; Burbelo et al., 2012) indicating the presence of all three main structural domains. One difference observed here was the lack of an intermediate SL that has previously been predicted to reside between SLI and SLII, 5' to the miR122 seed site (herein termed SLIA). Whilst this SL has previously been predicted for EqHV the Mfold predictions conducted here did not indicate the presence of SLIA. Furthermore SLIA demonstrates high sequence comparison with the, unpaired, miR122 seed site one of HCV and was therefore omitted from the structural prediction to maintain a more "HCV like" 5'UTR. In comparison with the HCV 5'UTR the secondary structure is largely similar throughout SLII and SLIII; however, 5' relative to SLII there are some marked differences. Most notably is the large and extended SLI of EqHV which is 57 nucleotides larger than HCV SLI, comparable in size to SLII, and the lack of SLIV. With regards to sequence specific features there are some interesting observations to be made. The HCV 5'UTR contains two miR122 seed sites; however, just one seed site, corresponding to seed site two, is conserved in the EqHV 5'UTR. This conservation indicates that miR122 plays a role in the EqHV lifecycle. However, this incomplete conservation may suggest that it either plays an alternative, or less important, role than in HCV.

Furthermore five out of the six nucleotides which form the apical loop of SLIIId are conserved between these two viruses (UUGGGC/U) with EqHV containing a C at position 6 in the loop and HCV a U. Whilst the apical loop is 100% conserved amongst HCV isolates, due to previous studies

demonstrating that the GGG motif interacts with the 40S ribosomal subunit during HCV translation initiation, and that only the GGG motif is conserved amongst other *Hepacivirus* and *Pestivirus* species, this study focussed only on the conserved GGG motif. Predicting the existence of a pseudoknot by conventional RNA secondary structure prediction software is difficult and often unreliable. As such, the existence of such a structure was not initially predicted by Mfold. However, by closely examining the sequence of this region in the context of the predicted secondary structure of SLI, SLII, and SLIII it became clear that the existence of a pseudoknot was possible. From the beginning of the stem of SLIIIe through to the end of the predicted pseudoknot base pairing region there is 84% sequence conservation between HCV and EqHV. However, the apical loop of SLIIIe is 100% conserved with HCV and 7 of the 8 nucleotides in the apical loop of SLIIIe are conserved. The predicted stem of the EqHV pseudoknot only exhibits 55% sequence identity with HCV. However, closer inspection revealed that the equivalent regions in these two viruses both have 9 nucleotide long sequences which are able to form canonical base pairs. Base pairing in this region would not preclude the formation of either SLIIIe or SLIIIe and would correctly position the conserved bases involved in the pseudoknot to form said interactions. Therefore a pseudoknot was modelled onto the EqHV 5'UTR using the HCV pseudoknot as a guide; this can be seen in Figure 3.2C.

Whilst it has been demonstrated to possess IRES function, to date, no functional attributes have been assigned to any individual domains of the EqHV 5'UTR. An in depth analysis of the structural features of the EqHV 5'UTR was therefore undertaken to assess how these contributed to its function as an IRES. Based upon the secondary structure that had been predicted (Figure 3.1) a panel of structural deletion mutants were designed which encompassed deletions of SLI, SLI+II, SLIII (excluding SLIIIe and f), SLIIIb, or SLIIId. Mfold was again utilised, following the design of said mutants, to predicted whether the deletion of any of these segments would impact the overall secondary structure of the 5'UTR and thus have effects on translation that were not specific to the structure being investigated (all Mfold predicted structures can be seen in Appendices Figures 8.1-8.8).

Deletions were introduced using Q5 SDM (NEB) according to the manufacturer's instructions, with mutagenesis primers designed in the online tool NEBaseChanger. Figure 3.2 is a schematic representation of all mutants which have been modelled on the predicted Mfold structure, and which also include the predicted pseudoknot. SLI is the main structural difference between the EqHV and HCV 5'UTRs and a deletion removing this SL has previously been made in the context of a bicistronic reporter system, and a chimeric EqHV/HCV SGR which contained the EqHV 5'UTR, and HCV 3'UTR and polyprotein. This deletion caused no significant effect on translation (Stewart et al., 2013). In this previous study the secondary structure of the EqHV 5'UTR was predicted to contain SLIA, as such deletion of SLI would result in a 5'UTR structurally similar to that of HCV, where SLIA is presumed to be homologous to HCV SLI. As this study does not predict the formation of SLIA, and utilises a consensus 5'UTR, this deletion (Δ SLI) was included to further investigate what effect SLI has on translation driven by the EqHV IRES.

SLIII has been shown, in HCV, to be the most important domain with regards to IRES-dependent translation. Therefore SLI+II were removed in combination, leaving only SLIII (Δ SLI+II), and SLIII was deleted (Δ SLIII), with SLIIIe, f, and the pseudoknot retained, to determine if this was also true for EqHV IRES function. SLIIIb, in HCV, has been demonstrated to interact with eIF3 during translation initiation, interacting with the ribosome binding face in order to sequester eIF3. This interaction prevents the formation of 43S complexes, increasing the availability of free 40S ribosomal subunits required for translation initiation from the HCV IRES (Hashem et al., 2013). This sub domain was therefore deleted (Δ SLIIIb) to investigate whether SLIIIb may have a similar function during translation from the EqHV IRES. Finally SLIIId, specifically the GGG motif located in the apical loop, has previously been demonstrated to interact with the 40S ribosomal subunit

during HCV translation initiation, positioning the ribosome at the AUG start codon. SLIId was therefore deleted (Δ SLIId) in its entirety to investigate whether translation from the EqHV IRES was able to proceed in its absence. This panel of deletions allows for a more detailed understanding of translation initiation from the EqHV IRES to be established to begin to understand if this proceeds in the same fashion as in HCV.

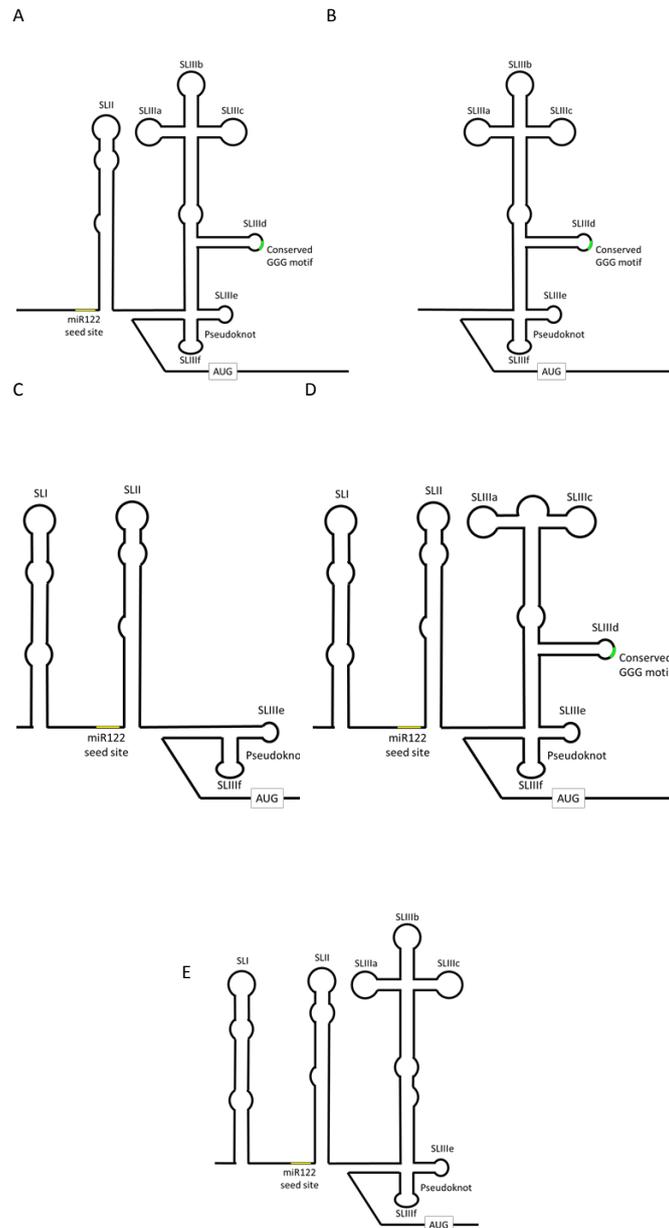


Figure 3.2: A schematic representation of the structural deletion mutations

Schematic representations of the predicted structures of A) Δ SLI, B) Δ SLI+II, C) Δ SLIII, D) Δ SLIIIb, and E) Δ SLIIId. The secondary structure of all 5'UTR deletions was predicted in Mfold, and a schematic representation, including the modelled pseudoknot, was drawn to represent each (as in Figure 3.1). The miR122 seed site is highlighted in yellow and the conserved GGG motif in green, the AUG start codon is denoted by text.

3.2.2 The execution of SHAPE experiments

SHAPE fragment analysis was conducted using capillary electrophoresis by DNA sequencing and services, Dundee University (a schematic can be seen in Figure 3.3). Data analysis was carried out using QuSHAPE (Karabiber et al., 2013) and structure predictions were carried out in RNAstructure (Reuter and Mathews, 2010) reading the SHAPE reactivities as a pseudo free energy constraint to produce an RNA secondary structure based upon the experimental data. Using the experimental set up described here we were able to produce high quantity, high quality, RNA which could be used to reproducibly obtain high quality SHAPE reactivity profiles covering the majority of the EqHV 5'UTR.

Initially gel-based SHAPE was utilised in order to confirm that elongation was being achieved using the designed primers (Appendices Figure 8.9) which were designed to bind in the luciferase gene ~100 bases downstream of the AUG start. Primers were radioactively labelled with γ ATP-³³P and used in place of fluorescently labelled primers as described in 2.4.1, using 5 mM NMIA rather than 10 mM. Extensions were analysed by separation on a SHAPE sequencing gel. Gels were exposed to a phosphor screen for one week and imaged using a FujiFilm FLA-500 imager (Figure 3.4). Whilst this process was not optimised to achieve high quality reads that would be suitable for SHAPE analysis, it was sufficient to demonstrate that the designed primers could successfully be used for elongation. As such these primers were synthesised with FAM and HEX fluorescent labels for use in high-throughput SHAPE using capillary electrophoresis.

Initial high-throughput SHAPE analysis was conducted using 5 mM NMIA, as was used in the gel-based SHAPE. However, this concentration of NMIA returned a sub-optimal trace in which reactivity was low toward the 5' end of the UTR and did not achieve sufficient signal to determine the individual SHAPE reactivities of these bases (Figure 3.5A). Therefore higher concentrations of NMIA were used to optimise reactivity and provide consistent reactivity through the entire UTR. Two concentrations (10 and 15 mM) of NMIA (Figures 3.5B and 3.5C) were trialled. This optimisation revealed that 15 mM NMIA resulted in too high a level of reactivity, resulting in a trace which frequently exceeded the detection limit and was therefore not considered acceptable. However, the use of 10 mM NMIA returned a trace which had consistently detectable reactivity, did not frequently exceed the detection limit, and maintained a reactivity profile through to the 5' end of the UTR. This concentration of NMIA was therefore taken forward to analyse the secondary structure of the EqHV 5'UTR, and mutants.

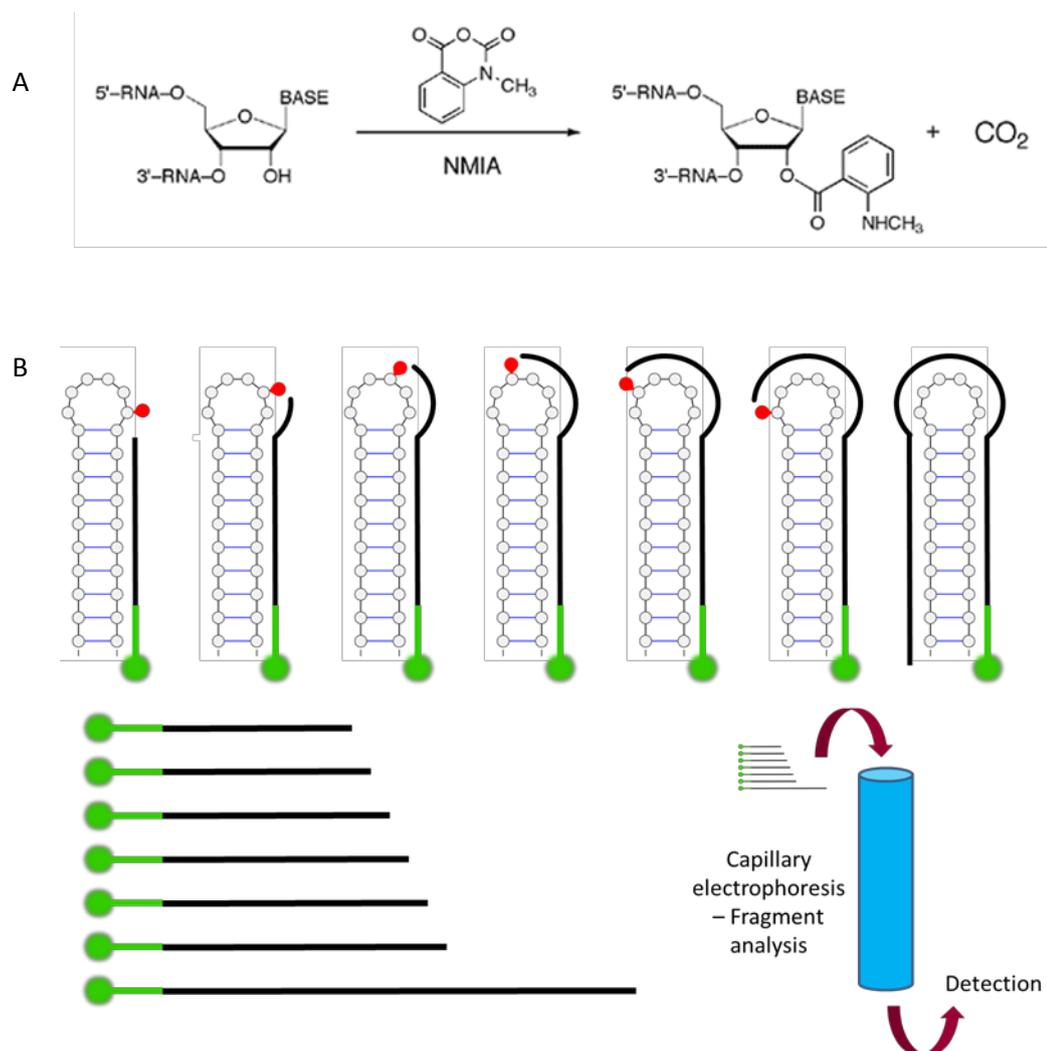


Figure 3.3: Schematic representation of SHAPE

A schematic representation of A) NMIA reacting with a flexible ribose moiety in RNA (Sheppard, 2005) and B) SHAPE experimental flow through. NMIA is represented in red, with fluorescently labelled primers shown in green. RT extensions are shown as black lines.

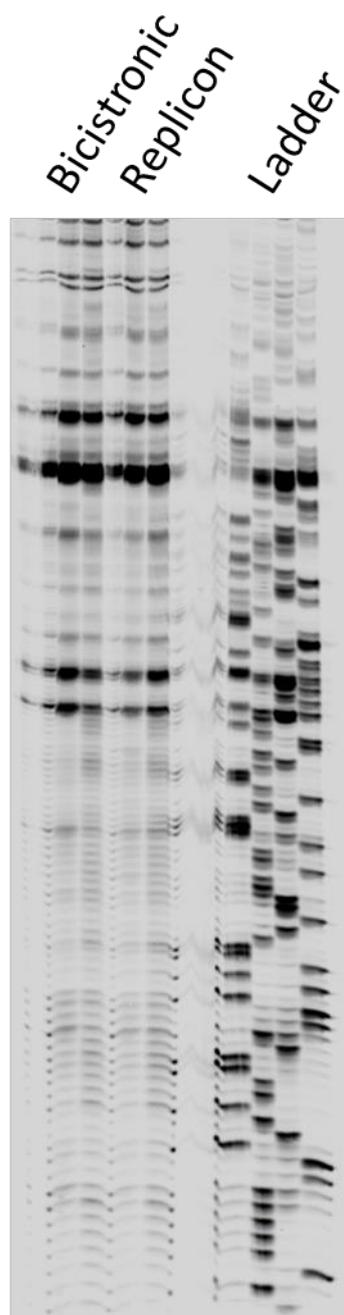


Figure 3.4: Determination of SHAPE primer extension using gel based SHAPE

SHAPE experiments using 5 mM NMIA were carried out on the full length 5'UTR in either the bicistronic or replicon constructs and extensions were separated in a sequencing gel in comparison to a ladder. Each band represents a primer extension termination event with darker bands representing sites of frequent termination.

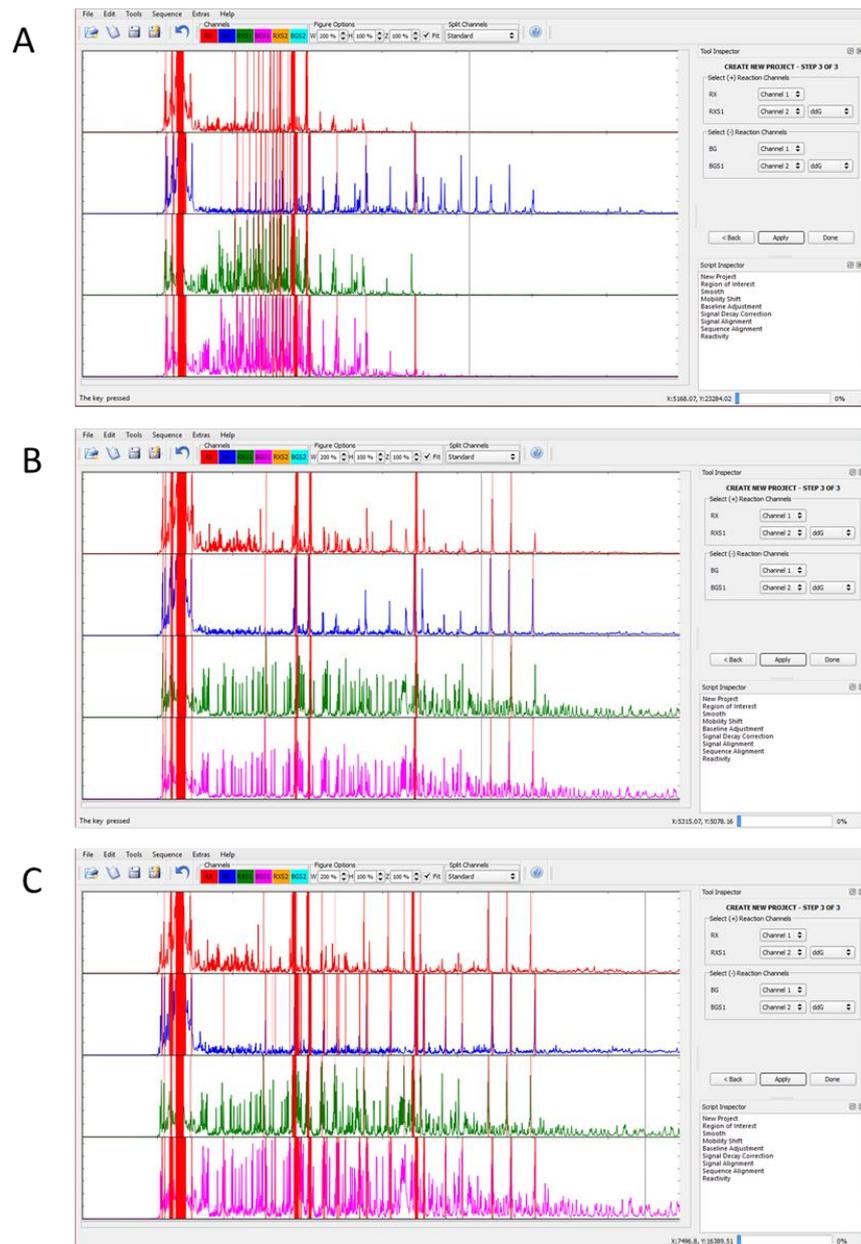


Figure 3.5: Optimisation of NMIA concentrations for high-throughput SHAPE

Capillary electrophoresis sequencing traces for the full length EqHV 5'UTR in the context of the replicon construct using A) 5 mM, B) 10 mM, and C) 15 mM NMIA. The top, red, trace is the NMIA (positive) reaction trace. The blue trace is the DMSO (negative) reaction trace. The green and pink traces represent the sequencing ladders for the positive and negative reactions respectively. Each peak represents an individual nucleotide and vertical red lines indicate a nucleotide for which the reactivity has exceeded the detection threshold.

3.2.3 SHAPE analysis of the WT 5'UTR

The WT 5'UTR of EqHV has been conclusively demonstrated to function as an IRES and is able to efficiently drive expression in the context of bicistronic reporter constructs, monocistronic expression constructs, and an SGR (Stewart et al., 2013; Scheel et al., 2015). However the RNA secondary structure has, to date and to our knowledge, only ever been predicted and never experimentally confirmed. Therefore determining the secondary structure of the EqHV 5'UTR served two purposes: Firstly to provide the first experimental description of the secondary structure of this viral IRES, and secondly to demonstrate that the UTR used in the experiments described in Chapter 3 matched that of the predicted structure, providing a reference structure with which to compare the 5'UTR mutants.

SHAPE experiments were carried out on the WT 5'UTR in both the SGR and bicistronic construct. SHAPE analysis was carried out on the UTR in both these constructs to ensure that the deletion mutations caused the same effect on the overall 5'UTR structure in both constructs. The values obtained from SHAPE were used in the prediction of RNA secondary structure as a pseudo free energy constraint in the prediction software RNAstructure (Reuter and Mathews, 2010).

Due to inaccuracies in structure prediction introduced by the presence of pseudoknots, and the predicted presence of a pseudoknot in the EqHV 5'UTR and the potential for this to disrupt structure prediction in its immediate vicinity, SLIIIe, f, and the pseudoknot were manually modelled, with the SHAPE reactivities mapped on to the structure. Whilst this approach discounted SHAPE reactivities in this region with regards to structure prediction, it was considered an appropriate method to tackle the issues caused by such structures, providing an accurate representation of RNA secondary structure in this region; especially when the sequence and structural similarities to HCV in this region are taken into account.

Figure 3.6 is an experimental model of the EqHV 5'UTR RNA secondary structure, in the context of an SGR. This was created using the experimental and prediction methodologies outlined in this chapter. Figure 3.7A is the same model with the SHAPE reactivity profile overlaid onto the structure (due to experimental limitations reactivity values for nucleotides 371-387 are not available). This, to our knowledge, represents the first experimentally confirmed model of the EqHV 5'UTR. The experimental structure agrees well with the predicted structure presented in Figure 3.1, with the exception of the presence of SLIA. Figure 4.3B shows an enlarged image of SLIA, and the miR122 seed site (₉₆CACUCC₁₀₁) which shows that the loop of SLIA was NMIA

reactive and would be predicted to be unpaired; confirming the existence of this intermediary SL. The miR122 seed site was NMIA reactive at its 5' end with a reduction in NMIA activity seen sequentially toward the 3' end.

Apart from SLIA, the structure of the EqHV 5'UTR agrees well with the overall predicted structure. Figures 3.7C and D show an enlarged view of SLIIIb, and SLIIId respectively. Both structural domains are here shown to form in the EqHV 5'UTR, with both structures matching extremely well with SHAPE reactivities. Both apical loops, ²⁴⁹ACUUU₂₅₃ and ³⁰⁷UUGGGC₃₁₂, showed extremely high SHAPE reactivities which match well with the structures. Finally SLIIIe, f, and the pseudoknot, are enlarged in Figure 3.7E. SLIIIe demonstrated intermediate SHAPE reactivities across its apical loop, agreeing with the structure, and ³⁵⁷ACCUGCCGG₃₆₅ showed low to intermediate reactivities, again agreeing with the structure; indicating they are base paired. There is, however, a lack of SHAPE reactivity in the apical loop of SLIIIf (excluding ³⁴⁹G and ³⁵⁰C) which appears to contradict the existence of this loop. However, low SHAPE reactivities in this region have previously been published for HCV (Angulo et al., 2016). ³⁴⁹GCGAG₃₅₃ are also predicted to form base pairs with ³⁶⁷CUCGU₃₇₁ as part of the pseudoknot and would therefore exhibit low NMIA reactivity in this confirmation. Furthermore, ³³⁶U-A₃₇₂ is 84% identical to the corresponding HCV sequence, with complete conservation in the apical loops of SLIIIe and SLIIIf (excluding ³⁵⁴G), and all of the bases which are predicted to form the pseudoknot. Due to this high level of conservation with the known structure of HCV, and the observation of similar SHAPE reactivities observed between the two viruses, this structure likely represents the biological secondary structure of this region of the EqHV 5'UTR.

The structure represented in Figure 3.6 and 3.7 are expected to represent an accurate depiction of the structure of the EqHV 5'UTR which has been modelled based on a combination of predicted secondary structure, experimental data, and comparison with conserved regions from HCV. This study presents, to our knowledge, the first experimental structure of the EqHV 5'UTR. Due to the accuracy of this model, and to allow for ease of comparison, the reactivity profiles of the WT bicistronic 5'UTR, and all mutants, were fitted onto the structure presented in Figure 4.2, or the respective predicted confirmation of each deletion mutant, based upon this structure and the Mfold analysis. The EqHV 5'UTR in the bicistronic reporter construct was also subject to SHAPE investigation using the same method outlined above. The results of these experiments can be seen in Figure 3.8. This analysis showed good agreement with the 5'UTR SHAPE reactivities from the SGR. Similarly, the apical loop of SLIA was highly reactive, and the miR122

seed site showed decreasing reactivity in the 5' to 3' direction, but was overall less reactive to NMIA than the SGR seed site. The apical loops of SLIIIb and SLIIId were both highly reactive, with some increase in reactivity in the base paired nucleotides, however this did not occur on a large scale and most likely represents experimental variation. Reactivities in SLIIIe, f, and the pseudoknot also agreed well with the data obtained from the SGR. It can be concluded that the 5'UTR secondary RNA structure does not differ significantly between the bicistronic construct and the SGR (a comparison of the SHAPE reactivity profiles between the SGR and bicistronic construct can be seen in Appendices Figure 8.10).

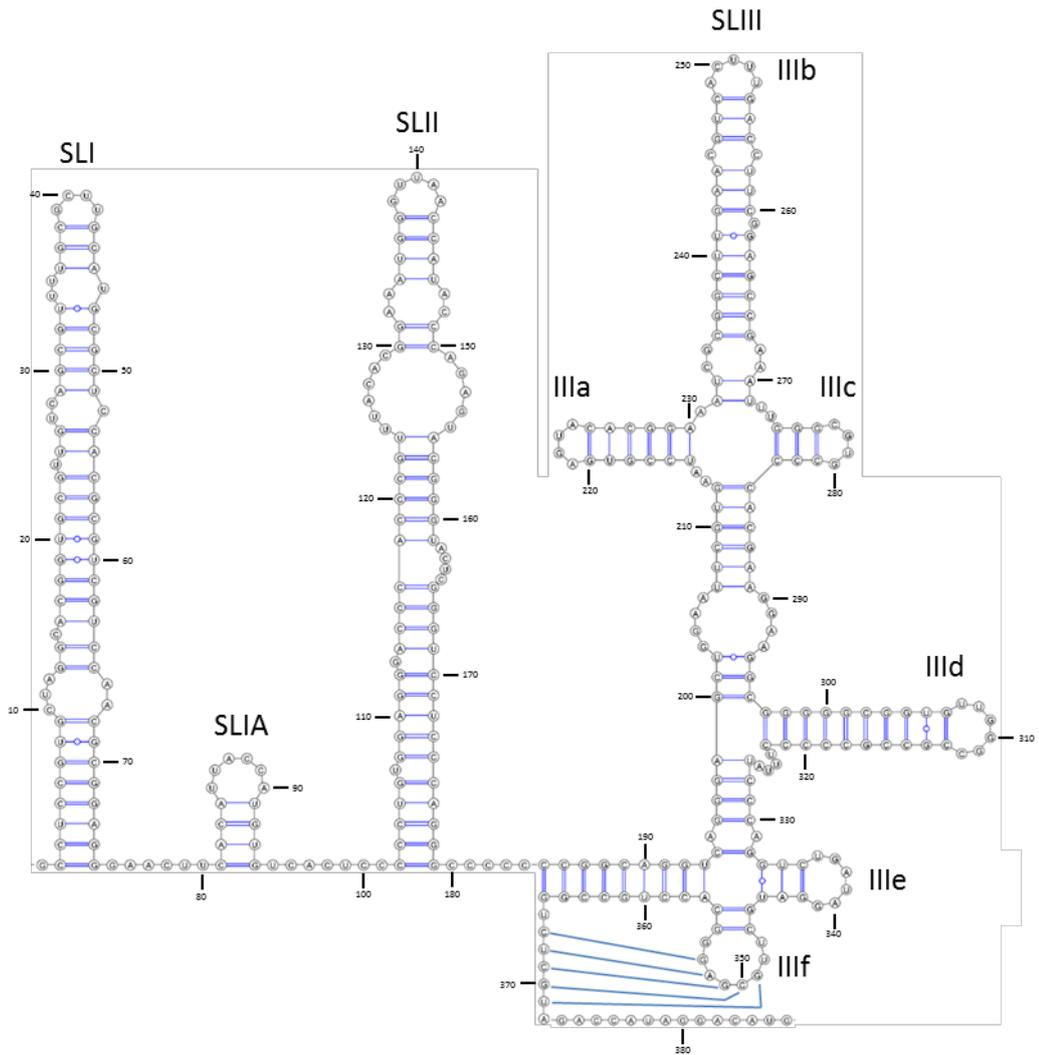


Figure 3.6: The secondary structure of the EqHV 5'UTR

A schematic representation of the EqHV 5'UTR RNA secondary structure. SHAPE analysis of the EqHV 5'UTR in the SGR was carried out and these data read as a pseudo free energy constraint in RNAstructure to predict the secondary structure. The pseudoknot was manually mapped and the base pairs predicted to form in this structure are represented as blue lines.

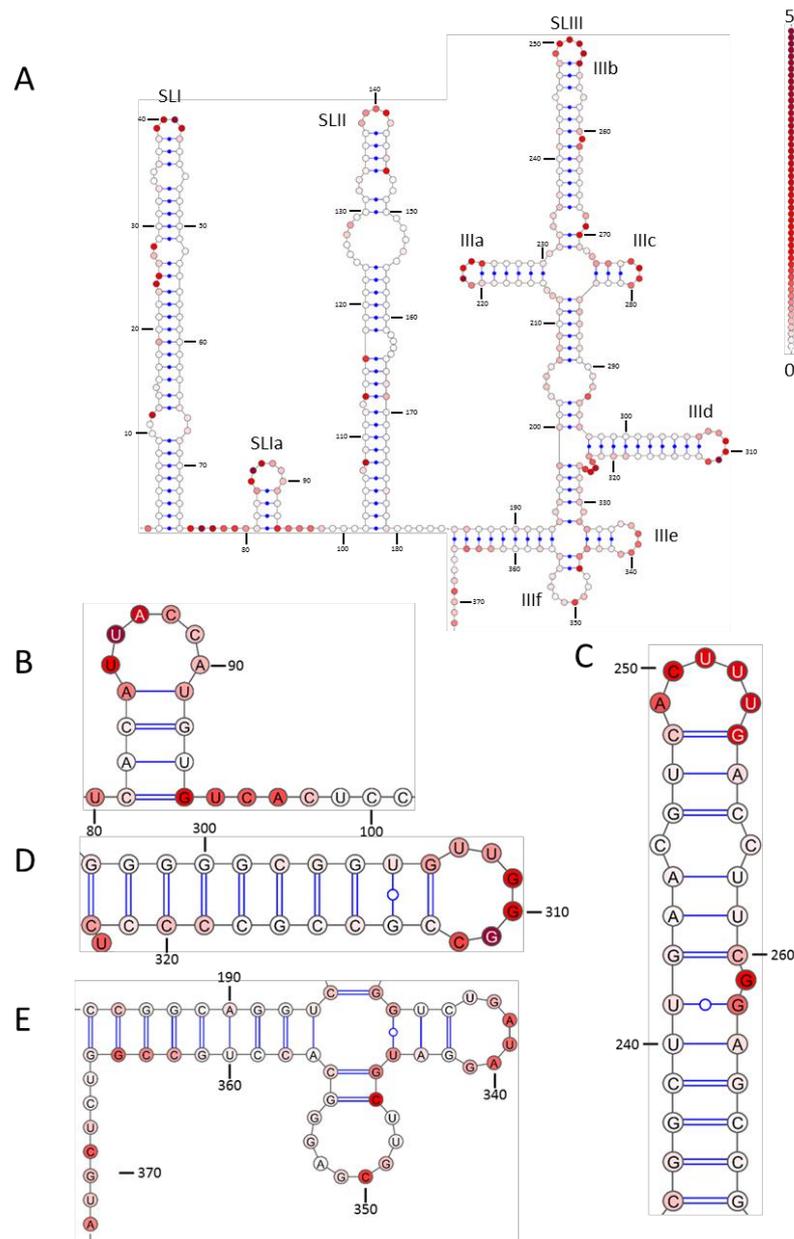


Figure 3.7: Experimental structure of the WT 5'UTR in SGR

The secondary structure of the A) WT 5'UTR B) SLIA and the miR122 seed site C) the apical portion of SLIIIB D) SLIIID and E) SLIIIE, f, and the pseudoknot in the SGR as predicted by RNA structure using experimentally obtained SHAPE reactivity values as a pseudo free energy constraint. SHAPE reactivities are represented on a colour scale from white (low reactivity – predicted paired) to red (high reactivity – predicted unpaired) a representative scale denoting reactivity increments of 0.1 is displayed. Every 10th nucleotide is number labelled and the number labelling in B, C, D, and E, corresponds to the number labelling in A.

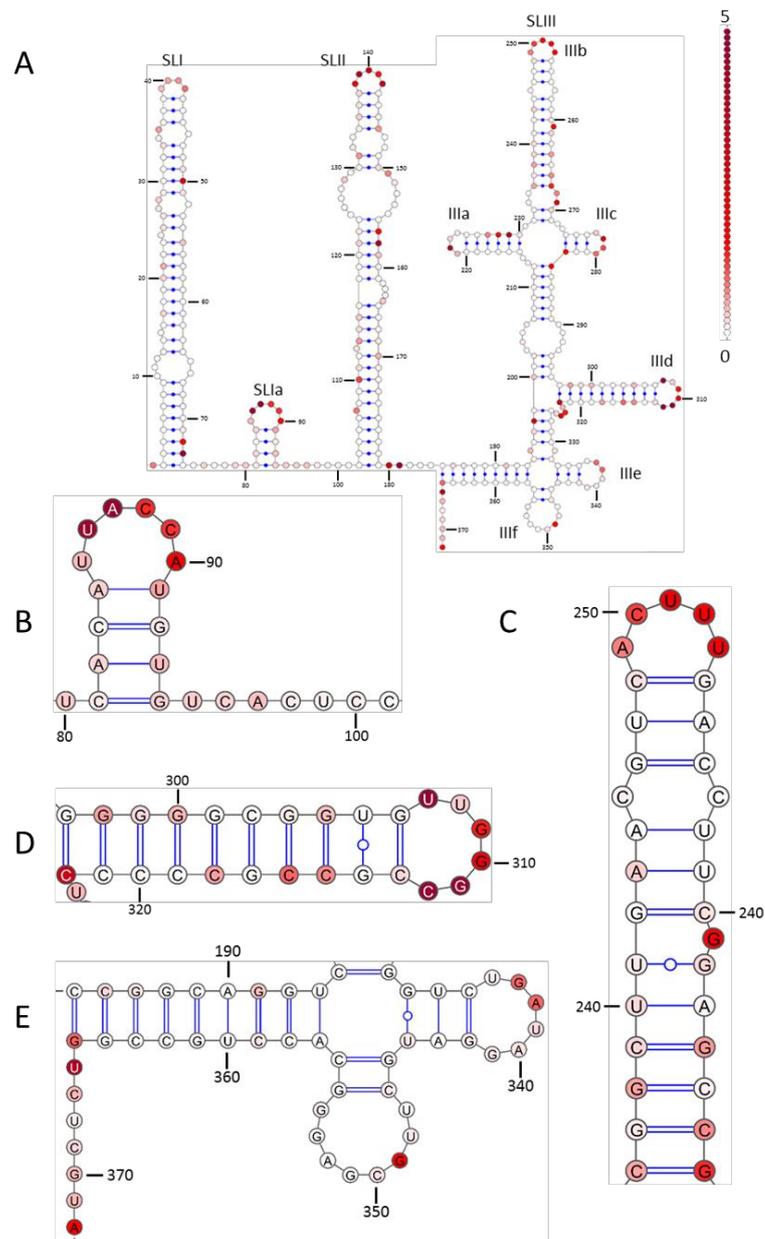


Figure 3.8: Experimental structure of the EqHV 5'UTR in the bicistronic construct

The secondary structure of the A) WT 5'UTR B) SLIIa and the miR122 seed site C) the apical portion of SLIIIb D) SLIIId and E) SLIIIf, and the pseudoknot in the bicistronic construct mapped onto the WT EqHV 5'UTR in the SGR as in Figure 4.3. SHAPE reactivities are represented on a colour scale from white (low reactivity – predicted paired) to red (high reactivity – predicted unpaired) a representative scale denoting reactivity increments of 0.1 is displayed. Every 10th nucleotide is number labelled and the number labelling in B, C, D, and E corresponds to the number labelling in A.

3.2.4 SHAPE analysis of the Δ SLI 5'UTR

In either the SGR (Figure 3.9) or bicistronic construct (Figure 3.10) the deletion of SLI did not appear to cause any major structural rearrangements. One striking observation is the overall structural similarity that the EqHV Δ SLI 5'UTR appears to share with the HVC 5'UTR. However, upon closer inspection of the sequence of SLIA, and the comparable HCV sequence, there appears to be no significant conservation. In fact SLIA more closely corresponds to the miR122 seed site 1 in HCV. This intermediary SL appears to constitute a divergence in structure, rather than a conservation. The Δ SLI 5'UTR retains all of the structural features as in the WT (excluding SLI which was deliberately removed) and this deletion does not appear to cause any aberrant effects on the overall structure. One interesting observation that can be made upon closer inspection of the miR122 seed site in the SGR (Figure 3.9B) is an apparent reversal in the trend observed in Figure 3.7B and 3.8B in that this site is less reactive at the 5' than the 3' end of the sequence. However, the data for all of the other structural features appears to be in good agreement. The miR122 seed site in the bicistronic construct (Figure 3.10B) retains the 5' to 3' loss of NMIA reactivity in the miR122 seed site seen in WT and the SHAPE reactivities in Figure 3.10 agree well with the SGR Δ SLI data. As with the WT there are some differences between the SGR and bicistronic reactivity data, these are not, however, predicted to cause any structural changes and are likely due to experimental variability.

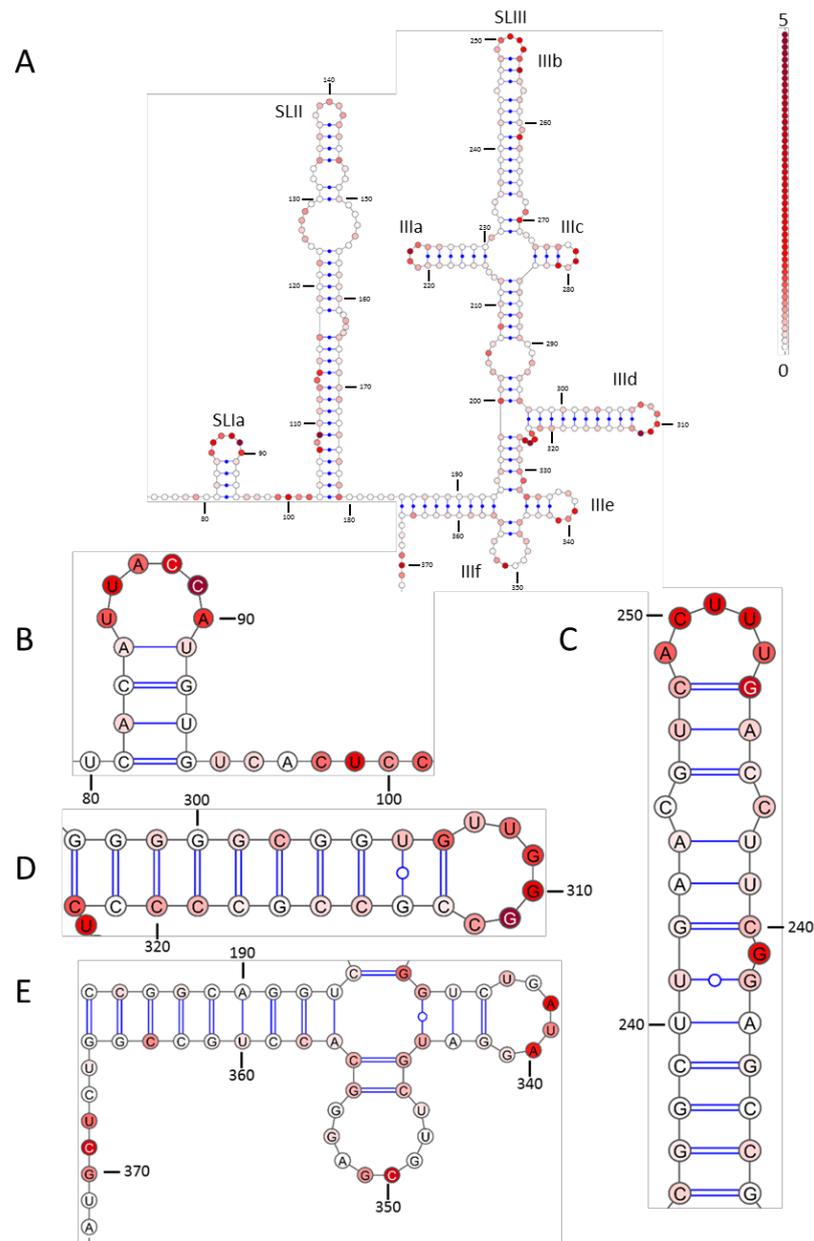


Figure 3.9: Experimental structure of EqHV Δ SLI 5'UTR in the SGR

The secondary structure of the A) Δ SLI 5'UTR B) SLIA and the miR122 seed site C) the apical portion of SLIIIB D) SLIIID and E) SLIIIE, f, and the pseudoknot in the SGR mapped onto the Δ SLI structure as predicted from the WT EqHV 5'UTR in the SGR in Figure 4.3. SHAPE reactivities are represented on a colour scale from white (low reactivity – predicted paired) to red (high reactivity – predicted unpaired) a representative scale denoting reactivity increments of 0.1 is displayed. Every 10th nucleotide is number labelled and the number labelling in B, C, D, and E corresponds to the number labelling in A.

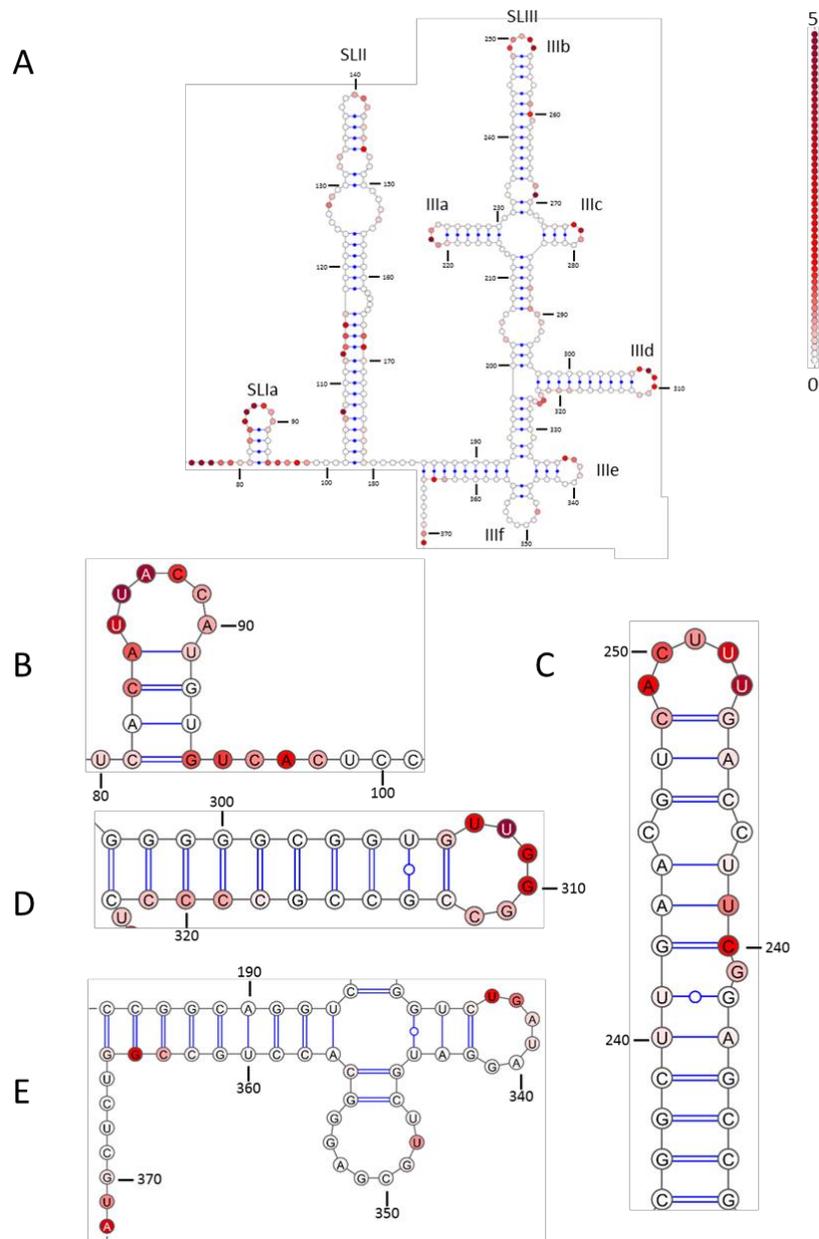


Figure 3.10: Experimental structure of EqHV Δ SLI 5'UTR in the bicistronic construct

The secondary structure of the A) Δ SLI 5'UTR B) SLIa and the miR122 seed site C) the apical portion of SLIIIb D) SLIIId and E) SLIIIe, f, and the pseudoknot in the bicistronic construct mapped onto Δ SLI structure as predicted from the WT EqHV 5'UTR in the SGR as in Figure 4.5. SHAPE reactivities are represented on a colour scale from white (low reactivity – predicted paired) to red (high reactivity – predicted unpaired) a representative scale denoting reactivity increments of 0.1 is displayed. Every 10th nucleotide is number labelled and the number labelling in B, C, D, and E corresponds to the number labelling in A.

3.2.5 SHAPE analysis of the Δ SLI+II 5'UTR

SHAPE reactivities for Δ SLI+II mapped extremely well onto the predicted structure for both the SGR (Figure 3.11) and bicistronic (Figure 3.12) constructs. Highly reactive bases largely mapped to unpaired regions, and low reactivity to paired regions. The patterns of reactivity also appeared to be largely similar between both constructs within the main structural regions which are focussed on in this study; excluding SLIA and the miR122 seed site which were deleted as part of this mutation. The apical loops of both SLIIIb (Figure 3.11B and 3.12B) and SLIIIc (Figure 3.11C and 3.12C) remained highly reactive, and the pattern of reactivity seen in SLIIIe, f, and the pseudoknot (Figure 3.11D and 3.12D) was similar to WT. These data indicate that the deletion of SLI+II in either the bicistronic construct, or in the SGR, resulted in a 5'UTR consisting only of SLIII and the pseudoknot as they are in the WT 5'UTR.

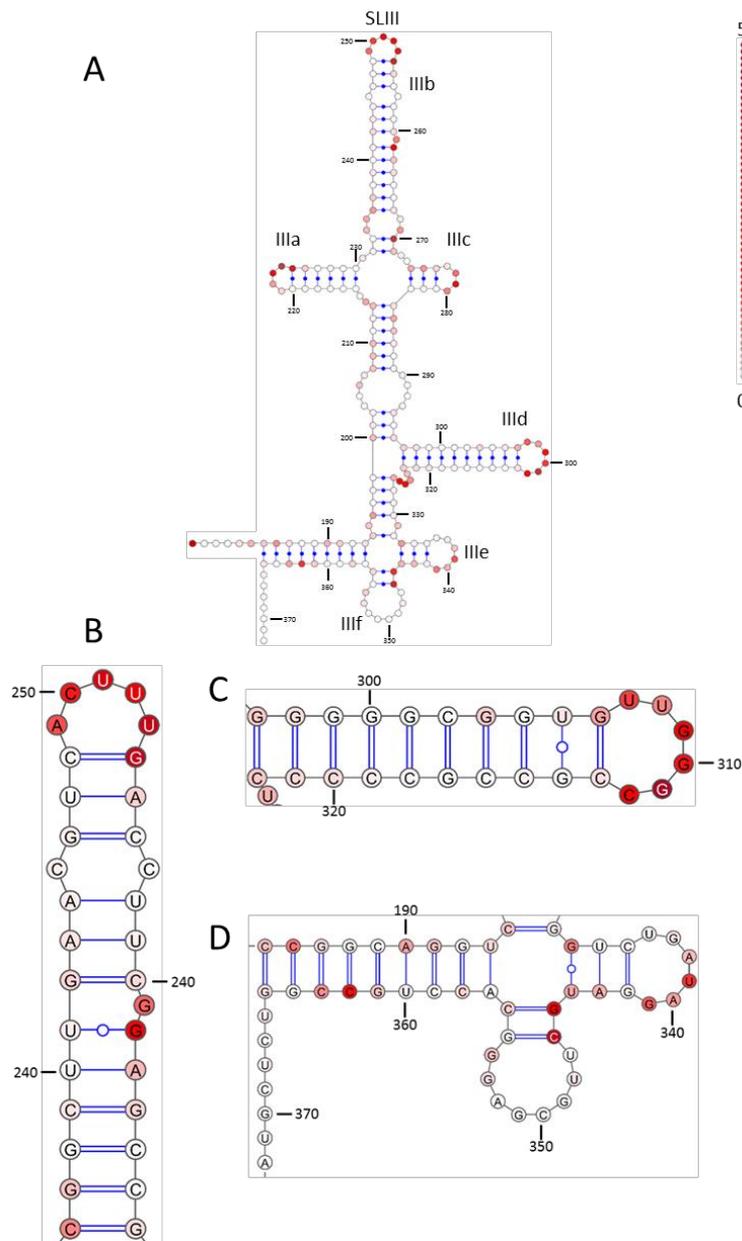


Figure 3.11: Experimental structure of EqHV Δ SLI+II 5'UTR in the SGR

The secondary structure of the A) Δ SLI+II 5'UTR B) the apical portion of SLIIIb C) SLIIId and D) SLIIIe, f, and the pseudoknot in the SGR mapped onto the Δ SLI+II structure as predicted from the WT EqHV 5'UTR in the SGR in Figure 4.3. SHAPE reactivities are represented on a colour scale from white (low reactivity – predicted paired) to red (high reactivity – predicted unpaired) a representative scale denoting reactivity increments of 0.1 is displayed. Every 10th nucleotide is number labelled and the number labelling in B, C, and D corresponds to the number labelling in A.

3.2.6 SHAPE analysis of the Δ SLIII 5'UTR

Whilst it was predicted to retain both SLI and SLII in the same confirmation as the WT, and care was taken to retain the sequences involved in the formation of the pseudoknot, it was possible that this large structural deletion could potentially disrupt the global structure of the 5'UTR; especially due to its proximity to the pseudoknot. However, across SLI+II SHAPE reactivities mapped well onto the structure obtained from the WT analysis, with well-defined areas of reactivity and non-reactivity which corresponded well with unpaired and paired regions, respectively, for both the SGR (Figure 3.13) and bicistronic (Figure 3.14) constructs. In both of these constructs SLIA was also maintained and the miR122 seed site showed intermediate reactivity across its entirety (Figures 3.13B and 3.14B). Although this does not completely agree with the data shown for the WT, reactivity here fits well with this region being single stranded and so the differences observed are not considered detrimental to the overall secondary structure. The apical loop of SLIIIe also demonstrated intermediate reactivity across its entirety and the apical loop of SLIII f was largely unreactive (Figures 3.13C and 3.14C), as it is for the WT (Figure 3.7E). There was some increase in SHAPE reactivity across the base paired sequence ${}_{220}\text{CUGCCG}_{226}$ which could potentially have induced a change to the structural organisation of the pseudoknot. However these values are indicative of intermediate to low reactivity, as are those in ${}_{227}\text{U}_{228}\text{U}_{231}\text{G}_{232}\text{U}$, therefore a complete disruption of base pairing is highly unlikely. For the bicistronic construct the pseudoknot region demonstrated a slight increase in reactivity (Figure 3.14C) but to a lesser extent than the SGR (Figure 3.13C), allowing for the conclusion that structure in this region was also retained.

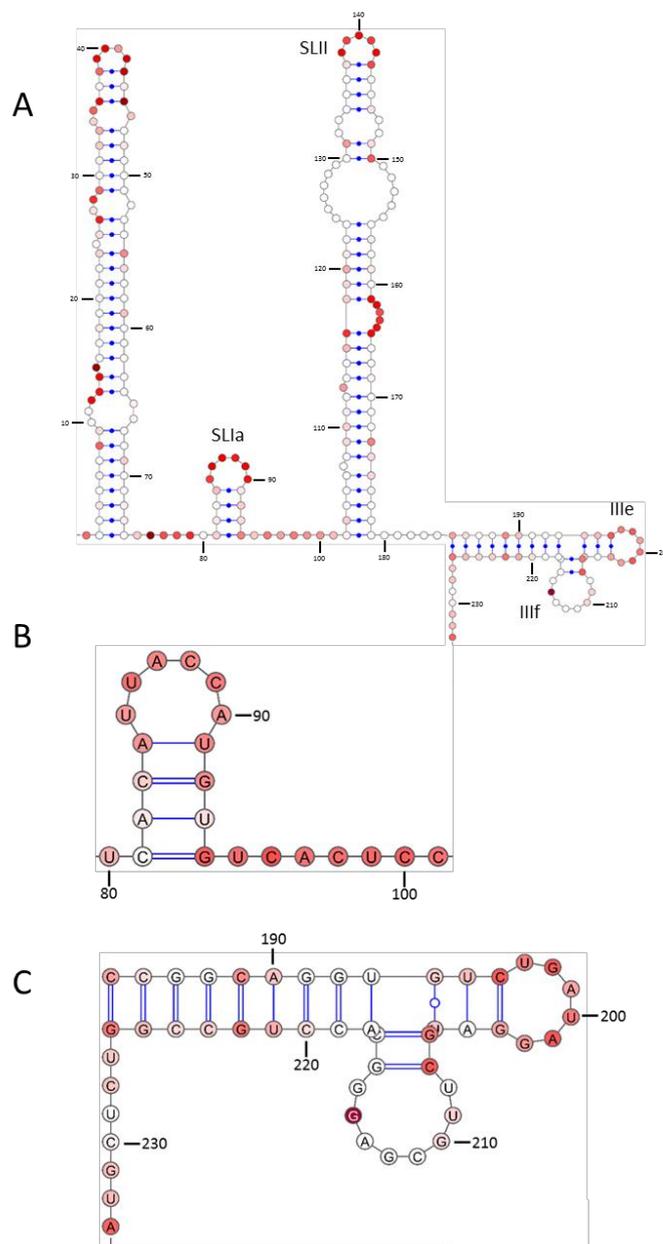


Figure 3.13: Experimental structure of EqHV Δ SLIII 5'UTR in the SGR

The secondary structure of the A) Δ SLIII 5'UTR B) SLIa and the miR122 seed site and C) SLIIIe, f, and the pseudoknot in the SGR mapped onto the Δ SLIII structure as predicted from the WT EqHV 5'UTR in the SGR in Figure 4.3. SHAPE reactivities are represented on a colour scale from white (low reactivity – predicted paired) to red (high reactivity – predicted unpaired) a representative scale denoting reactivity increments of 0.1 is displayed. Every 10th nucleotide is number labelled and the number labelling in B and C corresponds to the number labelling in A.

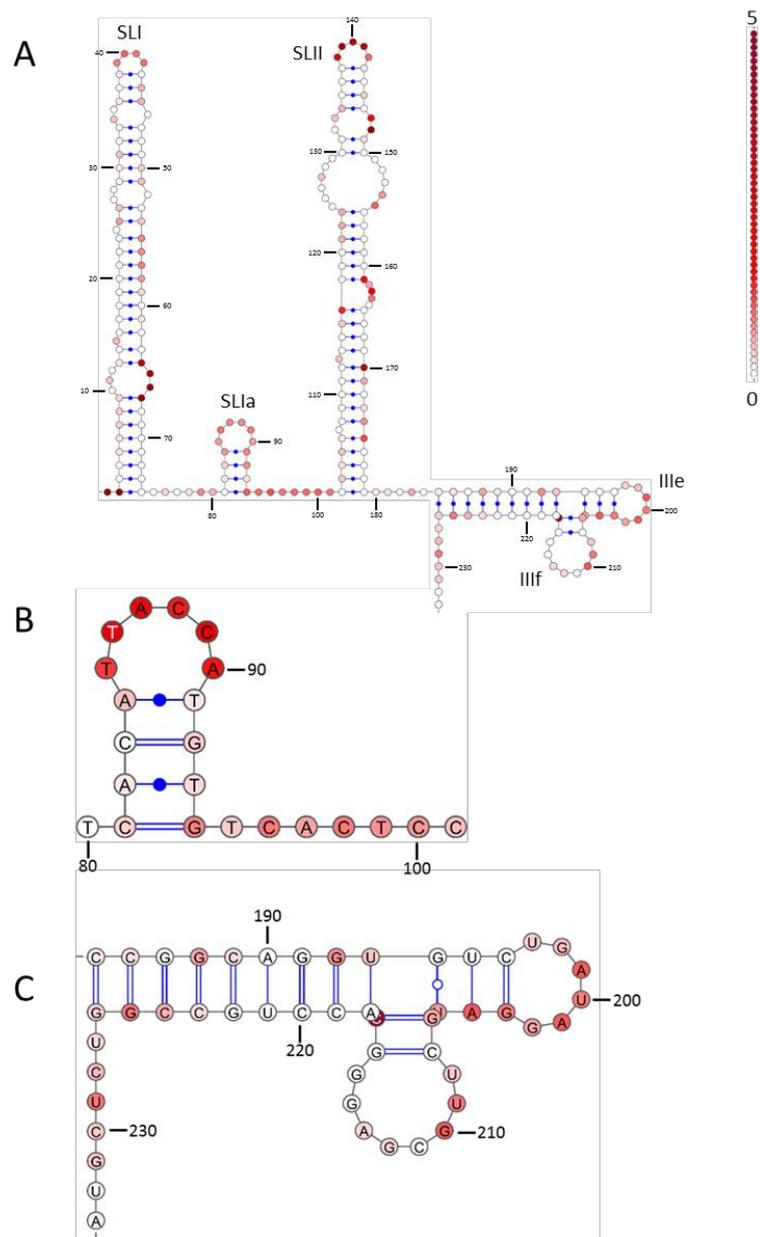


Figure 3.14: Experimental structure of EqHV Δ SLIII 5'UTR in the bicistronic construct

The secondary structure of the A) Δ SLIII 5'UTR B) SLIIa and the miR122 seed site and C) SLIIIe, f, and the pseudoknot in the bicistronic construct mapped onto the Δ SLIII structure as predicted from the WT EqHV 5'UTR in the SGR as in Figure 4.9. SHAPE reactivities are represented on a colour scale from white (low reactivity – predicted paired) to red (high reactivity – predicted unpaired) a representative scale denoting reactivity increments of 0.1 is displayed. Every 10th nucleotide is number labelled and the number labelling in B and C corresponds to the number labelling in A.

3.2.7 SHAPE analysis of the Δ SLIIIb 5'UTR

The SHAPE reactivities for both the SGR (Figure 3.15) and bicistronic (Figure 3.16) constructs were in good agreement with those seen across SLI and SLII for the WT (Figure 3.7), indicating maintenance of these structures in Δ SLIIIb. A reduction in the reactivity in the apical loop of SLII was observed in the SGR, however some of these bases still exhibited intermediate reactivity and, given the agreement demonstrated for the other regions in the 5' portion, it was not deemed that this reduction in reactivity would result in a significant structural rearrangement. SLIA was also maintained, displaying the characteristic reduction in reactivity in the miR122 seed site (Figures 3.15B and 3.16B) that is evident in almost all of the other constructs analysed. The apical loops of SLIIIa and c remained highly reactive in Δ SLIIIb, with a highly reactive single stranded region linking them. Furthermore, reactivities in SLIII d for both constructs (Figures 3.15C and 3.16C) agreed well with the WT (Figure 3.7D) and maintained high reactivity in the apical loop. SHAPE reactivity values for SLIII e and f were also in good agreement with those in the WT, as were those for the bases predicted to form the pseudoknot. These data show that the overall structure of SLIII is maintained upon the deletion of SLIIIb, with this deletion producing a truncated SLIII where SLIIIa and SLIIIc are linked by a single stranded region in place of SLIIIb.

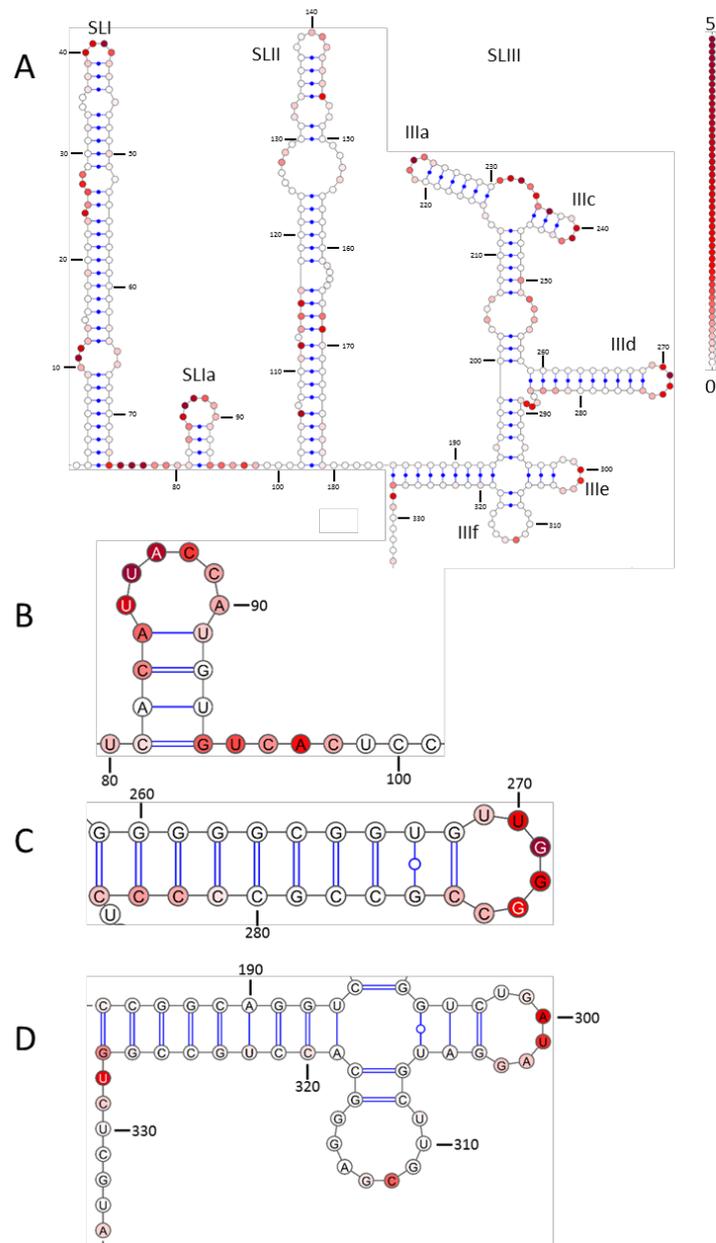


Figure 3.15: Experimental structure of EqHV Δ SLIIIb 5'UTR in the SGR

The secondary structure of the A) Δ SLIIIb 5'UTR B) SLIIa and the miR122 seed site C) SLIIId and D) SLIIIf, and the pseudoknot in the SGR mapped onto the Δ SLIIIb structure as predicted from the WT EqHV 5'UTR in the SGR in Figure 4.3. SHAPE reactivities are represented on a colour scale from white (low reactivity – predicted paired) to red (high reactivity – predicted unpaired) a representative scale denoting reactivity increments of 0.1 is displayed. Every 10th nucleotide is number labelled and the number labelling in B, C, and D corresponds to the number labelling in A.

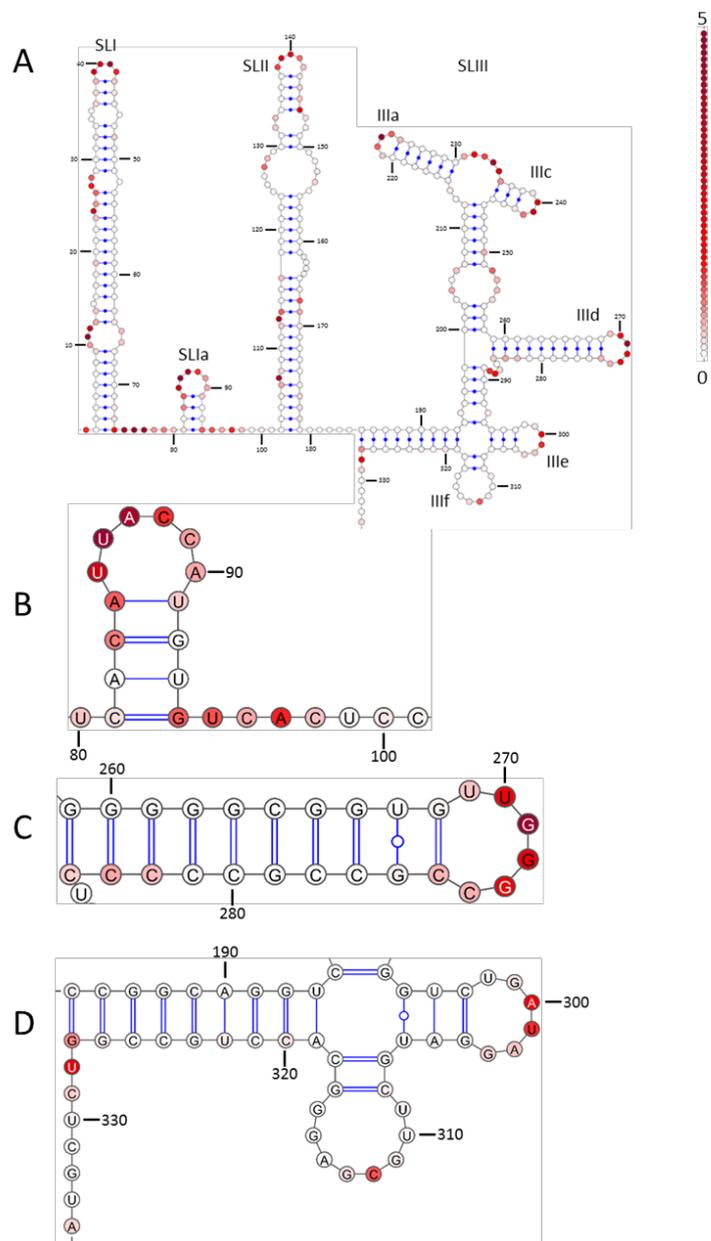


Figure 3.16: Experimental structure of EqHV Δ SLIIIb 5'UTR in the bicistronic construct

The secondary structure of the A) Δ SLIIIb 5'UTR B) SLIA and the miR122 seed site C) SLIIId and D) SLIIIe, f, and the pseudoknot in the bicistronic construct mapped onto the Δ SLIIIb structure as predicted from the WT EqHV 5'UTR in the SGR as in Figure 4.11. SHAPE reactivities are represented on a colour scale from white (low reactivity – predicted paired) to red (high reactivity – predicted unpaired) a representative scale denoting reactivity increments of 0.1 is displayed. Every 10th nucleotide is number labelled and the number labelling in B, C, and D, corresponds to the number labelling in A.

3.2.8 SHAPE analysis of the Δ SLIIId 5'UTR

The GGG motif has been shown to be involved in proper folding of the HCV 5'UTR (Jubin et al., 2000) therefore demonstrating correct folding of this 5'UTR in the absence of SLIIId, and the GGG motif, was essential.

For both the SGR (Figure 3.17) and bicistronic (Figure 3.18) constructs the SHAPE reactivity profile through SLI+II matched well with that of WT. SLIA was again maintained and this region and demonstrated a, now somewhat characteristic, reduction in reactivity from 5' to 3' of the miR122 seed site (Figure 3.17B and 3.18B). SLIIa, b, and c were all maintained, with high reactivities in their respective apical loops (a detailed view of SLIIb can be seen in Figures 3.17C and 3.18C). Similarly to Δ SLIIb, the deletion of SLIIId appeared to simply result in a single stranded "linker" region which bridged the gap where SLIIId previously was. This region was shown to be highly NMIA reactive in both constructs. Finally, SLIIe and SLIIIf appeared to be well maintained, with similar SHAPE reactivity profiles both between constructs and in comparison to WT (Figure 3.17D and 3.18D) and ${}_{185}\text{CCGGCAGGU}_{193}/{}_{357}\text{ACCUGCCGG}_{365}$ maintained low reactivity, indicating decreased structural freedom. These results demonstrate that the predicted secondary structure of the Δ SLIIId 5'UTR (Figure 4.13) is a good representation of its structure *in vitro*.

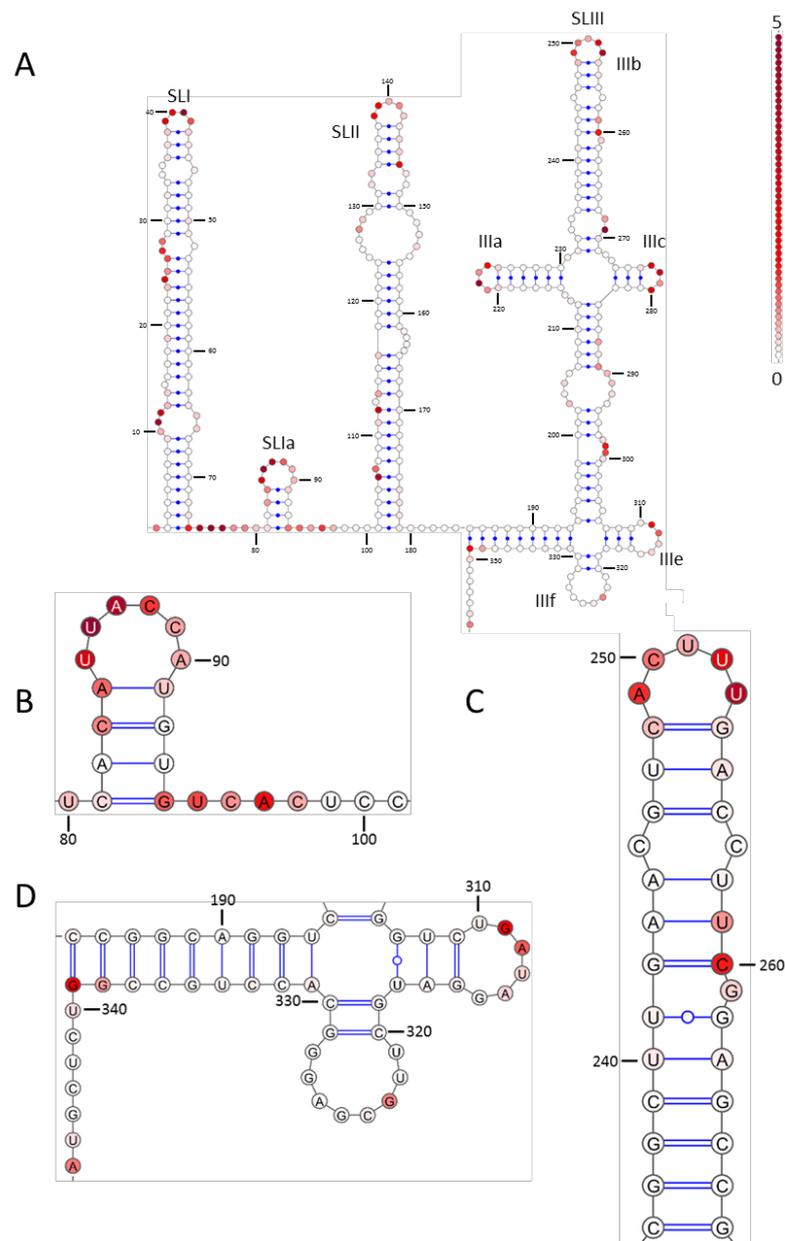


Figure 3.17: Experimental structure of EqHV Δ SLIId 5'UTR in the SGR

The secondary structure of the A) Δ SLIId 5'UTR B) SLIA and the miR122 seed site C) SLIIIb and D) SLIIIe, f, and the pseudoknot in the SGR mapped onto the Δ SLIId structure as predicted from the WT EqHV 5'UTR in the SGR in Figure 4.3. SHAPE reactivities are represented on a colour scale from white (low reactivity – predicted paired) to red (high reactivity – predicted unpaired) a representative scale denoting reactivity increments of 0.1 is displayed. Every 10th nucleotide is number labelled and the number labelling in B, C, and D corresponds to the number labelling in A.

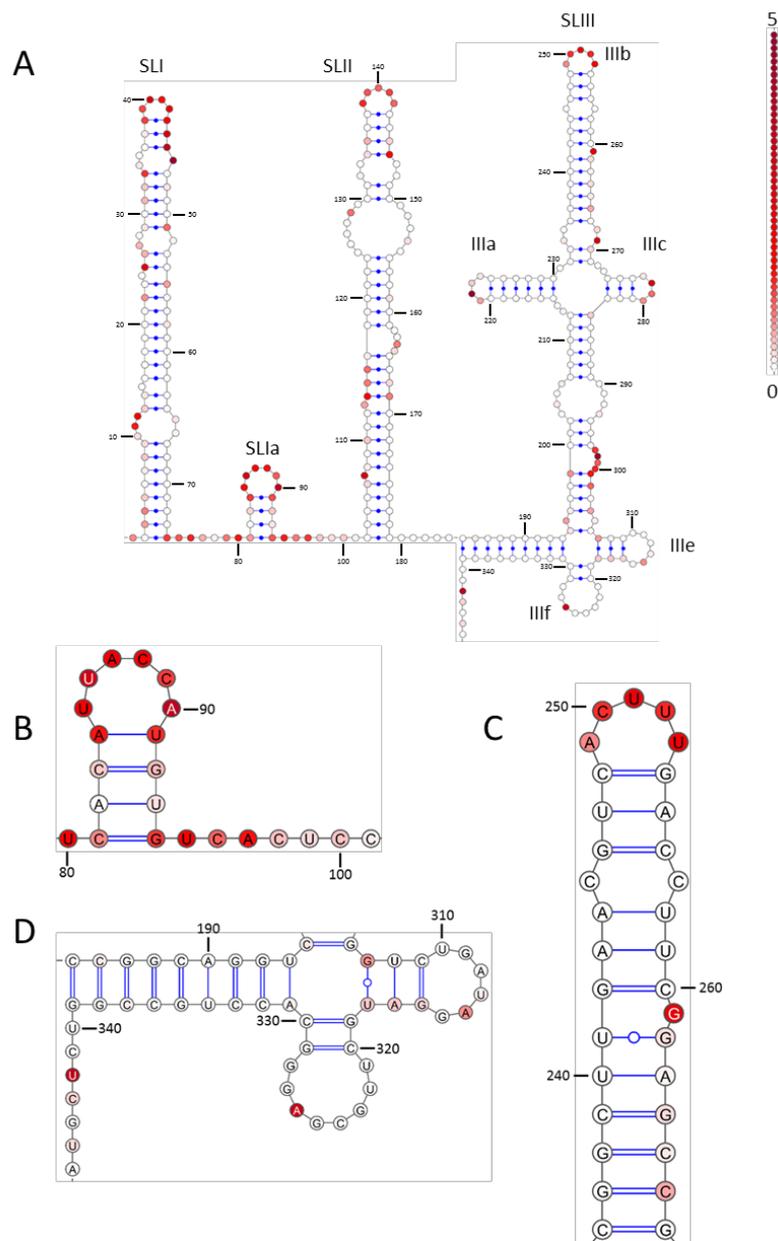


Figure 3.18: Experimental structure of EqHV Δ SLIIId 5'UTR in the bicistronic construct

The secondary structure of the A) Δ SLIIId 5'UTR B) SLIIa and the miR122 seed site C) SLIIb and D) SLIIe, f, and the pseudoknot in the bicistronic construct mapped onto the Δ SLIIId structure as predicted from the WT EqHV 5'UTR in the SGR as in Figure 4.13. SHAPE reactivities are represented on a colour scale from white (low reactivity – predicted paired) to red (high reactivity – predicted unpaired) a representative scale denoting reactivity increments of 0.1 is displayed. Every 10th nucleotide is number labelled and the number labelling in B, C, and D corresponds to the number labelling in A.

3.2.9 Investigating the activity of EqHV 5'UTR mutants in a bicistronic construct

In order to assay the IRES activity of the 5'UTR deletion mutants, each was introduced into an RLuc/FLuc pGL3 bicistronic expression construct by Q5 SDM (NEB). Bicistronic expression constructs are widely used for analysing IRES activity and express two reporter genes from which translation is driven in a cap and IRES-dependent fashion, respectively. Upon transfection into a cell the construct is transcribed into capped mRNA encoding both reporter genes, separated by the IRES under investigation. Translation of RLuc was driven by the m⁷G cap and FLuc by the IRES (Figure 3.19A).

To assess IRES activity from these constructs, 1.5x10⁵ Huh7, FHK, or 293T cells were transfected with 2 µg DNA and assayed for dual luciferase activity 24 hours post transfection (h.p.t). For each mutant, and cell line, the ratio of FLuc:RLuc was calculated and used as a measure of relative efficiency of translation from the IRES in relation to that of the m⁷G cap. A pGL3 plasmid which contained no IRES was used to control for read through from RLuc to FLuc in the absence of an IRES. Each experiment was subject to a minimum of N=3 biological repeats and a student's T test calculated to determine statistical significance.

ΔSLI showed no significant impact on translation in either Huh7 (Figure 3.19B), FHK (Figure 3.19C), or 293T (Figure 3.19D) cells. This is in good agreement with previously published data (Stewart et al., 2013) on the translational efficiency of the ΔSLI 5'UTR and indicates that SLI does not play a role in translation from the EqHV IRES. Whilst the effect of the ΔSLI deletion was consistent through all three cell lines which were tested, the ΔSLI+II deletion exhibited different effects in each cell line. Whilst ΔSLI+II significantly inhibited translation in both Huh7 and FHK cells, the degree of inhibition was far greater in FHK cells. The same deletion in 293T cells exhibited no significant reduction in translation; however there was a trend toward inhibition (p=0.059). Taken together these data indicate that the presence of SLI+II influences the ability of the EqHV 5'UTR to initiate translation. However, the variability in the effect that the ΔSLI+II deletion had on translation efficiency between the cells lines may indicate that SLII is not playing a direct role in translation initiation from the EqHV IRES.

The deletion of SLIII, however, caused an almost complete ablation of translation with >95% inhibition in all three cell lines tested. ΔSLIIIb, on the other hand, demonstrated much less of an effect on EqHV IRES function which, similarly to ΔSLI+II, was significant only in Huh7 and FHK

cells but not in 293T cells. This indicated that SLIIIb is not necessary for IRES function and that it may not be directly involved in translation initiation. Finally, Δ SLIIId exhibited complete ablation of translation with >99% inhibition in all three cell lines tested. Taken together these data suggest that SLI+II, and SLIIIb, are not required for EqHV IRES mediated translation but that SLIII, specifically SLIIId, is necessary.

In order to investigate the effect of miR122 on translation from the EqHV IRES, FHK and 293T cells, which do not express miR122, were transduced to overexpress this liver specific microRNA. The transduced cells were selected, and maintained, in 5 μ g/ml puromycin to ensure that a population of cells expressing miR122 was maintained. To confirm that these puromycin resistant cells were expressing miR122 and had not simply acquired puromycin resistance, two reporter plasmids, pGL3 MCS and pGL3 1225, were utilised. pGL3 1225 is a miR122 responsive plasmid which contains a perfect miR122 seed site within the luciferase reporter gene, which is cleaved upon binding to miR122 resulting in downregulation of expression. pGL3 MCS does not contain such a site and was used to control for any differences in transfection efficiency that may arise from the over expression of miR122. Both constructs were transfected into FHK 122 and 293T 122 cells, harvested 24 h.p.t, and assayed for luciferase activity. Figure 3.20 represents a minimum of N=3 biological repeats for both cell lines and demonstrates that expression of miR122, in either cell line, did not have any significant effect on transfection efficiency (Figure 3.20A). Both the FHK 122 and 293T 122 cell lines were confirmed to express miR122, as evidenced by the significant reduction in luciferase activity in the miR122 expressing cell lines (Figure 3.20B).

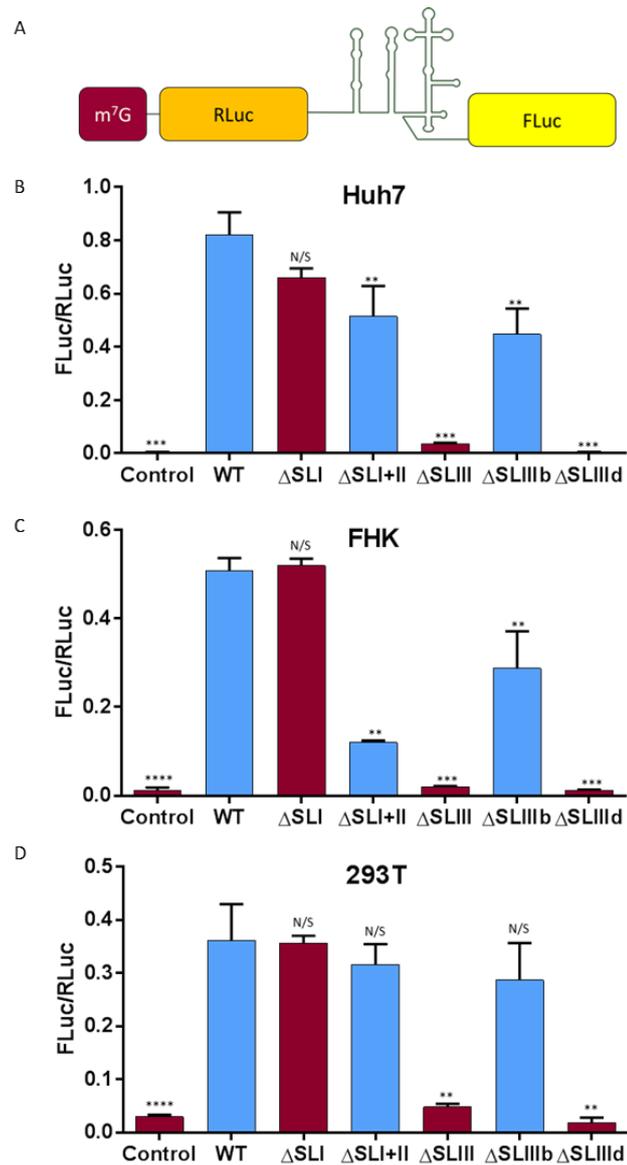
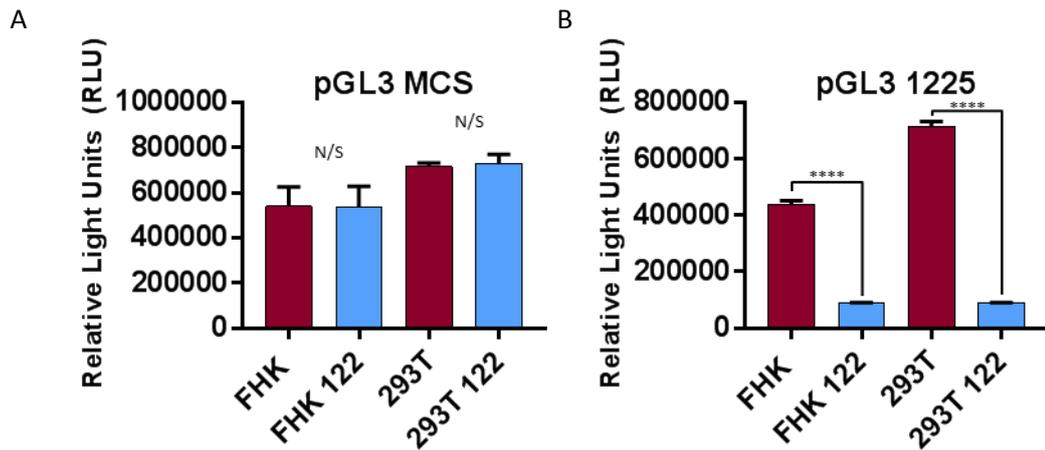


Figure 3.19: EqHV mutant IRES activity in an RLuc/FLuc bicistronic construct

A) A schematic representation of the pGL3 bicistronic construct containing the WT EqHV 5'UTR. B, C, and D) pGL3 N (where N represents either WT or mutant EqHV, or a control plasmid containing no IRES) was transfected into B) Huh7 C) FHK, or D) 293T cells, harvested 24 h.p.t and assayed for dual luciferase activity. All values are represented as a ratio of FLuc:RLuc and this is used as a representative measure of IRES activity. All data was subject to a student's T test in comparison to the WT value for each respective cell line. Stars represent statistically significant values and N/S represents values which were not significant.



A) A control luciferase reporter *pGL3 MCS* or B) the *miR122* responsive *pGL3 1225* was

Figure 3.20: Assay of *miR122* expression in FHK and 293T cells

transfected into FHK, FHK 122, 293T, and 293T 122 cell lines and assayed 24 h.p.t for luciferase activity. Data was subject to a student's T test comparing each construct against its corresponding transfection into the 122 cell line. Stars highlight statistically significant differences whilst N/S represents values which were not significant.

Following confirmation that the 122 cell lines were expressing miR122 all 5'UTR mutants, WT, and control plasmid, were transfected into the FHK 122 and 293T 122 cell lines, harvested 24 h.p.t, and assayed for dual luciferase activity as previously described (Figure 3.21 A and B). The same trends were observed in the miR122 expressing cell lines as was observed in the parental cell lines (Huh7 cells were omitted from these experiments as they endogenously express miR122). However it could have been possible that the expression of miR122 was effecting each mutant to an equal extent, thus masking any effect of miR122 on translation. Therefore the translation efficiency of the WT 5'UTR in the parental cell lines was compared to its efficiency in the miR122 expressing cell lines in both FHK, and 293T cells (Figures 3.21C and D). These data showed that efficiency of translation from the WT IRES is not influenced by the presence of miR122 in either FHK or 293T cells, indicating that miR122 is unable to influence translation efficiency from the EqHV IRES in the context of a bicistronic expression system.

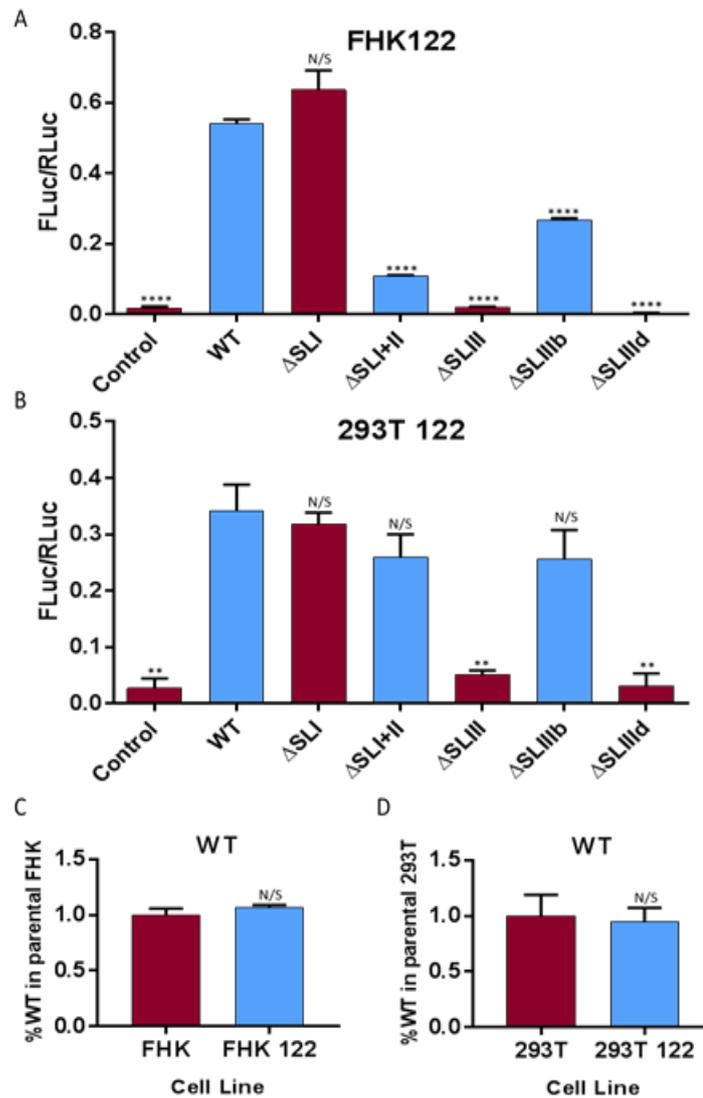


Figure 3.21: The effect of miR122 on EHV IRES activity in a bicistronic construct

pGL3 N (where *N* represents either WT or mutant EHV, or a control plasmid containing no IRES) was transfected into A) FHK 122 and B) 293T cells, harvested 24 h.p.t, and assayed for dual luciferase activity. All values are represented as a ratio of FLuc:RLuc and this is used as a representative measure of IRES activity. All data was subject to a student's *T* test in comparison to the WT value for each respective cell line. Stars represent statistically significant values and *N/S* represents values which were not significant. WT translation values from C) FHK and FHK 122 and D) 293T and 293T 122 cells were compared. Values were normalised to % of the WT value in the parental cell line and a student's *T* test carried out to determine statistical significance. *N/S* represents non-significant values.

3.2.10 Investigating the activity of EqHV 5'UTR mutants in an SGR

Whilst there are many advantages to assaying IRES activity in the bicistronic system, there are also many disadvantages to this method of analysis (as discussed in 3.1). Notably, from the observations made in 3.2.9, the EqHV 5'UTR IRES activity did not seem to be influenced by the presence of miR122, as has previously been reported (Scheel et al., 2015). The authors also reported an enhancement of translation from the EqHV 5'UTR IRES in the presence of the EqHV 3'UTR, which is not present in the bicistronic construct.

It is common practice in the field of HCV research, when conducting replication assays using SGRs, to assay an early time point, typically 4-6 hours post electroporation (h.p.e), as a measure of input translation to control for electroporation efficiency. It has previously been shown that the EqHV 5'UTR is able to efficiently drive expression of a reporter gene in the context of an SGR, which is detectable 5 h.p.e. Each of the 5'UTR mutants described in Figure 3.2, and the WT, were therefore introduced into pSGR NZCI (described in 5.2.4), which contains a codon optimised CpG/UpA low FLuc reporter gene (a schematic of which can be seen in Figure 3.22A) in order to create an alternative means of assaying EqHV IRES activity. This construct represents a more physiologically relevant system by which to investigate the EqHV 5'UTR as it is located at the proximal 5' end of the construct as it would be in the viral genome, as opposed to its internal position in the bicistronic construct. The 3'UTR is also present in this construct, unlike in the bicistronic construct, ~10 Kb downstream of the 5'UTR; again mirroring the organisation of the viral genome.

All structural deletion mutants, and the WT 5'UTR, were synthesised by GENEART and introduced into the pSGR by molecular cloning. An XbaI site directly upstream of the 5'UTR and an AscI site located at the beginning of the CpG/UpA low luciferase reporter gene were utilised to introduce the 5'UTR into pSGR NZCI. In order to assay IRES activity in this context SGR RNA was *in vitro* transcribed from linearised DNA using the ribomax T7 transcription kit (promega) according to manufacturer's instructions. Huh7 and FHK (by electroporation), and 293T cells, were all transfected with 2 µg RNA, harvested at 6 h.p.t and assayed for firefly luciferase activity. A mock transfection was carried out with RNase free water to control for background luciferase activity and each experiment was subject to a minimum of N=3 biological repeats; a student's T test was carried out to determine statistical significance.

In either Huh7 (Figure 3.22B), FHK (Figure 3.22C), or 293T (Figure 3.22C) cells Δ SLI did not demonstrate any significant effect on translation from the EqHV 5'UTR IRES; indicating SLI is not involved in EqHV translation. Deletion of SLI+II (Δ SLI+II) demonstrated a significant reduction in translation compared to WT in all three cell lines tested, however this effect was enhanced in Huh7 cells (~25% efficiency) compared to both FHK and 293T cells (~50% efficiency). These data indicate that the absence of SLI+II has a large impact on translation; however the relative inhibition of translation, compared with significant size of the deletion created for this mutant, may suggest that the impact of the Δ SLI+II deletion may not be due to a direct involvement in translation. In all cell lines tested here deletion of SLIII, SLIIIb, or SLIIIc caused a complete ablation of translation, equivalent to mock transfection levels. Taken together these data indicate that SLIII is necessary, and SLIII and the pseudoknot are sufficient, for EqHV 5'UTR IRES function. Δ SLI+II is still able to initiate translation whilst Δ SLIII, Δ SLIIIb, or Δ SLIIIc completely ablate translation. Therefore it can be concluded that SLIII, and more specifically SLIIIb and SLIIIc all play direct and necessary roles in the initiation of translation from the EqHV IRES in the context of an SGR.

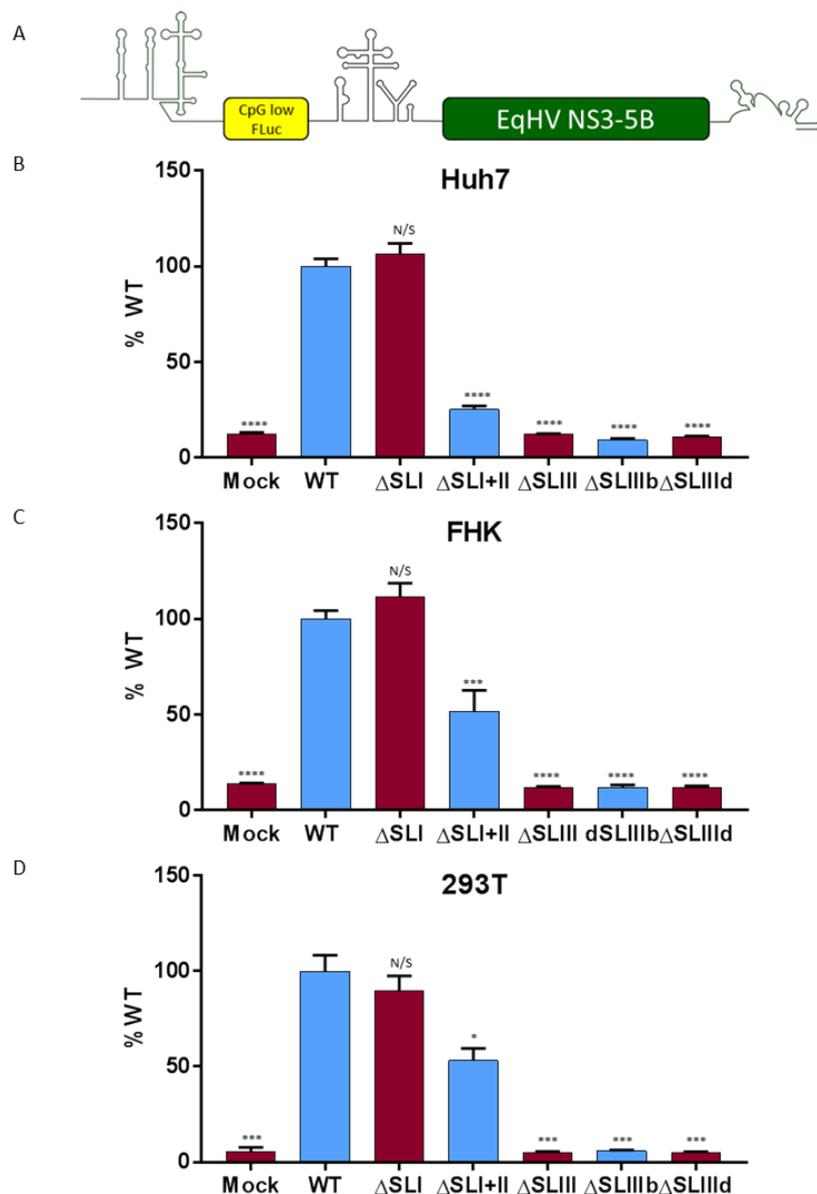


Figure 3.22: EqHV mutant IRES activity in an SGR

A) A schematic representation of SGR NZCI containing the WT EqHV 5'UTR. B, C, D) SGR NZCI N (where N represents either the WT or mutant EqHV 5'UTR) or mock (no RNA) was transfected into B) Huh7 C) FHK and D) 293T cells. Cells were harvested 6 h.p.t and assayed for firefly luciferase activity. All values are represented as a percentage of the RLU for the WT EqHV 5'UTR and this is used as a representative measure of IRES efficiency (with WT taken as 100% activity). All data was subject to a student's T test in comparison to the WT value for each respective cell line. Stars represent statistically significant values and N/S represents values which were not significant.

In order to assay the effect of miR122 on translation from the EqHV IRES the miR122 expressing cell lines described in 3.2.9 were transfected with all 5'UTR mutants and the WT EqHV 5'UTR, or mock treated, as described above (Figure 3.23A and B). The same trends were observed in the miR122 expressing cell lines as was observed in the parental cell lines (Huh7 cells were omitted from these experiments as they endogenously express miR122). However it could have been possible that the expression of miR122 was effecting each mutant to an equal extent, thus masking any differences. Therefore the translation efficiency of the WT 5'UTR in the parental cell lines was compared to its efficiency in the 122 cells (Figure 3.23C and D). This analysis showed that the overexpression of miR122 in both FHK and 293T cells resulted in a significant enhancement of translation to 228% and 129% of WT in the parental cell lines, respectively. In light of this result the miR122 seed site was changed to the corresponding miR124 seed site by Q5 SDM, and introduced into the full length 5'UTR (termed NZCI 124). SGR NZCI 124 RNA was transfected into FHK and 293T (+/- 122) cells in order to assess whether the effect that miR122 had on translation was mediated by an interaction with the miR122 seed site (Figure 3.24). In both the FHK 122 and 293T 122 cell lines NZCI 124, unlike the WT, was unresponsive to the presence of miR122 when compared to its IRES activity in the parental cell lines. Taken together with Figure 3.23C and D, these data indicate that EqHV IRES activity is enhanced by the presence of miR122, and that enhancement is mediated by the miR122 seed site located directly upstream of SLII.

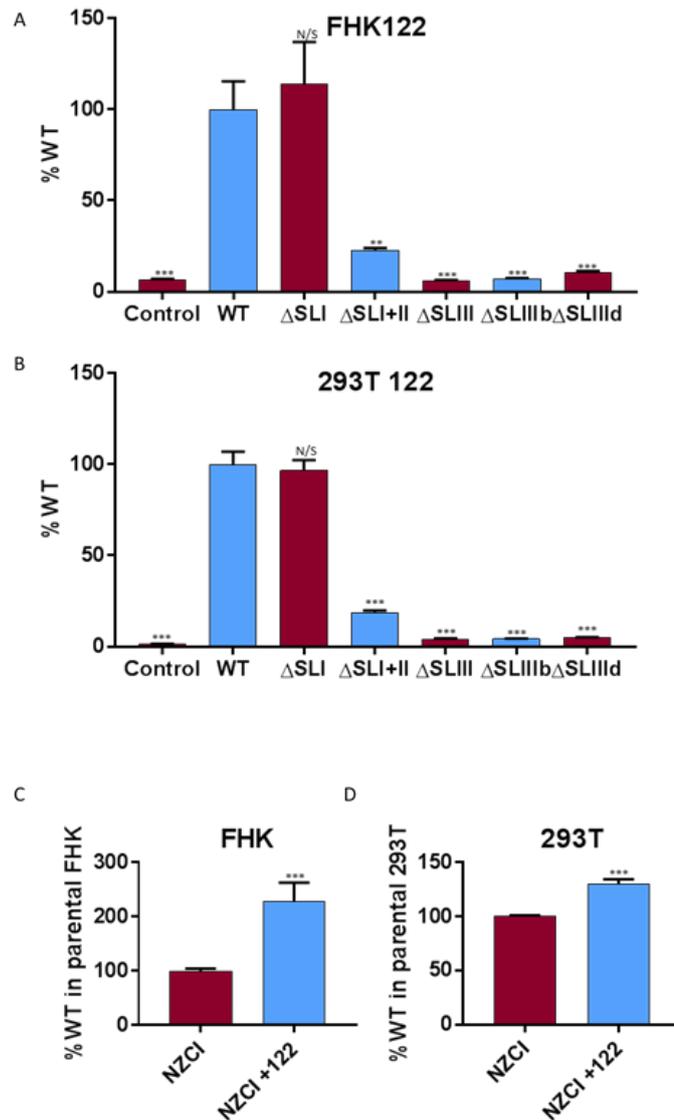


Figure 3.23: The effect of miR122 on EqHV IRES activity in an SGR

A) FHK 122 and B) 293T 122 cells were transfected, with 2 μ g RNA or 2 μ l RNase free water (mock) and assayed for firefly luciferase activity 6 h.p.e/t. All values are represented as a percentage of the RLU for the WT EqHV 5'UTR and this is used as a representative measure of IRES efficiency (with WT taken as 100% activity). All data was subject to a student T test in comparison to the WT value for each respective cell line. Stars represent statistically significant values and N/S represents values which were not significant. C, D) WT translation values from C) FHK and FHK 122 and D) 293T and 293T 122 cells was compared. Values were normalised to % of the WT value in the parental cell line and a student's T test carried out to determine statistical significance.

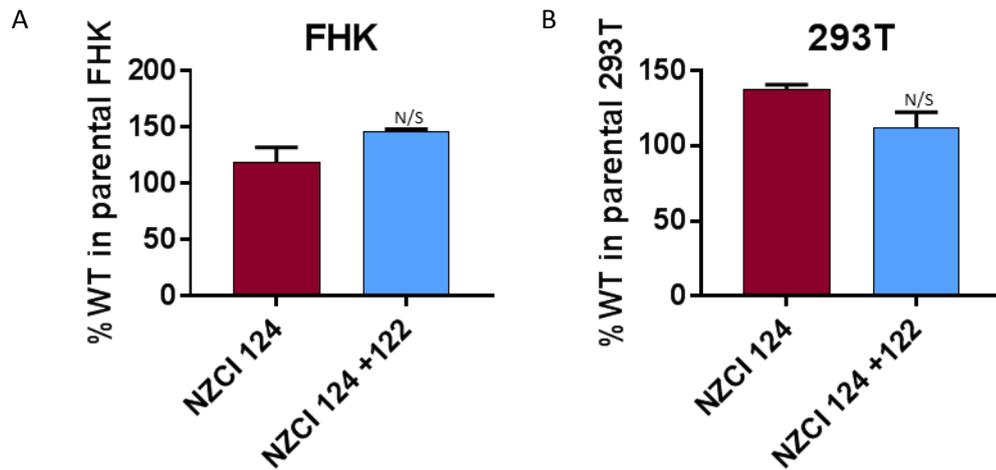


Figure 3.24: The effect of miR122 on NZCI 124

A) FHK and FHK 122 and B) 293T and 293T 122 cells were transfected with 2 μ g SGR NZCI 124 and assayed for firefly luciferase activity 6 h.p.t and normalised to NZCI 124 in the respective parental cell line. All data was subject to a student's T test in comparison to the WT value for each respective cell line to determine statistical significance. N/S represents values which were not significant.

3.3 Discussion

3.3.1 Predicting the structure of the EqHV 5'UTR, and structural mutants

The secondary structure of the EqHV 5'UTR predicted for this study agrees well with previously published predictions (Kapoor et al., 2011; Burbelo et al., 2012), indicating the presence of three main structural domains (SLI, SLII, and SLIII) and a pseudoknot. However, the presence of the putative SLIA was omitted here due to the structural divergence from HCV that this would represent and, as such, did not directly influence decision making with regards to the deletions which were made. Furthermore, due to limitations of RNA structure prediction software (discussed in 3.1) it was not possible to predict the existence of a pseudoknot within the EqHV 5'UTR. However using previously published EqHV 5'UTR predictions, and the HCV 5'UTR as a model, it was possible to map the EqHV 5'UTR sequence onto a potential pseudoknot.

To probe the role that the structural domains of the EqHV 5'UTR played in translation a panel of deletion mutants were created which systematically removed major structural components, based on the predicted structure. To reduce the possibility that these deletions exhibited indirect effects on translation by causing major structural changes in the 5'UTR, the secondary structure of each mutant was predicted in the same manner as the full 5'UTR. These predictions indicated the deletions would only exhibit the desired local effect on the secondary RNA structure and provided confidence that effects on translation would be directly influenced by each respective deletion.

3.3.2 Experimental determination of the EqHV 5'UTR RNA secondary structure

This chapter describes the first experimentally confirmed structure of the EqHV 5'UTR. In good agreement with the predicted structure (Figure 3.1) the EqHV 5'UTR contains three main structural domains (SLI, SLII, and SLIII). Whilst the existence of a pseudoknot has not been experimentally confirmed the high level of conservation with the corresponding region of the HCV 5'UTR (84% sequence identity across ₃₃₆U-A₃₇₂), SHAPE reactivities which resemble those previously published for the HCV pseudoknot (Angulo et al., 2016), the conservation of function (discussed below), and the conservation of structure, especially throughout SLIII, exhibited between these two viruses provides compelling evidence for the existence of this structure in the EqHV 5'UTR. Therefore the existence of a pseudoknot in the EqHV 5'UTR is highly probable.

Using experimental SHAPE data to inform structure predictions based on free energy is a highly accurate method of secondary structure prediction, with 97% of the *E. coli* 16S ribosome RNA base pairs being predicted, in comparison to the accepted structure, using this method (Deigan et al., 2009). This provides confidence that Figure 3.6 represents the correct confirmation of the RNA secondary structure of the EqHV 5'UTR. This section will discuss the structure of the WT 5'UTR, as was experimentally confirmed from the SGR, as it most closely resembles the 5'UTR positioning in virus and therefore is most likely to represent the 5'UTR as it exists in nature.

3.3.3 The structure of EqHV SLI, SLIA, and SLII

SLI is a large, 73 nucleotide, stem loop spanning $_{2}C-G_{75}$ with an apical loop, $_{39}G-U_{42}$, which demonstrated high SHAPE reactivity. Interestingly, the bulges in SLI show low SHAPE reactivity and would thus be considered to be predicted base paired. However, the relatively slow reaction rate of NMIA could explain this apparent disparity between the SHAPE data and the structure (Merino et al., 2005). If the bulges in SLI are transiently, and rapidly, switching between two confirmations, i.e. paired and unpaired, NMIA may not have enough time to properly react with the unpaired confirmation. These regions would therefore exhibit low SHAPE reactivity in this experimental set up. The high reactivity of the SLI apical loop suggests that any conformational changes occurring within SLI do not impact its overall stability.

This study has also confirmed the existence of SLIA, a small stem loop which resides in between SLI and SLII, 5' of the miR122 seed site ($_{81}C-G_{94}$). The intervening region between SLI and SLIA ($_{76}G-U_{80}$) is unpaired and the SHAPE reactivity profile agrees well with this. There is a short stem of 4 bases followed by a 6 nucleotide apical loop ($_{85}U-A_{90}$) with the following four nucleotides closing the stem. The nucleotides in the apical loop show high to intermediate reactivity, with the base paired nucleotides in the stem exhibiting low reactivity, indicating that this structure is stable within the 5'UTR. The single stranded region between SLIA and SLII, which contains the miR122 seed site ($_{96}C-C_{101}$), however appears to have a less straightforward relationship between its existence as a single stranded region in the structure, its SHAPE reactivity profile, and its role in translation. The miR122 seed site was demonstrated to interact with miR122 to enhance translation, and therefore is available for binding and would be predicted to be unpaired. However, whilst the structure shown in Figure 4.3B shows the miR122 seed site as single stranded, the SHAPE reactivity profile for this region indicates that $_{96}CAC_{98}$ are NMIA reactive, and therefore flexible and predicted unpaired, but that $_{99}UCC_{101}$ are unreactive and

thus demonstrate properties of base paired nucleotides with regards to their SHAPE reactivities. However, given the structures that exist either side of the miR122 seed site (SLIA and SLII) and that the unpaired bases available to interact with the miR122 site would not be able to pair with this sequence to extend these SLs, and the functional data demonstrating that the miR122 seed site is available to interact with miR122, it is highly likely that this site is unpaired.

SLII spans 79 nucleotides, ₁₀₁C-G₁₇₉. It contains two bulges, one large and one small, a mismatched loop on the right hand side of the stem, and a 5 nucleotide apical loop (₁₃₈G-A₁₄₂). Like SLI the bulge, and flexible, regions of the stem have low NMIA reactivity. However, due to the high reactivity demonstrated in the apical loop, this SL is also likely to be highly stable. There is a possibility that these single stranded regions do not exhibit reactivity toward NMIA due to RNA breathing as discussed for SLI, and similarly that any conformational changes occurring within SLII do not impact its stability.

3.3.4 The structure of EqHV SLIII

SLIII is the largest of the structural features of the EqHV 5'UTR spanning 164 nucleotides ₁₉₄C-G₃₃₂, and contains 6 sub domains that make up its overall structure (SLIIIa, b, c, d, e, and f). The paired region ₁₈₅CCGGCAGGU_{193/357}ACCUGCCGG₃₆₅ (Figure 4.3E) was not considered to be a part of SLIII as these bases are predicted to play a role in the formation of the pseudoknot and are, as such, classified within this structural feature.

SLIIIa (₂₁₅U-A₂₃₀) is here demonstrated to consist of a stem of 6 base pairs with an apical loop of 4 bases (₂₂₁A-A₂₂₄), the stem has a low reactivity to NMIA, indicative of a base paired region, with the apical loop exhibiting high reactivity. SLIIIb (₂₃₂A-U₂₇₁) is the largest sub domain of SLIII spanning 41 nucleotides. SLIIIb contains one large bulge towards the base of the stem, and an apical loop (₂₄₉A-U₂₅₃) consisting of 6 nucleotides which exhibit high reactivity to NMIA. SLIIIC (₂₇₄G-C₂₈₃) sits opposite SLIIIa and consists of a three base pair stem and a 4 nucleotide apical loop (₂₇₆C-G₂₈₀) which also demonstrates high NMIA reactivity similar to SLIIIa and SLIIIb.

SLIIId (₂₉₇G-C₃₂₂) has a 10 base pair stem with an apical loop consisting of 6 nucleotides (₃₀₇U-C₃₁₂), crucially this apical loop contains the highly conserved GGG motif. This study demonstrates that this motif is indeed presented in the apical loop of SLIIId, as was predicted, and that it is highly reactive to NMIA. These data provide further evidence towards a conserved mechanism of translation initiation between EqHV and HCV as deletion of this domain causes a complete

ablation in translation. It was hypothesised that this effect was due to the removal of the GGG motif which was likely interacting with the 40S ribosomal subunit to position it at the AUG start. Therefore the demonstration that this motif is presented in the apical loop, and is single stranded, means that it is available to interact with the 40S subunit and may represent a conserved mechanism of hepaciviral translation initiation.

SLIIIe (₃₃₃G-U₃₄₄) consists of a three base pair stem with a 6 nucleotide apical loop (₃₃₆U-G₃₄₁). Reactivities throughout the stem and loop agree well with the structure as shown. Finally SLIII f (₃₄₅G-G₃₅₆) possesses the shortest stem of the structural domains of SLIII of just two base pairs. The apical loop (₃₄₇U-G₃₅₄) is 8 nucleotides in length and is largely NMIA unreactive, however this loop has been shown to be relatively NMIA unreactive in published SHAPE data for the HCV IRES (Angulo et al., 2016) and is predicted to interact with downstream sequences to form the pseudoknot, which is described in 3.3.5.

Whilst the large majority of the SHAPE reactivity profile across SLIII agrees well with the structure, especially across base paired regions and in the apical loops, bulges within SLIII, similarly to SLI and SLII, are largely unreactive or exhibited low to intermediate reactivity. As discussed for SLI and SLII the highly structured nature of the rest of SLIII, coupled with the high NMIA reactivity of all of the apical loops, indicate that SLIII is a stable structure which folds as is shown in Figure 3.6. The most likely cause of the discrepancies between structure and SHAPE data exhibited here is RNA breathing which does not allow NMIA enough time to fully react with the bases in the bulges, rendering them NMIA unreactive. However, the overall structure and reactivity of the apical loops would suggest that SLIII is highly stable and the potential internal rearrangements in structure do not affect the overall confirmation of SLIII.

3.3.5 The structure of the EqHV pseudoknot

The pseudoknot remains a predicted structure, and has not been experimentally confirmed in this study, however there is strong evidence for its existence from the SHAPE analysis, and confirmed structure, of the 5'UTR. The pseudoknot as predicted consists of a stem of 9 base pairs forming between ₁₈₅CCGGCAGGU₁₉₃/₃₅₇ACCUGCCGG₃₆₅, and 6 nucleotides from the apical loop of SLIII f ₃₄₈UGCGAG₃₅₃ interacting with ₃₆₇CUCGUA₃₇₂. The EqHV sequence encompassing the start of the SLIIIe apical loop to the end of the predicted pseudoknot (₃₃₆U-A₃₇₂) shares 84% sequence conservation with the HCV sequence from the equivalent region and 100% across ₃₄₈UGCGAG₃₅₃ and ₃₆₇CUCGUA₃₇₂; predicted to base pair as part of the EqHV pseudoknot and

known to form such interactions in the HCV pseudoknot. Whilst the stem only displays 55% sequence identity across the two viruses it is important to note that whilst they may not maintain a conserved sequence, each virus has a perfectly complementary 9 nucleotide sequence located in the same equivalent location within the 5'UTR. SHAPE reactivities across this region provide further evidence that a pseudoknot is being formed. ¹⁸⁵CCGGCAGGU_{193/357}ACCUGCCGG₃₆₅ has no to low NMIA reactivity across its entirety agreeing with base pairing at these sites.

SLIIIf demonstrates a low SHAPE reactivity profile throughout its apical loop, excluding ³⁵⁰C. However if the pseudoknot is forming as predicted ³⁴⁸UGCGAG₃₅₃ of the SLIIIf apical loop would be base paired with ³⁶⁷CUCGUA₃₇₂, therefore low SHAPE reactivity in this region is in good agreement with the existence of the pseudoknot. In this configuration ³⁴⁷U and ³⁵³G would be not be base paired and thus should exhibit high NMIA reactivity. However, low NMIA reactivity has previously been demonstrated for HCV, which is known to form a pseudoknot, and therefore low reactivity of these bases does not exclude the formation of such a structure. Based on the SHAPE data presented here, conservation between EqHV and HCV throughout the pseudoknot region, and that the other secondary structure features which have been experimentally confirmed and do not preclude the formation of the pseudoknot, it is hypothesised that this structure exists in the EqHV 5'UTR as shown in Figure 3.6.

The data obtained from the SHAPE analysis of the EqHV 5'UTR, combined with structure prediction using this data as a pseudo free energy constraint, and conserved structural, sequence, and functional, features between EqHV and HCV has resulted in the first experimentally confirmed secondary structure model of the EqHV 5'UTR (Figure 3.6). This structure is extremely likely to contain a pseudoknot at its 3' end, prior to the AUG start. This complements the functional study described in Chapter 3, supporting the hypotheses made regarding the process of EqHV IRES mediated translation, and provides further evidence to support the hypothesis of a conserved mechanism of hepaciviral translation.

3.3.6 EqHV 5'UTR Δ SLI and Δ SLI+II structural deletions have no overall impact on 5'UTR secondary structure

To confirm that the functional effects that structural deletions had on EqHV IRES mediated translation were specifically caused by deletion of a structure, and not more general structural effects that the deletion caused, it was necessary to confirm the structures of the WT and

mutant EqHV 5'UTRs in each construct. This also confirmed that differences observed in translation from the bicistronic and SGR constructs was not due to differential UTR folding in these different contexts. The structure of the 5'UTR in the SGR is extensively discussed in 3.3.2-3.3.5. Comparisons of mutant 5'UTR structures will be made to the 5'UTR structure in the SGR as this represents the most physiologically relevant context in which the structure of the 5'UTR was mapped. A model for each mutant was created based on the Mfold prediction (appendices Figures 8.1-8.8), and combined with the experimentally confirmed structure to create an updated prediction of mutant structures from those depicted in Figure 3.2. SHAPE reactivity values were then mapped onto these structures in order to assess whether they significantly differed from the WT across the relevant regions. Structural predictions using the experimental SHAPE values as pseudo free energy constraints were also generated, and agreed well with those depicted here; for ease of comparison, as no significant structural changes were predicted, the confirmed WT structure with the mutations manually introduced, and SHAPE values mapped on, are here shown.

The WT 5'UTR SHAPE reactivities for the bicistronic construct mapped well onto the structure determined from the SGR WT 5'UTR. No major differences between the two constructs were observed and the overall secondary structure is not altered when the 5'UTR was in the context of the bicistronic construct. Importantly, the miR122 seed site, SLIIIb, SLIIId, and the pseudoknot exhibited extremely similar SHAPE reactivity profiles. Some nucleotides exhibited increased reactivity in the bicistronic UTR, however there were no defined regions that exhibited major differences in reactivity, and thus no structural differences are expected. Importantly, when taking into consideration the functional data, the miR122 seed site exhibited very little change between the SGR and bicistronic constructs. Absolute reactivity was slightly lower in the bicistronic than in the SGR, however there was no change in classification of nucleotides throughout and nucleotides predicted to be paired/unpaired in the SGR showed the same trends in the bicistronic construct. These data demonstrate that the difference in miR122 responsiveness observed between these two constructs is due to the positioning of the 5'UTR, rather than different structural arrangements. This supports the hypothesis that the enhancement of translation upon expression of miR122 is due to Xrnl protection.

Δ SLI had no impact on IRES function in comparison to the WT 5'UTR and it was therefore not expected to cause any major structural changes outside of the deletion of SLI. This was

confirmed by SHAPE as all of the major structural features displayed reactivity profiles which resembled the WT 5'UTR and were suggestive of a maintenance of structure.

Δ SLI+II had an extremely variable effect on IRES efficiency, which ranged from mean efficiencies of ~23%-100% of WT. Two hypotheses were posed to explain this variation in function: Firstly that SLII may have an indirect role to play in translation and therefore the effect of its deletion had varying effects on translation. Secondly that Δ SLI+II caused the remaining structure to adopt two, or more, different conformations with different translation competency, with the ratio of these different conformations favoured under different conditions. However, the SHAPE reactivity profile was consistent with that of SLIII in WT, and across the two constructs. If different conformations of the 5'UTR were present large differences would have been expected to be seen between the SHAPE reactivity profiles of Δ SLI+II and the WT 5'UTR, with regions adopting different NMIA reactivity profiles between different conformations. As no such regions were present these data rule out the possibility that Δ SLI+II causes the 5'UTR to adopt a number of different conformations and supports the hypothesis of an accessory role of SLII in translation. It does not however account for why the variation occurs and why the Δ SLI+II 5'UTR exhibits such variability. A more detailed investigation into the role of SLII in translation, and the host cell factors with which it interacts, would be needed to identify differences between the cell lines that may account for the variable translation phenotypes observed.

3.3.7 The effect of SLI+II on EqHV 5'UTR IRES function

The role of structural features of the EqHV 5'UTR IRES were here investigated in two complimentary systems. In order to fully appreciate the effect on translation that each deletion had, and subsequently the role that structure plays in translation, these data must be looked at in combination.

Δ SLI, regardless of the cell line or construct in which it was tested, caused no significant impact on translation. Therefore it is possible to conclude that, in agreement with the work published by Stewart *et al* this SL, which is the only major structure of the EqHV 5'UTR not present in HCV, does not have a role to play in translation initiation (Shi and Lai, 2006). It has been demonstrated that the HCV SLI functions in replication and that domains II-IV of the HCV 5'UTR constitute the IRES (Friebe *et al.*, 2001; Lukavsky, 2009). Due to the structural similarities observed between EqHV and HCV, and the data presented here, and published by others (Stewart *et al.*, 2013; Scheel *et al.*, 2015), a conclusion naturally follows that SLI functions during EqHV replication in

similar way to that of HCV SLI. However, without a system to study replication of EqHV, this cannot yet be confirmed. Furthermore, it is possible that SLI of EqHV and HCV do not represent equivalent structures within the respective 5'UTRs. It may be the case that the EqHV SLIA represents the equivalent structure to HCV SLI and, if function is conserved, it is SLIA which plays a role in EqHV replication with SLI carrying out another, unknown, function. Whilst this hypothesis does not rule out a role for EqHV SLI in replication, it is something that would need to be considered in future experiments.

Whilst the data for Δ SLI was in agreement across all five cell lines, and both constructs, the data for Δ SLI+II was less straightforward. Translation efficiency varied greatly between cell lines and constructs alike. Δ SLI+II exhibited 61% mean activity of WT in Huh7 cells transfected with the bicistronic construct compared with 25% mean activity in Huh7 cells electroporated with the SGR, a similar trend was seen in the 293T cells where Δ SLI+II caused no significant effect on translation in the bicistronic construct but a 51% reduction in activity in the SGR. The converse was observed in FHK cells as mean IRES activity rose from 23% of WT in the bicistronic construct to 51% in the SGR in this cell line. There was no consistent effect on translation observed between each method of delivery, i.e. transfection or electroporation, and therefore this can be ruled out as the cause of the observed differences in IRES activity. The degree of variability between constructs and cell lines for this mutant means that definitive conclusions regarding its role in translation are difficult to draw. There was no consistent trend to be observed between either the construct in which the UTR was assayed, or which cell line was assayed. As such, different cell host factors or construct specific effects do not appear to be the cause of the variation. However, one definitive conclusion that can be drawn from this data is that SLI+II do not play an essential role in translation as, whilst their deletion resulted in mean translation efficiency ranging from 23-100% of WT, a complete ablation of translation was not observed under any conditions. When this data is combined with our, and others, findings that SLI is not involved in EqHV IRES mediated translation it may be the case that SLII, whilst not essential, serves an accessory function which enhances translation from the EqHV IRES. The data presented here is in agreement with previous studies demonstrating that HCV SLII may not be required for IRES function (Wang et al., 1994; Reynolds et al., 1996; Kamoshita et al., 1997; Kolupaeva et al., 2000). However, the inhibitory effect on IRES activity in the absence of SLII may suggest that a conserved function may exist between HCV and EqHV in contacting the ribosome.

The role that miR122 may be exhibiting in translation from the Δ SLI+II 5'UTR is discussed in 3.3.10.

3.3.8 Structural deletions in EqHV 5'UTR SLIII have no overall impact on 5'UTR secondary structure

Δ SLIII caused complete ablation of translation in all cell lines tested and there are two potential explanations for this phenotype: Firstly that SLIII is necessary for translation and its deletion therefore ablates this function. Secondly that the deletion of this large SL caused a structural rearrangement that abolished translation, especially given the close spatial relationship between SLIII and the pseudoknot. However, when mapped onto the Δ SLIII structure the data supports the hypothesis that SLIII is necessary for IRES function. SLI+II are clearly retained with the stems, and apical loops, showing reactivity profiles in good agreement with the WT. The flexible region in SLII did become highly reactive upon the deletion of SLIII, compared to unreactive in the WT. If the hypothesis regarding breathing and formation of unstable structural intermediates which are not conducive to NMIA reactivity is true, SLIII could be mediating such events and as such its deletion stabilises the confirmation as shown in Figure 3.7. However, as SLII is not necessary for IRES function, increased flexibility in this region is highly unlikely to cause complete ablation of translation. Interestingly these data indicate a role for SLIII effecting the flexibility of SLII, suggesting potential effects on tertiary structure caused by SLIII.

Due to the proximity of the structures it was most likely that, if they were to exist, aberrant effects of Δ SLIII on the 5'UTR structure would affect the pseudoknot. It is known that the pseudoknot plays an extremely important role in HCV translation, positioning the AUG start in the mRNA binding cleft (Berry et al., 2010). Great care was taken in the design of this mutant to preserve SLIIIe and SLIII f, and the initial stem of the pseudoknot, to preserve correct folding. However, it could not be ruled out that such a large deletion would have aberrant effects on the formation of the pseudoknot and thus be responsible for the ablation of translation. The SHAPE data in this region of Δ SLIII was in good agreement with that which was seen for the WT, and the slight observed differences do not appear large enough to elicit any major effect on the structure of the pseudoknot. Interestingly the miR122 seed site exhibits higher SHAPE reactivity throughout the sequence for the Δ SLIII mutant compared to WT, indicating that this region may be more flexible than in the WT 5'UTR. This could be being mediated by the increased flexibility in SLII. However a functional attribute cannot be assigned to this change in reactivity as translation does not occur from this mutant.

Δ SLIIIb exhibited variable effects between the two constructs tested, and it appeared as if SLIIIb was not essential to IRES function in the bicistronic construct, but necessary in the SGR. These different translation phenotypes could have been attributed either to different folding of this mutant in the two constructs resulting in different phenotypes, or to differential requirements for structural features in the bicistronic and SGR constructs; whereby the function of SLIIIb was not completely necessary in the bicistronic construct due to 40S ribosomal subunits being released from the end of cistron one and accumulating at the Δ SLIIIb IRES. This spatial arrangement could substitute for the proposed function of SLIIIb in binding eIF3 to prevent 43S accumulation and promote free 40S, assuming conserved function between EqHV and HCV. If the latter were the case then this would indicate that SLIIIb is indeed necessary for IRES function and that the ability of the Δ SLIIIb mutant to initiate translation was merely an artefact of the experimental set up. Figure 3.15 and Figure 3.16 demonstrate that SHAPE reactivities between the two constructs for the Δ SLIIIb mutant are largely similar. SLIIIa and SLIIIc are retained and SLIIIb is replaced by a single stranded “linker” region which is highly reactive to NMIA. The overall structure of SLIII is retained and the deletion of SLIIIb does not appear to have any significant impact on the pseudoknot. Some differences in reactivities were observed in SLII in the SGR; however, this study has already demonstrated that SLII is not necessary for IRES function and so slight effects seen here would not be expected to ablate translation.

Finally Δ SLIIId caused complete ablation of translation in both the bicistronic and SGR constructs. It was hypothesised that this was due to the removal of the GGG motif which is proposed to interact with the 40S ribosomal subunit to position it at the AUG start codon for translation. However, it was possible that this ablation of translation was merely due to improper folding of the 5'UTR mediated by this deletion. The GGG motif in the apical loop of SLIIId has been previously demonstrated to impact correct folding of the HCV 5'UTR (Jubin et al., 2000). If the former of these two hypothesis was true it would provide further evidence that there exists a conserved mechanism of hepaciviral translation initiation. Determining the structure of this mutant 5'UTR was therefore essential. Upon inspection of the reactivity profile of Δ SLIIId it can be seen that it is in good agreement with the WT 5'UTR across all of the main structural elements investigated in this chapter. The overall structure of SLIII, and the pseudoknot, appear to be maintained, as do SLI, SLII, and SLIA. There are some slight differences to be observed in SLI and SLII. As for previous mutants these slight differences in reactivity profile do not appear to cause any overall structural changes to the UTR. Given their location in the 5' half of the UTR, which is

not necessary for IRES function, they are highly unlikely to be able to completely ablate translation. As expected the deletion of SLIIId, similarly to Δ SLIIb, resulted in a single stranded, highly NMIA reactive, “linker” region where SLIIId used to reside. This region already possessed some flexibility due to the single stranded region directly downstream of SLIIId, and thus it would not be expected that a minor increase in flexibility here would be able to ablate translation. The retention of all of the other major structural features, especially with regards to the pseudoknot and SLIII, indicates that no rearrangements of structure have occurred and confirms a specific block to translation mediated by the deletion of SLIIId. This finding provides further evidence that SLIIId, specifically the GGG motif, is interacting with the 40S ribosomal subunit to initiate translation; a function conserved between EqHV and HCV.

3.3.9 The effect of SLIII on EqHV 5'UTR IRES function

Δ SLIII consistently ablated translation to mock/control levels across all cells lines irrespective of the construct in which it was assayed. These data show that SLIII plays a direct and essential role in translation from the EqHV IRES, as it does in HCV. Whilst no sequence or structural deletions were made to SLIIIe or f, when the Δ SLIII data is viewed in combination with that obtained for Δ SLI+II it would appear as if SLIII and the pseudoknot comprise the minimal EqHV IRES.

Translation efficiency from Δ SLIIb exhibited variation between Huh7, FHK, and 293T cells for the bicistronic construct with regards to the efficiency of translation; however, in the context of the SGR, translation was completely ablated in all three cell lines. For the bicistronic construct mean translation efficiencies were 54% and 56% of WT in Huh7 and FHK cells respectively and, although not statistically significant, 80% of WT in 293T cells. Whilst the degree of efficiency was variable between these three cell lines it is evident that the deletion of SLIIb did not completely ablate translation in the context of the bicistronic construct. This is in stark contrast to the complete lack of translation exhibited from this mutant in the same cell lines when it was in the context of the SGR. This poses many questions about the role of this structure in EqHV translation as data from the bicistronic assays would suggest that it is non-essential whereas the SGR data indicates that it is. The disagreement between the two sets of data may be due to experimental differences in regards to the two constructs. The proximity of the UTR to the cap-dependent first cistron could result in the accumulation of initiation factors, and ribosomal

subunits, which would be able to interact with the remaining structural elements of the 5'UTR that would not have been able to, or which would have occurred far less frequently, in the SGR.

During translation from the HCV IRES SLIIIb has been shown to recruit eIF3 by interacting with its ribosome binding face, thus preventing the accumulation of 43s complexes and promoting the availability of 40S subunits (Sun et al., 2013; Hashem et al., 2013). If this same mechanism were to be taking place in EqHV then proximity of the 5'UTR to a site of active translation i.e. cistron one, could be acting to carry out this function in the absence of SLIIIb. In this scenario, following the detection of the stop codon of cistron one, recruitment of the release factor, and dissociation of the two ribosomal subunits from the mRNA, there would be an accumulation of 40S ribosomal subunits at the end of cistron one and thus at the Δ SLIIIb UTR. This could therefore avoid the requirement of eIF3 interactions with SLIIIb, and allow translation to proceed from this mutant IRES. However when the 5'UTR was located in its more natural position, at the proximal 5' end of the SGR, SLIIIb may be absolutely required to either free 40S from eIF3, or recruit eIF3 and the 40S subunit to the 5'UTR for translation initiation and therefore, in its absence, translation would be ablated.

Finally translation from Δ SLIIId (SLIIId contains a conserved GGG motif in its apical loop) was completely ablated in all cell lines tested in both the bicistronic, and SGR constructs (Figure 3.19 and Figure 3.22). The conserved GGG motif could be the key factor in understanding the clearly essential nature of SLIIId in translation from the EqHV 5'UTR IRES, and our understanding of translation initiation in HCV allows us to draw such conclusions. The GGG motif in the apical loop is completely conserved amongst HCV isolates and is conserved in the related *Hepaci-* and *Pestiviruses* (Jubin et al., 2000). In HCV this motif has been shown to directly interact with the 40S subunit, positioning it at the AUG start during translation initiation (Matsuda and Mauro, 2014; Malygin et al., 2013; Angulo et al., 2016). Due to the structural similarities observed between the 5'UTR of these two closely related viruses (Figure 3.1), the observed similarities in relation to the function of 5'UTR structural domains, and the presence of the conserved GGG motif both in sequence and position in the UTR, it is not unreasonable to assume a conserved mechanism of translation between EqHV and HCV. If this is the case then the Δ SLIIId deletion would remove the site of interaction between the EqHV 5'UTR and the 40S subunit, therefore preventing ribosome recruitment to the AUG start and preventing the formation of a translation complex.

Taken together these data show that SLI is not involved in translation initiation from the EqHV 5'UTR IRES and, whilst not necessary for IRES function SLII may play a role in efficient translation initiation. Furthermore these data suggest that SLIII and the pseudoknot comprise the necessary minimal EqHV IRES. There is likely a conserved mechanism of translation initiation between HCV and EqHV.

3.3.10 The role of miR122 in translation from the EqHV 5'UTR IRES – a tale of two constructs

miR122 is a liver specific microRNA which is essential to HCV replication, however it has also been demonstrated to play a role in translation and RNA stability (Jangra et al., 2010; Thibault et al., 2015; Jopling et al., 2005). Whilst HCV contains two miR122 seed sites within its 5'UTR, EqHV only contains one, located directly 5' of SLII. This corresponds to seed site two in HCV (Figure 3.1). It has previously been demonstrated that the overexpression of miR122, in cell lines in which it is not endogenously expressed, upregulates translation from the EqHV 5'UTR IRES which can be knocked down using LNAs (Scheel et al., 2015). However, whilst this effectively demonstrated the effect of miR122 on EqHV translation, the site of interaction with the UTR was merely assumed. Therefore this study aimed to firstly confirm the enhancing effect of miR122 on EqHV translation, and secondly demonstrate the site of interaction.

Upon transfection of bicistronic constructs into cells engineered to overexpress miR122 no obvious effect could be observed (Figure 3.21). In order to confirm whether or not miR122 was exhibiting an effect on translation the WT 5'UTR translation data from said miR122 expressing cells was normalised to the corresponding data from the parental cell line (Figure 3.21). This showed that no significant difference in translation efficiency could be seen upon the expression of miR122 for either FHK or 293T cells. One hypothesis for this lack of responsiveness is that translation from the EqHV 5'UTR IRES is simply not influenced by the presence of miR122; however this explanation seemed extremely unlikely given published data, and the array of functional similarities observed here between HCV and EqHV. It was therefore much more likely that the miR122 is unable to carry out the function which enhances EqHV translation when the 5'UTR is assayed in the bicistronic construct.

To answer this question the same miR122 expressing cell lines were assayed using the SGR. When a comparison of the WT values was conducted between the cell lines, as described above, for the SGR it became clear that the presence of miR122 significantly enhanced translation (Figure 3.23). This indicated that the effect of miR122 was dependent on the construct in which

it was assayed. When Scheel et al demonstrated miR122 mediated enhancement of EqHV translation they did so in a monocistronic reporter, in which the 5'UTR is located at the proximal 5' end. It was therefore likely that the effect of miR122 on translation is dependent on the positioning of the 5'UTR. It is therefore hypothesised that miR122 may be exhibiting a stabilising effect on the EqHV 5'UTR, potentially protecting it from the Xrn1 endonuclease, as has been demonstrated for HCV (Thibault et al., 2015).

It was also observed that the efficiency of Δ SLI+II appeared to decrease in the miR122 expressing cells in comparison to the WT, from ~50% mean translation efficiency in FHK and 293T cells to ~20% in FHK 122 and 293T 122 cells (Figure 3.22 and Figure 3.23). It was possible that the presence of miR122 was actively inhibiting translation. However this deletion mutant was the only one that exhibited such an effect, and was also the only one in which the deletion removed the miR122 seed site. This therefore meant that the most likely explanation was that the observed reduction in efficiency was due to the lack of a miR122 seed site. This would result in translation from the WT 5'UTR being enhanced by miR122 with Δ SLI+II not, resulting in a perceived reduction in translation efficiency compared with WT. When the Δ SLI+II values from miR122 expressing cells are compared to the WT values from the respective parental cell lines they demonstrate a mean efficiency of 46% and 43% in FHK and 293T cells, respectively, compared with 52% and 53% for Δ SLI+II values from the parental cell lines. However, the differences between the sets of data obtained from the parental and miR122 expressing cells is not statistically significant for either cell line. This confirms the hypothesis that Δ SLI+II was simply unresponsive to the presence of miR122 rather than actively downregulated.

To confirm that these observations were caused by the lack of the miR122 seed site, rather than another, unknown, feature which resided within SLI+II a miR122 to miR124 seed site mutant was created and assayed in the FHK and 293T parental and miR122 expressing cells (Figure 3.24). No significant difference was observed in translation efficiency from SGR NZCI 124 in the presence of miR122 in either cell line. This confirmed that enhancement of translation from the EqHV 5'UTR IRES is mediated by an interaction between miR122 and the miR122 seed site. This can be ablated by complete removal, or mutation, of this seed site. However whether the activity of miR122 is dependent on the location of this seed site was not investigated in this study and further work would be required to demonstrate this.

Taken together these data show that miR122 mediated enhancement of EqHV 5'UTR IRES activity is dependent on two factors, the positioning of the 5'UTR at the proximal end of the construct from which translation is being measured, and the presence of a miR122 seed site, which is naturally located directly 5' of SLII.

This structural study has provided the first experimental description of the secondary structure of the EqHV 5'UTR and has confirmed the existence of three main structural domains SLI, SLII, and SLIII(a ,b ,c ,d ,e , and f). The data in this study also indicate that a pseudoknot is highly likely to form at the 3' end of the EqHV 5'UTR. Furthermore it has been demonstrated that structural deletions of major 5'UTR structures do not affect overall folding and structure, and confirms that translation phenotypes observed for these mutants are specific to the function of the deleted structural domain. Therefore, when taken together with the functional analysis of the EqHV 5'UTR it is possible to conclude that SLI+II are not necessary for translation, and that SLIII and SLIIId are absolutely necessary. The necessity of SLIIIb has not yet been confirmed, however the hypothesis that SLIIIb is only not necessary in the bicistronic construct due to the internal location of the IRES, and its proximity to the end of cistron one, provides a reasonable explanation for this. Further investigation to confirm this hypothesis is required.

Chapter 4 EqHV 5'UTR interactions with the host cell translational machinery

4.1 Introduction

The data presented in Chapter 3 demonstrated that SLIII is the main determinant of translation from the EqHV 5'UTR IRES. Furthermore, these data demonstrated an essential role for SLIIIb and SLIIId in translation. A further analysis of the role which sequence plays in the function of SLIII was therefore carried out to investigate the ability of EqHV SLIII to interact with the host cell translational machinery.

It is known that eIF3 and the 40S ribosomal subunit specifically interact with HCV SLIIIb and SLIIId, respectively. The ribosomal binding face of eIF3 interacts with the apical loop, and a mismatched loop of SLIIIb (Sun et al., 2013; Hashem et al., 2013). A more specific determination of interaction has been demonstrated between the 40S ribosome and HCV SLIIId. The conserved GGG motif in SLIIId specifically interacts with ¹¹¹⁶CCC₁₁₁₈ of the 40S subunit (Matsuda and Mauro, 2014; Malygin et al., 2013). Mutations made within the apical loops of HCV SLIIIb and SLIIId have been demonstrated to significantly impair translation activity from the HCV IRES. It was hypothesised that the apical loops of EqHV SLIIIb and SLIIId also interact with eIF3 and the 40S ribosomal subunit and that these interactions could be interrupted by apical loop mutations.

Section 4.2.1 describes EqHV SLIIIb and SLIIId apical loop mutants and the effects they have on translation from the EqHV IRES.

SHAPE footprinting is a method of SHAPE analysis (described in 4.2.1) in which a predicted interacting partner is introduced into the reaction, following folding and before the addition of the SHAPE reagent. This method has been used to predict eIF3 interactions with HCV SLIIIb (Sun et al., 2013) and HCV SLIIId interactions with the 40S ribosome (Angulo et al., 2016). Furthermore, analysis of RNA-protein interactions requires minimal additional optimisation over that which is required for the conventional SHAPE experiments. Therefore if a SHAPE experimental protocol has been previously optimised for a specific RNA, as was the case in this study, SHAPE footprinting provides a relatively straightforward experimental progression to provide a more detailed insight into the characteristics of the RNA in question.

Section 4.2.2 of this Chapter describes SHAPE footprinting experiments carried out on the WT EqHV 5'UTR, and the apical loop mutants, to assess their ability to interact with purified eIF3 and 40S ribosomal subunit.

Whilst SHAPE footprinting provides good evidence for a protein-RNA interaction, this approach utilises purified proteins for the analysis and is therefore not suitable to identify a large number of interacting partners. Furthermore this technique infers interaction by assessing the ability of a protein to protect unpaired nucleotides from reaction with the SHAPE reagent. This means that proteins that associate with a protein in a complex, and therefore are not involved in a direct interaction with the RNA, would not be able to be identified. To identify which components of the translational machinery are associated with the EqHV 5'UTR, tandem mass tag mass spectrometry (TMT MS) was utilised (Thompson et al., 2003)

Section 4.2.3 describes pull-down assays performed using biotinylated EqHV 5'UTR as bait, and the interacting partners identified by TMT MS.

4.2 Results

4.2.1 Sequence specific mutations in EqHV SLIIIb and SLIIId disrupt translation

Functional analysis had demonstrated that SLIIIb, and SLIIId, play key roles in translation from the EqHV IRES. It was hypothesised that this was due to interactions with eIF3 and the 40S ribosomal subunit, respectively. Apical loop mutations of SLIIIb and SLIIId in HCV have previously been demonstrated to reduce translation efficiency to 50%, and <5% respectively, and these loops have previously been demonstrated to interact with eIF3 and the 40S subunit. (Jubin et al., 2000; Barría et al., 2009; Kieft et al., 2001; Sun et al., 2013; Angulo et al., 2016). To investigate whether the apical loops of EqHV SLIIIb and SLIIId were mediating interactions with host cell factors during translation initiation, causing inhibition of translation upon their deletion, sequence specific mutations were introduced into SLIIIb (₂₅₀CUU₂₅₂ to ₂₅₀GUC₂₅₂) and SLIIId (₃₀₉GGG₃₁₁ to ₃₀₉AGU₃₁₁). ₃₀₉GGG₃₁₁ to ₃₀₉AGU₃₁₁ is a mutation that has been previously characterised for HCV and ablates both translation and 40S ribosomal subunit interaction. Mutations were introduced into the SGR by Q5 SDM (NEB). As described in 3.2.10 RNA was *in vitro* transcribed and transfected into Huh7 (Figure 4.1A), FHK (Figure 4.1B) and 293T (Figure 4.1C) cells, and assayed for FLuc activity 6 h.p.t. All experiments were conducted to N=3 and a student's T test calculated for all data in comparison to WT translation. Data are represented as a percentage of the WT IRES activity, which is set at 100%. The GUC mutant caused a significant impairment of translation to 40% of WT in the Huh7 cells, 76% in FHK cells, and 64% in 293T cells. Unlike the Δ SLIIIb mutation, GUC did not cause a complete ablation of translation in any of the cell lines tested, indicating that the overall structure of SLIIIb is more important to EqHV translation than the sequence of the apical loop. The AGU mutant, on the other hand, caused an almost complete ablation of translation in all three cell lines tested to a mean efficiency of 13% of WT in Huh7, 17% in FHK, and 10% in 293T cells. Indicating that the sequence of the apical loop of SLIIId is a key determinant of EqHV IRES activity.

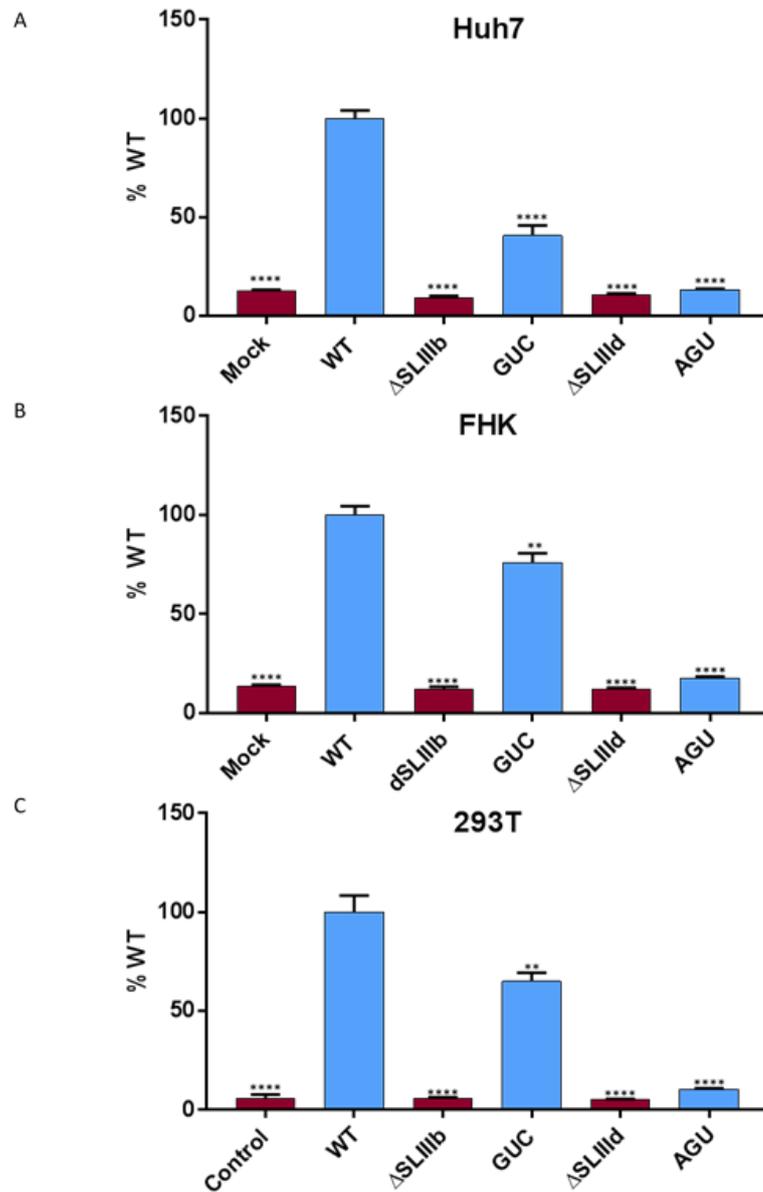


Figure 4.1: Translation activity of SLIIIb and SLIIId apical loop mutations

NZCI N (where N represents either WT or mutant EqHV) or mock (no RNA) was transfected into A) Huh7 B) FHK and C) 293T cells, harvested 6 h.p.t and assayed for firefly luciferase activity. All values are represented as a percentage of the RLU for the WT EqHV 5'UTR and this is used as a representative measure of IRES efficiency (with WT taken as 100% activity). All data was subject to a student's T test in comparison to the WT value for each respective cell line. Stars represent statistically significant values.

4.2.2 SHAPE footprinting analysis

The functional data in Figure 4.1 suggested that the GUC and AGU mutations could be disrupting SLIIIb and SLIIIId interactions with the host cell translational machinery. Based on our knowledge of the HCV IRES it was hypothesised that the interacting partners were eIF3 and the 40S ribosomal subunit, respectively. To test this hypothesis SHAPE footprinting analysis of SLIII was conducted with either purified eIF3 or 40S ribosomal subunit (provided by N. Locker, University of Surrey). In a conventional SHAPE experiment the reactivity of any given base is solely dependent upon the flexibility of its 2'OH group. However, the addition of purified protein, before the addition of NMIA, alters this. If the protein and RNA interact, the reactivity of each base toward NMIA becomes dependent not only upon the flexibility of its 2'OH, but also upon protection by the interacting protein. Therefore if a previously NMIA reactive base interacts with the protein it will preclude NMIA binding and shift from a reactive to unreactive base. A schematic representation of this altered SHAPE protocol can be seen in Figure 4.2.

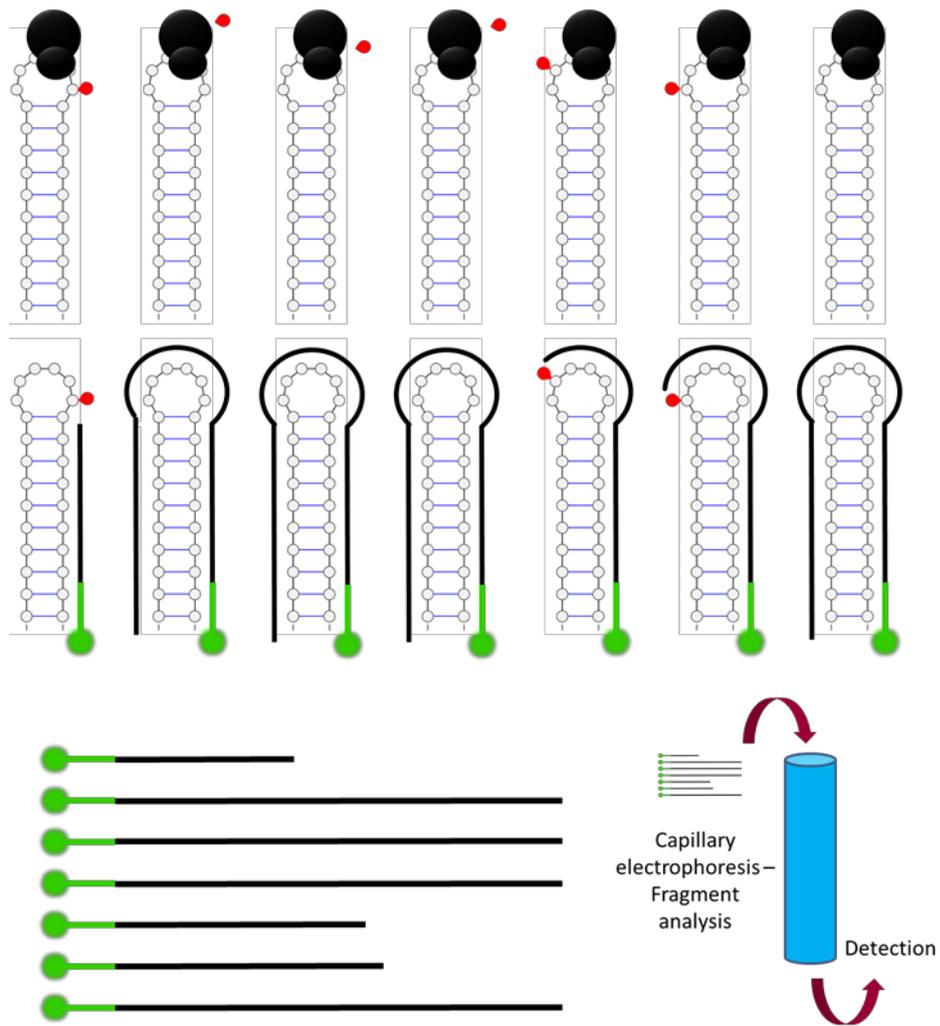


Figure 4.2: A schematic representation of the SHAPE experimental procedure

A schematic representation of the SHAPE footprinting experimental flow through with the addition of protein following RNA folding and before addition of NMIA. NMIA is represented in red, with fluorescently labelled primers shown in green. RT extensions are show as black lines.

4.2.3 The EqHV 5'UTR interacts with eIF3 and the 40S ribosomal subunit

Using SHAPE footprinting the ability of SLIIIb and SLIIId to interact with eIF3 and the 40S ribosomal subunit was investigated. EqHV SGR containing the WT 5'UTR was *in vitro* transcribed and subject to SHAPE footprinting analysis with 300 nM of purified eIF3 (Figure 4.3), or 40S ribosomal subunit (Figure 4.4). All experiments were conducted to N=3 and a student's T test, in comparison to no protein, was conducted for each base in SLIIIb (for eIF3 footprinting) or SLIIId (for 40S footprinting) to determine if any statistically significant change in NMIA reactivity had occurred.

The apical loop of SLIIIb (₆₅ACUUU₆₉) is highly NMIA reactive when analysed in a traditional SHAPE experiment (Figures 5.3A and B). However, upon the addition of eIF3 NMIA reactivity is significantly reduced across all 5 bases in the apical loop, in comparison to no eIF3 (Figure 4.3C and D and Table 4.1). No other statistically significant changes in NMIA reactivity are observed in SLIIIb in the footprinting assay. These data indicate that eIF3 is able to specifically interact with the apical loop of EqHV SLIIIb.

Similarly the apical loop of SLIIId (₁₂₃UUGGGC₁₂₈) is highly NMIA reactive in a traditional SHAPE experiment (Figure 4.4A and B and Table 4.2). However, upon the addition of 40S ribosomal subunit NMIA reactivity is significantly reduced across all 6 bases in the apical loop, in comparison to no 40S (Figure 4.4C and D and Table 4.2). These data indicate that the 40S ribosomal subunit is able to specifically interact with the apical loop of EqHV SLIIId. Interestingly, there is also a statistically significant increase at ₁₂₉C, the first base 3' of the apical loop. This is similar to the non-significant increase ($p=0.13$) in NMIA reactivity exhibited at ₇₁A (increase in mean NMIA reactivity of 0.2 to 2.6) of SLIIIb (Figure 4.3D). Both increases in reactivity occur at bases immediately downstream of the site of interaction and could therefore be caused by residual bound protein on the apical loops causing extension of the RT to terminate and thus giving a false positive NMIA reactivity.

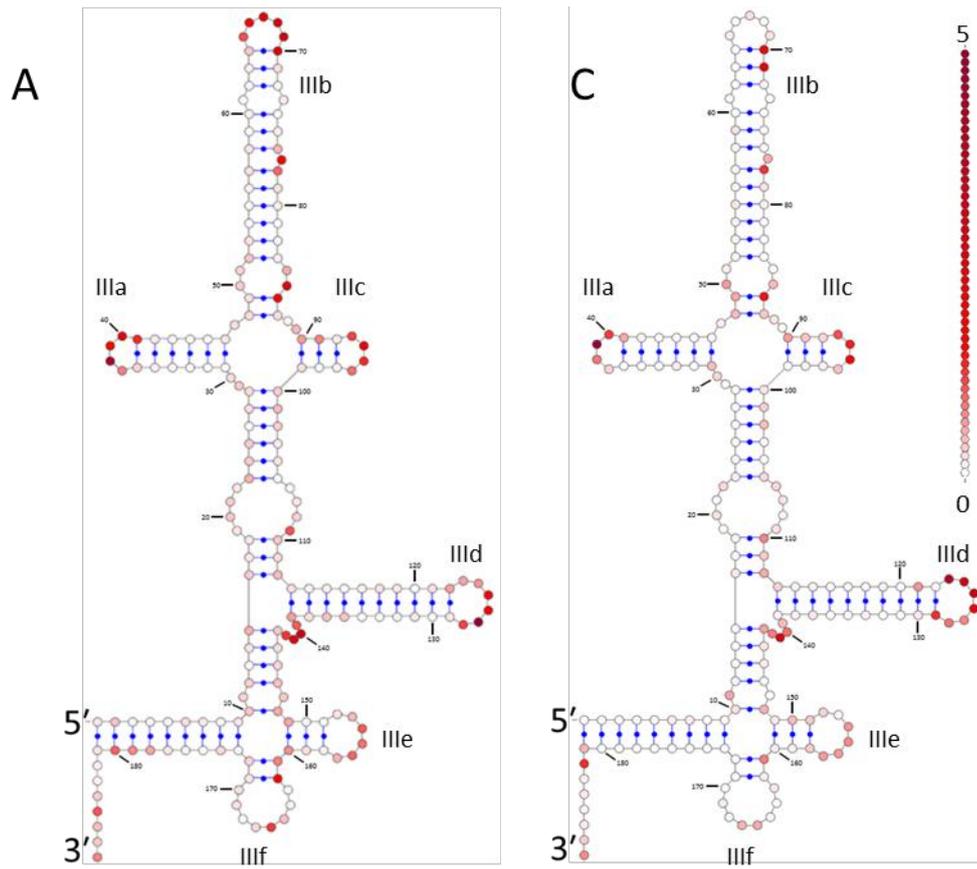


Figure 4.3: Shape footprinting reveals an interaction between the apical loop of EqHV SLIIIb and eIF3

A) SHAPE analysis of SLIII of the EqHV 5'UTR B) Enlarged image of SLIIIb from the SHAPE analysis in A. B) eIF3 SHAPE footprinting analysis of SLIII of the EqHV 5'UTR C) Enlarged image of SLIIIb from the footprinting analysis in C. Red stars indicate bases which exhibited a statistically significant reduction in NMIA reactivity when footprinting with eIF3 was conducted, compared to no eIF3. SHAPE reactivities are represented on a colour scale from white (low reactivity – predicted paired) to red (high reactivity – predicted unpaired) a representative scale denoting reactivity increments of 0.1 is displayed. Number labelling corresponds to domain III only and number labelling in B and D correspond to number labelling in A and C.

Nucleotide	WT no eIF3	WT +eIF3	Difference	p value
65 A	1.033	0.173	(0.860)	0.020
66 C	1.563	0.050	(1.513)	0.005
67 U	2.027	0.257	(1.770)	0.027
68 U	2.227	0.070	(2.157)	0.009
69 U	2.693	0.223	(2.470)	0.015

Table 4.1: SHAPE reactivity values for the apical loop of WT SLIIIb +/- eIF3

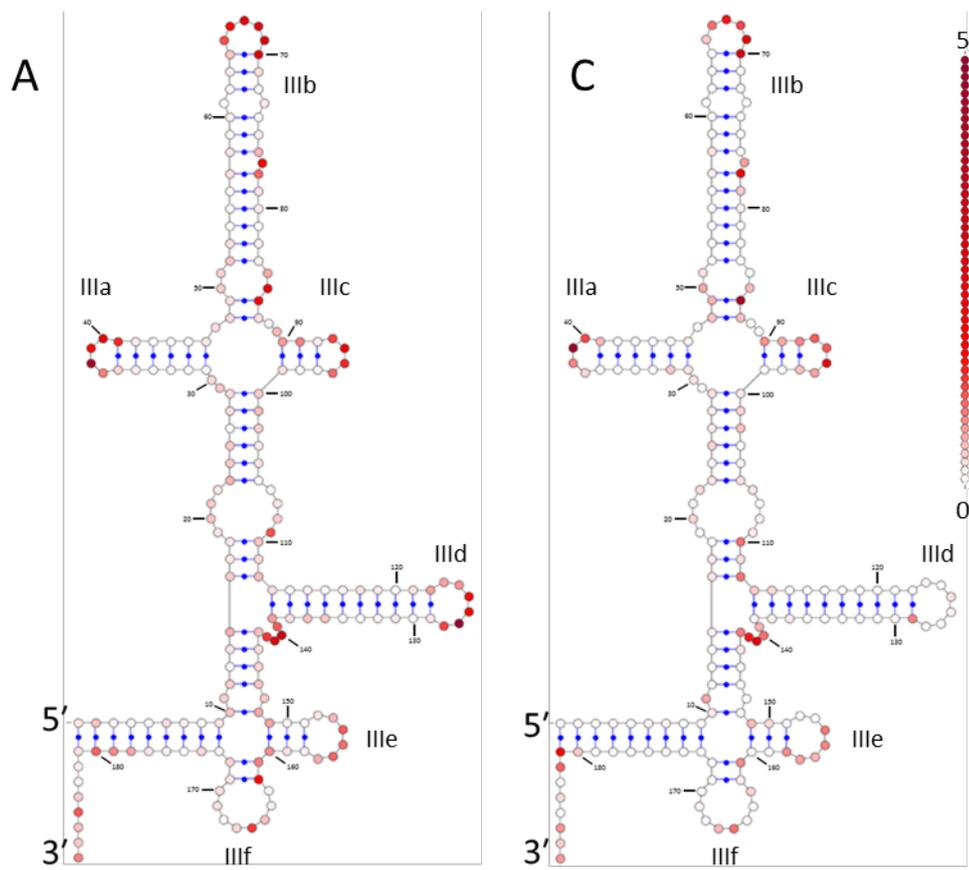


Figure 4.4: Shape footprinting reveals an interaction between the apical loop of EqHV SLIIId and the 40S ribosomal subunit

A) SHAPE analysis of SLIII of the EqHV 5'UTR B) Enlarged image of SLIIId from the SHAPE analysis in A. C) 40S ribosomal subunit SHAPE footprinting analysis of SLIII of the EqHV 5'UTR D) Enlarged image of SLIIId from the footprinting analysis in C. Red stars indicate bases which exhibited a statistically significant reduction in NMIA reactivity when footprinting with 40S was conducted, compared to no 40S; green stars represent an increase. SHAPE reactivities are represented on a colour scale from white (low reactivity – predicted paired) to red (high reactivity – predicted unpaired) a representative scale denoting reactivity increments of 0.1 is displayed. Number labelling corresponds to domain III only and number labelling in B and D correspond to number labelling in A and C.

Nucleotide	WT no 40S	WT +40S	Difference	p value
122 G	0.553	0.000	(0.553)	0.183
123 U	0.540	0.117	(0.423)	0.046
124 U	0.620	0.000	(0.620)	0.011
125 G	1.577	0.253	(1.323)	0.009
126 G	1.580	0.100	(1.480)	0.025
127 G	3.590	0.027	(3.563)	0.045
128 C	1.097	0.103	(0.993)	0.001
129 C	0.137	1.020	0.880	0.010

Table 4.2: SHAPE reactivity values for the apical loop of WT SLIIId +/- 40S ribosomal subunit

4.2.4 Mutations in SLIIIb and SLIIId can disrupt EqHV 5'UTR interaction with eIF3 and the 40S ribosomal subunit

Both eIF3 and the 40S ribosomal subunit significantly altered the NMIA reactivity of bases in the apical loops of SLIIIb and SLIIId and can be said to be exhibiting an interaction. It was therefore hypothesised that the reduction, and complete ablation, of translation from the GUC and AGU mutations (Figure 4.1), was caused by these mutations disrupting the interactions demonstrated in Figures 4.3 and 4.4. To test this hypothesis, these mutants were subject to the same SHAPE footprinting assays as in 4.2.3.

Unlike the WT (Figure 4.3) the SLIIIb GUC mutant exhibited very little change in NMIA reactivity across the apical loop of SLIIIb in eIF3 SHAPE footprinting. Only ¹²⁶C exhibited any significant change. However, whilst this base did exhibit a significant reduction in mean NMIA reactivity compared to WT, from 2.68 to 0.77, a SHAPE reactivity value over 0.7 is widely considered to be highly reactive. Therefore ¹²⁶C is still considered to be highly reactive to NMIA and as such would not be considered to be protected from NMIA in this assay (Figure 4.5 and Table 4.3). These data indicate that eIF3 is not able to interact with the GUC mutant to protect the apical loop from NMIA reaction. The GUC mutant disrupts the interaction between the apical loop of SLIIIb, and eIF3.

Similarly, the SLIIId AGU mutant exhibited no significant difference in NMIA reactivity in all bases in its apical loop, apart from two, in a footprinting assay using 40S ribosomal subunit compared to WT. The change in NMIA reactivity seen in ¹²⁴U actually represents an increase in NMIA reactivity. As with ¹²⁶C, ¹²⁴U and ¹²⁷U are still both considered to be highly reactive to NMIA in both the WT, and the AGU footprinting assay (Figure 4.6 and Table 4.4). Therefore, whilst there is a significant change, the differences seen here do not impact the conclusion that the 40S ribosomal subunit is not able to interact with the SLIIId AGU mutant.

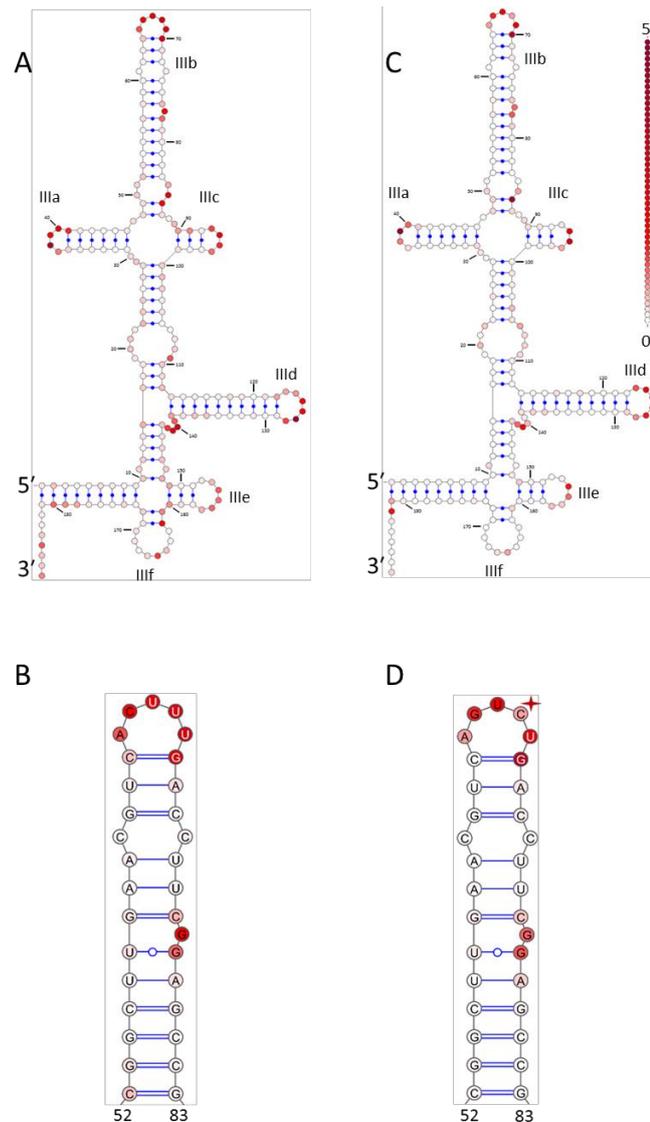


Figure 4.5: A mutation in the apical loop of EHV SLIIIb prevents interaction with eIF3

A) SHAPE analysis of SLIII of the EHV 5'UTR B) Enlarged image of SLIIIb from the SHAPE analysis in A. C) eIF3 SHAPE footprinting analysis of SLIII of the EHV GUC 5'UTR D) Enlarged image of SLIIIb from the footprinting analysis in C. Red stars indicate bases which exhibited a statistically significant reduction in NMIA reactivity when footprinting with eIF3 was conducted, compared to WT no eIF3. SHAPE reactivities are represented on a colour scale from white (low reactivity – predicted paired) to red (high reactivity – predicted unpaired) a representative scale denoting reactivity increments of 0.1 is displayed. Number labelling corresponds to domain III only and number labelling in B and D correspond to number labelling in A and C.

Nucleotide	WT no eIF3	GUC +eIF3	Difference	p value
65 A	1.033	0.790	(0.243)	0.173
66 C/G	1.563	1.807	0.240	0.224
67 U	2.027	2.367	0.340	0.401
68 U/C	2.227	0.770	(1.457)	0.015
69 U	2.693	2.873	0.180	0.293

Table 4.3: SHAPE reactivity values for the apical loop of GUC SLIIIb +eIF3 in comparison to WT without eIF3

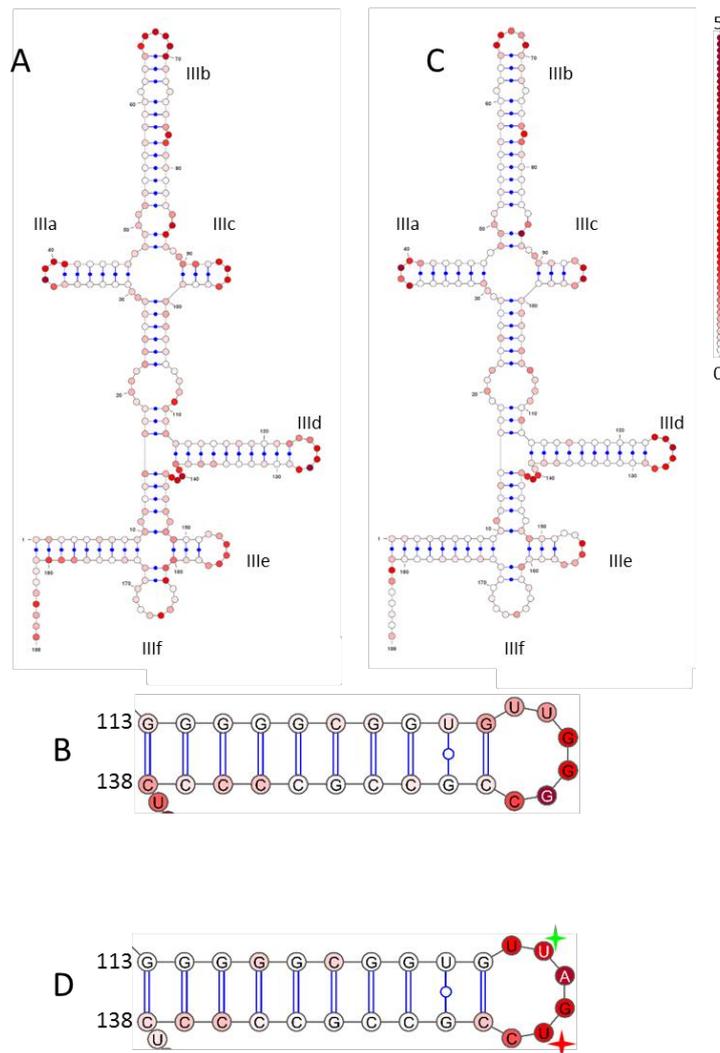


Figure 4.6: A mutation in the apical loop of EqHV SLIIId prevents interaction with the 40S ribosomal subunit

A) SHAPE analysis of SLIII of the EqHV 5'UTR B) Enlarged image of SLIIId from the SHAPE analysis in A. C) 40S ribosomal subunit SHAPE footprinting analysis of SLIII of the EqHV AGU 5'UTR D) Enlarged image of SLIIId from the footprinting analysis in C. Green stars indicate bases which exhibited a statistically significant increase in NMIA reactivity when footprinting with 40S was conducted, compared to WT no 40S; red stars indicate a decrease. SHAPE reactivities are represented on a colour scale from white (low reactivity – predicted paired) to red (high reactivity – predicted unpaired) a representative scale denoting reactivity increments of 0.1 is displayed. Number labelling corresponds to domain III only and number labelling in B and D correspond to number labelling in A and C.

Nucleotide	WT no 40S	AGU +40S	Difference	p value
122 G	0.553	0.000	(0.553)	0.069
123 U	0.540	3.670	3.130	0.288
124 U	0.620	3.240	2.620	0.041
125 G/A	1.577	2.190	0.613	0.070
126 G	1.580	1.050	(0.530)	0.185
127 G/U	3.590	0.960	(2.630)	0.019
128 C	1.097	2.110	1.013	0.402
129 C	0.137	0.110	(0.027)	0.074

Table 4.4: SHAPE reactivity values for the apical loop of AGU SLIIIb +40S ribosomal subunit in comparison to WT no 40S

4.2.5 The EqHV 5'UTR interacts with host cell translational machinery necessary for HCV translation

To confirm the interactions observed in 4.2.3, and to determine other mammalian initiation factors with which the EqHV 5'UTR interacts, pull-down assays were conducted using Huh7 cell lysate and the EqHV 5'UTR RNA as bait. The identity of interacting partners was then determined by tandem mass tagged (TMT) mass spectrometry (MS [TMT MS]) (University of Bristol). Pull downs were carried out using the Pierce™ Magnetic RNA-Protein Pull-Down Kit according to the manufacturer's instructions. Three groups of proteins are described here: eIFs (Table 4.1), 40S ribosomal (Table 4.2), and 60s ribosomal (Table 4.3) proteins. Only proteins identified at an abundance ratio of sample:bead only control >1 were considered to have been specifically pulled down by the WT 5'UTR and are displayed here.

As discussed in 1.2.3 the HCV IRES interacts with a subset of the canonical initiation factors, and the 40S and 60S ribosomal subunits, to initiate translation in a cap-independent fashion. This process does not require eIF4G or eIF4E, and absolutely requires eIF2, eIF3, eIF5B, and the ribosomal subunits to initiate translation. The functional analysis of the EqHV IRES described in Chapter 3 indicates that the function of the individual IRES domains is conserved between these two closely related viruses, and that the mechanism of translation initiation may be conserved. Therefore the EqHV IRES would also require interaction with the same eIFs as HCV, and 40S and 60S subunits, to initiate translation. The TMT MS data shown in Table 4.5, Table 4.6, and Table 4.7 are in good agreement with this hypothesis and demonstrate that the EqHV IRES does not interact with eIF4G or eIF4E, does not interact with the eIFs required for HCV IRES mediated translation initiation, as well as the 40S and 60S ribosomal subunits required to initiate elongation.

This analysis demonstrates the ability of the EqHV 5'UTR to pull-down, and therefore interact with, the complete set of eIFs and ribosomal subunits required for HCV IRES mediated translation. These data provide compelling evidence for a conserved mechanism of translation initiation between HCV and EqHV (discussed in 4.3.3).

Protein FDR Confidence	Description	Abundance Ratio
High	Eukaryotic translation initiation factor 3 subunit I	2.009
High	Eukaryotic translation initiation factor 5A-1	1.792
High	Eukaryotic translation initiation factor 3 subunit F	1.703
High	Eukaryotic translation initiation factor 3 subunit D	1.661
High	Eukaryotic translation initiation factor 5	1.579
High	Eukaryotic translation initiation factor 3 subunit C-like protein	1.548
High	Eukaryotic translation initiation factor 3 subunit E	1.4755
High	Eukaryotic translation initiation factor 3 subunit A	1.4745
High	Eukaryotic translation initiation factor 3 subunit H	1.4325
High	Eukaryotic translation initiation factor 2A	1.3585
High	Eukaryotic translation initiation factor 5B	1.356
High	Eukaryotic translation initiation factor 3 subunit L	1.352
High	Eukaryotic translation initiation factor 3 subunit M	1.329
High	Eukaryotic translation initiation factor 3 subunit J	1.259
High	Eukaryotic translation initiation factor 3 subunit B	1.2535
High	Eukaryotic translation initiation factor 2 subunit 3	1.2255
High	Eukaryotic translation initiation factor 2 beta	1.212
High	Eukaryotic translation initiation factor 2 subunit 1	1.195

Table 4.5: eIF interacting partners of the EqHV 5'UTR

A list of eIFs identified to interact with the EqHV 5'UTR from pulldown-TMT MS analysis. FDR confidence is the False discovery rate confidence (high=1% medium=5%). Abundance ratio is the ratio protein abundance for the WT 5'UTR pull-down to the bead only control.

Protein FDR Confidence	Description	Abundance Ratio
High	40S ribosomal protein S29	2.68
High	40S ribosomal protein S30	2.56
High	40S ribosomal protein S12	2.34
High	40S ribosomal protein S11	1.91
High	40S ribosomal protein S13	1.77
High	40S ribosomal protein S9	1.74
High	40S ribosomal protein S18	1.7
High	40S ribosomal protein S25	1.68
High	40S ribosomal protein S2	1.66
High	40S ribosomal protein S24	1.64
High	40S ribosomal protein S26	1.63
High	40S ribosomal protein S16	1.55
High	40S ribosomal protein S17	1.54
High	40S ribosomal protein S4	1.53
High	40S ribosomal protein S15	1.52
High	40S ribosomal protein S3a	1.42
High	40S ribosomal protein S15a	1.4
High	40S ribosomal protein S27	1.4
High	40S ribosomal protein S8	1.4
High	40S ribosomal protein S18	1.39
High	40S ribosomal protein S20	1.39
High	40S ribosomal protein S3	1.39
High	40S ribosomal protein S14	1.34
High	40S ribosomal protein S6	1.28
High	40S ribosomal protein SA	1.27
High	40S ribosomal protein S7	1.21
High	40S ribosomal protein S27	1.14
High	40S ribosomal protein S5	1.13

Table 4.6: 40S ribosomal interacting partners of the EqHV 5'UTR

A list of 40S ribosomal proteins identified to interact with the EqHV 5'UTR from pulldown-TMT MS analysis. FDR confidence is the False discovery rate confidence (high=1% medium=5%). Abundance ratio is the ratio protein abundance for the WT 5'UTR pull-down to the bead only control.

Protein FDR Confidence	Description	Abundance Ratio
Medium	60S ribosome subunit biogenesis protein NIP7 homolog	3.5155
High	60S ribosomal protein L37a	2.1245
High	60S ribosomal protein L35	1.8765
High	60S ribosomal protein L34	1.8545
High	60S ribosomal protein L10	1.8475
High	60S ribosomal protein L27	1.8475
High	60S ribosomal protein L18a	1.8225
High	60S ribosomal protein L13a	1.8055
High	60S ribosomal protein L22	1.777
High	60S ribosomal protein L27a	1.77
High	60S ribosomal protein L36	1.7505
High	60S ribosomal protein L32	1.7465
High	60S ribosomal protein L19	1.713
High	60S ribosomal protein L37	1.7
High	60S ribosomal protein L8	1.6875
High	60S ribosomal protein L24	1.6745
High	60S ribosomal protein L10a	1.653
High	60S ribosomal protein L7a	1.6285
High	60S ribosomal protein L3	1.6175
High	60S ribosomal protein L10	1.615
High	60S ribosomal protein L35a	1.6015
High	60S ribosomal protein L6	1.5355
High	60S ribosomal protein L26	1.5135
High	60S ribosomal protein L23a	1.475
High	60S ribosomal protein L26-like 1	1.3825
High	60S ribosomal protein L12	1.3245
High	60S ribosomal protein L30	1.287
High	60S ribosomal protein L11	1.247
High	60S ribosomal protein L36a	1.152
High	60S ribosomal protein L13	1.044

Table 4.7: 60S ribosomal interacting partners of the EqHV 5'UTR

A list of 60s ribosomal proteins identified to interact with the EqHV 5'UTR from pulldown-TMT MS analysis. FDR confidence is the False discovery rate confidence (high=1% medium=5%). Abundance ratio is the ratio protein abundance for the WT 5'UTR pull-down to the bead only control.

4.3 Discussion

4.3.1 eIF3 interacts with the apical loop of EqHV 5'UTR SLIIIb during IRES mediated translation

As discussed in 3.3.8 the proximity of the EqHV IRES to the end of cistron one in the bicistronic construct may have been compensating for the function of SLIIIb in the Δ SLIIIb mutant IRES, therefore only the SGR was used to further investigate the role of SLIIIb in EqHV IRES mediated translation. Here a $_{250}\text{CUU}_{252}$ to $_{250}\text{GUC}_{252}$ mutation was introduced into the apical loop of SLIIIb to investigate if this mutation could also knock down translation in the same way as Δ SLIIIb (Figure 4.1). Whilst this mutation did not completely ablate translation it did reduce activity to 40-70% of WT. eIF3 has previously been demonstrated to protect the apical loop of HCV SLIIIb in SHAPE footprinting. The translation efficiency of a deletion of the apical portion of HCV SLIIIb resulted in 34% translation efficiency of WT, similar to that observed in Huh7 cells for the GUC mutant (Sun et al., 2013). This study also demonstrated $_{214}\text{AAU}_{216}$, which resides within a mismatched loop in the HCV SLIIIb, interacts with eIF3. However this loop is not conserved within EqHV SLIIIb in the structure described in this study (Figure 4.2). Therefore, it was hypothesised that the reduced IRES efficiency seen in the GUC mutant was caused by this mutant's inability to interact with eIF3 via the apical loop.

SHAPE footprinting analysis confirmed that eIF3 was indeed interacting with the apical loop of EqHV SLIIIb, causing a significant reduction in NMIA reactivity in this region. However, unlike studies conducted on eIF3-HCV 5'UTR interactions, no significant reduction in NMIA reactivity was observed for any other base in the apical portion of SLIII. This evidence suggests that eIF3 exclusively interacts with the apical loop of EqHV SLIIIb. However, the apical loops of SLIIIa, and SLIIIc, are completely conserved between HCV and EqHV, suggesting that they share a conserved function. Mutation of these SLs in HCV reduced translation to <10% of WT with the deletion of SLIIIa resulting in <5.5 fold affinity for eIF3. These data therefore raise some interesting questions about the role of SLIIIb in interacting with eIF3 in EqHV IRES-dependent translation.

HCV SLIIIb has been extensively shown to interact with eIF3 to facilitate efficient translation. The SHAPE footprinting conducted here confirmed that eIF3 specifically interacts with the apical loop of EqHV SLIIIb, so it was therefore of no surprise that deletion of this subdomain ablates translation. However the $_{250}\text{GUC}_{252}$ mutation, which disrupts the interaction of eIF3 and SLIIIb, does not completely block translation. It is therefore possible that another interaction site exists. Given the lack of any significant alteration to NMIA reactivity throughout SLIII upon eIF3 binding

it may be the case that eIF3 is interacting with a structural motif of SLIII which is in a base paired confirmation. This would preclude NMIA from reacting in the natural confirmation and thus protection by eIF3 would not be observable by this method. Furthermore, the SLIIIabc junction has been demonstrated, for HCV, to be an important determinant of eIF3 interaction and translation efficiency (Odreman-Macchioli et al., 2000). The EqHV Δ SLIIIb mutant encompassed a deletion of the entirety of SLIIIb, therefore disrupting the SLIIIabc junction and this could provide an explanation for the differences observed between Δ SLIIIb and ${}_{250}\text{GUC}_{252}$ in their ability to initiate translation. Whilst these data suggest a role for the SLIIIabc junction in EqHV translation, it does appear to rule out any specific interactions between SLIIIa, or SLIIIC, and eIF3.

It is hypothesised that both sequence and structural features are involved in EqHV SLIII – eIF3 interactions during translation initiation. eIF3 specifically interacts with the apical loop of SLIIIb to enhance IRES activity, but this is not the main determinant of translation with regards to eIF3. There are two hypotheses for what this main determinant may be. Either eIF3 interacts with SLIIIb in a structure specific manner; or eIF3 interacts with the SLIIIabc junction. A more detailed mutational analysis to investigate these possibilities, as well as the roles of SLIIIa and SLIIIC, could answer these questions. Alternative methods of RNA structure probing, such as enzymatic cleavage analysis, could also be used to further our understanding of EqHV 5'UTR interactions with eIF3.

4.3.2 40S ribosomal subunit interacts with ${}_{309}\text{GGG}_{311}$ during IRES mediated translation

As discussed in 3.3.2 the Δ SLIIId mutation completely ablated translation from the EqHV IRES and this was hypothesised to be due to the absence of the conserved GGG motif within the apical loop (Figure 4.2). In order to test this a ${}_{309}\text{GGG}_{311}$ to ${}_{309}\text{AGU}_{311}$ mutation, previously demonstrated to reduce HCV translation to ~10% of WT translation was introduced into pSGR NZCI (Jubin et al., 2000; Otto and Puglisi, 2004; Barría et al., 2009). The data shown in Figure 4.1 demonstrate that a similar effect can be observed for mutation of the GGG triplet in EqHV, with translation efficiency <20% of WT, and as low as 10%.

This GGG motif has been demonstrated, in HCV, to interact with ${}_{1116}\text{CCC}_{1118}$ of the 40S ribosomal subunit, exhibiting a protective effect in SHAPE footprinting. Conversely the 40S ribosome has been demonstrated to have a protective effect on ${}_{266}\text{GGG}_{268}$ in the HCV apical loop (Angulo et al., 2016) with the corresponding HCV AGU mutant negating these effects. It was therefore hypothesised that, due to the similar functional effects demonstrated for this mutation, the

EqHV_{309GGG311} motif also interacts with the 40S subunit. SHAPE footprinting analysis confirmed that 40S ribosomal subunit had a protective effect on the apical loop of EqHV SLIIId, with all bases in the apical loop exhibiting a significant reduction in NMIA reactivity in this experiment. No bases exhibited a loss of NMIA reactivity upon the addition of 40S subunit to the AGU mutant 5'UTR. Whilst the corresponding experiments for protection of _{1116CCC1118} by EqHV SLIII were not conducted, the structural, sequence, and functional conservation of these bases between EqHV and HCV is strongly indicative that _{309GGG311} of EqHV SLIIId interacts with _{1116CCC1118} of the 40S ribosomal subunit.

Interestingly, for both eIF3 and 40S SHAPE footprinting, an increase in NMIA reactivity was observed in bases immediately downstream of the interaction site. However it is believed that this is an artefact of the experimental setup. RNA was not phenol chloroform extracted to remove protein following NMIA reaction. This step was not included due to loss of sample and downstream detection issues. It was expected that the high temperature used for RT would prevent any protein RNA interactions forming during the extension and therefore not affect the experiment. However, low levels of binding during the extension could have occurred and thus halted RT progression, resulting in a high number of transcription stops at this point and a false positive NMIA peak. This appears to be the most likely cause of the increase in NMIA reactivity and this explanation fits well with the lack of significance exhibited in the eIF3 footprinting. Furthermore it is not believed that this artefact had an impact on the overall data obtained from the footprinting analysis as there was a clear and reproducible decrease in NMIA activity in the apical loops. Protein binding detection as a false positive NMIA signal would cause bases to appear highly NMIA reactive.

4.3.3 The EqHV 5'UTR interacts with all of the factors required for HCV translation initiation

The HCV IRES requires only a subset of initiation factors to efficiently initiate translation of its polyprotein. The 40S ribosomal subunit is directly recruited to the IRES through a specific interaction with SLIIId, and correct folding of the pseudoknot correctly positions the AUG start in the mRNA binding cleft. eIF3 is recruited through a specific interaction with SLIIIb, and this ternary complex (TC) associates with eIF2/GTP/Met-tRNA to form the 48S particle. Following hydrolysis of GTP eIF2 releases Met-tRNA, and is itself released from the 48S complex. eIF5B then facilitates a further GTP hydrolysis reaction which allows binding of the 60S subunit, resulting in the assembly of an 80S ribosome capable of initiating translation (Ji, 2004). As can

be observed in Table 4.1 the EqHV WT 5'UTR is able to pull-down a number of eIF3 subunits, including eIF3a which has been demonstrated to be the most important in HCV IRES activity (Sun et al., 2013). A number of eIF2 subunits, and eIF5B, were also pulled down. This demonstrated that the EqHV 5'UTR was interacting all of the initiation factors involved in the expression of the HCV polyprotein from the IRES. Interestingly eIF1A was not shown to associate with the EqHV 5'UTR. This interaction was demonstrated for HCV by (Jaafar et al., 2016) who also suggested an updated mechanism of translation initiation whereby the HCV binds an assembled pre initiation complex (PIC).

The EqHV 5'UTR is also shown to be associated with a number of 40S (Table 4.2) and 60S (Table 4.3) ribosomal proteins. These data therefore demonstrate that the EqHV 5'UTR is specifically interacting with eIF3 and the 40S ribosomal subunit, and that the 80S ribosome is highly likely to assemble on the EqHV AUG start following eIF2 and eIF5b binding.

4.3.4 A predicted model of EqHV translation initiation

In this study a complimentary functional and structural analysis of the EqHV 5'UTR IRES was undertaken which demonstrated that SLI is not involved in IRES mediated translation, SLII is not essential but enhances translation, and that SLIII is absolutely necessary. The apical loop of SLIIIb interacts with eIF3 and whilst this interaction is not necessary for IRES function the presence of SLIIIb is crucial. SLIIIc is absolutely required for IRES activity and this is due to a conserved GGG motif which interacts with the 40S ribosome. The EqHV 5'UTR also interacts with all of the factors involved in HCV translation initiation.

Based upon the data presented here it is proposed that EqHV and HCV share a conserved mechanism of translation initiation. The data presented here provides a compelling argument for this functional conservation and this also potentially represents a conserved mechanism of hepaciviral translation initiation (Figure 1.1B). However, investigations into the mechanisms employed by the other newly discovered hepacivirus is needed to confirm this.

Chapter 5 Rational modification of an EqHV sub-genomic replicon to enhance replicative capacity

5.1 Introduction

Sub-genomic replicons (SGRs) are widely used to study an array of positive-sense RNA viruses including picorn-, alpha-, and flavi- viruses. The principles of an SGR are that the structural proteins of a given virus are replaced with a reporter gene so that virus replication can be studied in the absence of virus particle production. Such constructs have provided extremely useful tools to research dangerous pathogens which are required to be handled only at high containment levels. This has increased research access to, and therefore knowledge of, these viruses as their SGRs can be handled at lower containment levels than infectious virus.

SGRs are a widely used system to study the replication of HCV, and the function of its non-structural proteins. The HCV genome is ~10 Kb in length with 5' and 3' UTRs. Expression of the large polyprotein is driven by the HCV 5'UTR IRES and is cleaved by cellular, and viral, proteases to yield 10 proteins. The first four constitute the structural genes (core, E1, E2, and p7) and the remaining six the non-structural genes (NS2, NS3, NS4A, NS4B, NS5A, NS5B). The HCV proteins NS3-5B constitute the minimal components for HCV replication. The HCV SGR consists of the 5'UTR, which drives expression of a reporter gene, followed by the EMCV IRES which drives expression of the non-structural genes NS3-5B, and finally the HCV 3'UTR. A GDD-GND/N mutation in the active site of NS5B is commonly used in the field as a replication deficient negative control (Lohmann et al., 1999).

Replication of the HCV SGR is assayed in two main ways: stable and transient. Stable replication (also referred to as stable cell lines) utilises an antibiotic resistance gene in place of the structural genes which, following transfection of IVT HCV SGR RNA into cells, can be selected to maintain a population of cells which harbour the SGR. Transient replication utilises a reporter gene, commonly firefly luciferase (FLuc), with chemiluminescent or fluorescent properties which can be quantified to assess levels of replication. Both systems have inherent advantages and disadvantages: stable cell lines can often support the replication of viral isolates for which replication cannot be detected in a transient assay, and can be used to select for mutations which enhance replication in mammalian cells. However they are time consuming to create, and select against deleterious mutations. On the other hand transient SGR assays allow for the study of viral growth kinetics, and the effects of deleterious or replication suppressive mutants, but virus isolates with low levels of replication may not be able to be detected transiently.

Traditionally the reporter gene used for HCV SGRs has been Feo, a fused firefly luciferase/neomycin phosphotransferase gene (**FLuc/Neo**), as this allows stable cell lines to be selected for whilst also providing a system to easily detect active replication both in the transient system and in the stable cell line. However, other reporter genes such as puromycin resistance, blasticidin resistance, fluorescent proteins, and Neo and Fluc in isolation have also been used. Recently it was discovered that a CpG and UpA low FLuc gene was able to increase detection of transient HCV SGR replication by 100 fold (Atkinson et al., 2014).

The advent of the HCV SGR represented a huge leap forward in HCV research and has facilitated a litany of discoveries that have culminated in the development of pan-genotypic direct acting antivirals (DAAs) which have ushered in a new era of IFN-free HCV therapy with remarkable success (reviewed in Zhang et al., 2016). However, HCV still remains a global health concern and, currently, no models exist for comparative study. EqHV is the most closely related virus to HCV, but exhibits no significant pathology in its natural host, equines. It therefore has the potential to provide illuminating data regarding how, and why, such closely related viruses exhibit such disparity in their disease causing capabilities. Therefore I set out to establish an EqHV SGR capable of efficient replication in mammalian cell culture to provide such a system.

Sections 5.2.1, 5.2.2, and 5.2.3 of this Chapter describe the construction and assay of a “first generation” EqHV sub-genomic SGRs using isolate EF369_11J; GenBank accession no. JX948116.1. Sections 5.2.4, and 5.2.5 of this chapter describe the construction and assay of SGRs derived from the consensus NZPI sequence (GenBank accession no. KP325401).

5.2 Results

5.2.1 First generation EqHV, and chimeric, SGRs

The original EqHV SGR was constructed in the image of the widely used HCV SGRs described in 5.1, and had an identical organisation (Figure 5.2A). The NS3-5B region of EqHV (isolate EF369_11J; GenBank accession no. JX948116.1) was PCR amplified and introduced into the JFH-1 HCV SGR construct, in place of the HCV NS3-5B region by H. Stewart and C. Walter (University of Leeds). The reporter gene used here was the Feo gene, so that both transient and stable replication could be investigated. The EMCV IRES drove translation of EqHV NS3-5B, and this was followed by the HCV 3'UTR. This construct was given a three letter designation H-E-H with each letter representing the 5'UTR, polyprotein, and 3'UTR, with H representing the HCV sequence and E the EqHV sequence (Figure 5.2A). Therefore the construct designated pSGR H-E-H would contain the HCV 5'UTR, EqHV polyprotein, and the HCV 3'UTR. This nomenclature is used throughout this chapter to refer to all of the chimeric SGRs constructed. In order to investigate replication of the EqHV SGRs it was first necessary to create a replication deficient clone equivalent to those used for HCV. The accepted replication deficient control for HCV replication is an NS5B RNA dependent RNA polymerase (RdRP) active site (GDD) mutation to GND/GNN. EqHV also contains the classical GDD motif in its NS5B RdRP, therefore a GDD → GNN mutation was introduced by Q5 SDM (NEB) to create a replication deficient control for replication assays. It was hypothesised that, like for many HCV isolates, EqHV would require culture adaptive mutations to replicate in cell culture. It has become common to artificially introduce adaptive mutations into HCV sequences in an attempt to facilitate their replication in cell culture. A common mutation, S2204I, has been demonstrated to significantly enhance HCV replication. Alignment of the EqHV and HCV polyprotein sequences revealed a conserved serine in the EqHV NS5A protein (S2180) equivalent to S2204 (Figure 5.1). An S→I mutation was introduced at this point (S2180I) to try and achieve EqHV replication in cell culture. To test the replicative ability of SGR H-E-H 2 µg IVT RNA was transfected by electroporation into Huh7, or Huh7.5, cells and assayed for firefly luciferase activity at 4, 24, 48, and 72 hours post transfection (h.p.t) (Figure 5.2B and C). It was clear to see that SGR H-E-H was unable to replicate in either Huh7, or Huh7.5, cells as luciferase activity declined from 4 to 72 hours with no detectable signal above that of the replication deficient negative control GNN at any time point. A representative replication graph for SGR JFH (SGR H-H-H), and a representative RNA gel of input RNA, can be seen in Appendices Figure 8.11 A and B respectively).

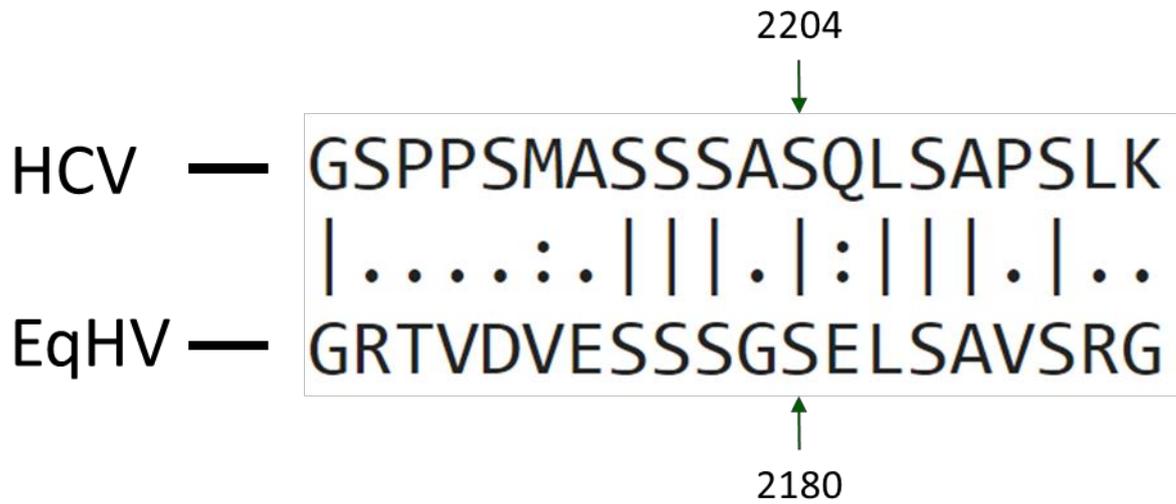


Figure 5.1: HCV S2204 is conserved in EqHV at a.a position 2180

The NS5A low complexity sequence (LCS) 1 of the HCV H77 reference sequence (Accession number AF011753.1) was aligned with the corresponding region of the NZPI EqHV a.a. sequence (Accession number KP325401) using the online tool Emboss Needle (https://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html) . The conserved serine (S) residue at positions 2204, and 2180 are indicated.

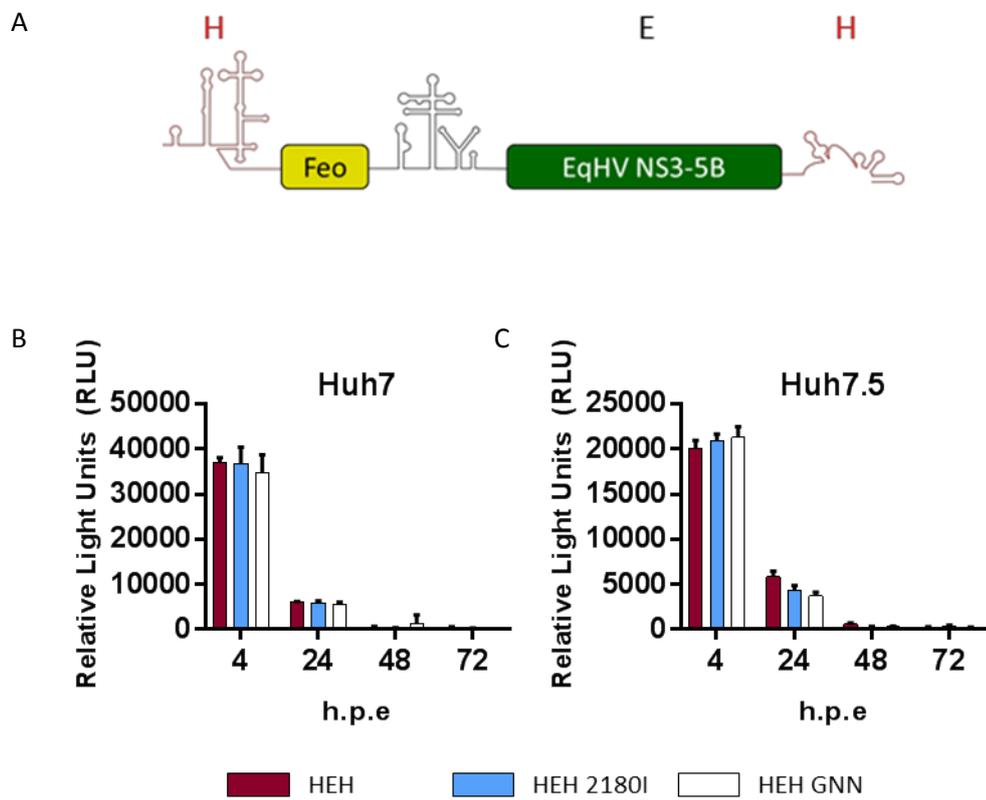


Figure 5.2: Transient replication of SGR H-E-H in human hepatoma cells

A) A schematic representation of SGR H-E-H B) SGR H-E-H, S2180I, and GNN, RNA was transfected into Huh7 or C) Huh7.5 cells and assayed for firefly luciferase activity at 4, 24, 48, and 72 h.p.t.

In order to create a complete EqHV SGR it was necessary to introduce the 5' and 3' UTRs into the H-E-H construct. The first of these to be introduced was the 3'UTR. The EqHV 3'UTR sequence (GenBank accession no. JX948116.1) was synthesised in a geneblock (GENEART) along with the 3' end of the coding region for NS5B to include an Sfil site (within NS5B) and an XbaI site (downstream of the UTR) to facilitate molecular cloning of the 3'UTR into the H-E-H construct.

The 3'UTR was introduced into the H-E-H, S2180I, and GNN SGRs to create an SGR which consisted of the HCV 5'UTR, the EqHV NS3-5B region, and the EqHV 3'UTR (termed H-E-E) (Figure 5.3A). SGR H-E-E, H-E-E S2180I, and H-E-E GNN were transfected into Huh7 (Figure 5.3B), and Huh7.5 (Figure 5.3C), cells and assayed for firefly luciferase activity at 4, 24, 48, and 72 h.p.t. Figure 5.3B and 6.3C demonstrate that no transient replication of either H-E-E or H-E-E S2180I could be detected up to 72 h.p.e. Firefly luciferase activity consistently decreased from 4 to 72 hours h.p.e with no detectable signal above that of the replication deficient negative control GNN from 24 h.p.e onwards.

To fully characterise the replicative ability of chimeric HCV/EqHV SGRs the EqHV 5'UTR, and a Δ SLI mutant, were introduced into pSGR H-E-H, 2180I, and GNN. The full EqHV 5'UTR, and one excluding SLI, had previously been isolated and characterised for their translational ability (Stewart et al., 2013), and two XmnI sites, located in the backbone of the plasmid, and in the EMCV IRES, were used to introduce these 5'UTRs into pSGR H-E-H, S2180I and GNN, to create an SGR consisting of the EqHV 5'UTR, EqHV NS3-5B coding region, and the HCV 3'UTR (termed E-E-H) (Figure 5.4A).

SGR E-E-H was transfected into Huh7 (Figure 5.4B), and Huh7.5 (Figure 5.4C), cells and assayed for firefly luciferase activity up to 72 h.p.e. Figure 5.4 clearly demonstrates that SGR E-E-H, Δ SLI, and S2180I, were unable to replicate in either cell line. Like H-E-E, firefly luciferase activity consistently decreased from 4 to 72 hours h.p.e with no detectable signal above that of the replication deficient negative control GNN from 24 h.p.e onwards.

Put together these data demonstrate that chimeric EqHV/HCV SGRs were not able to replicate in the human hepatoma cell lines Huh7 and Huh7.5.

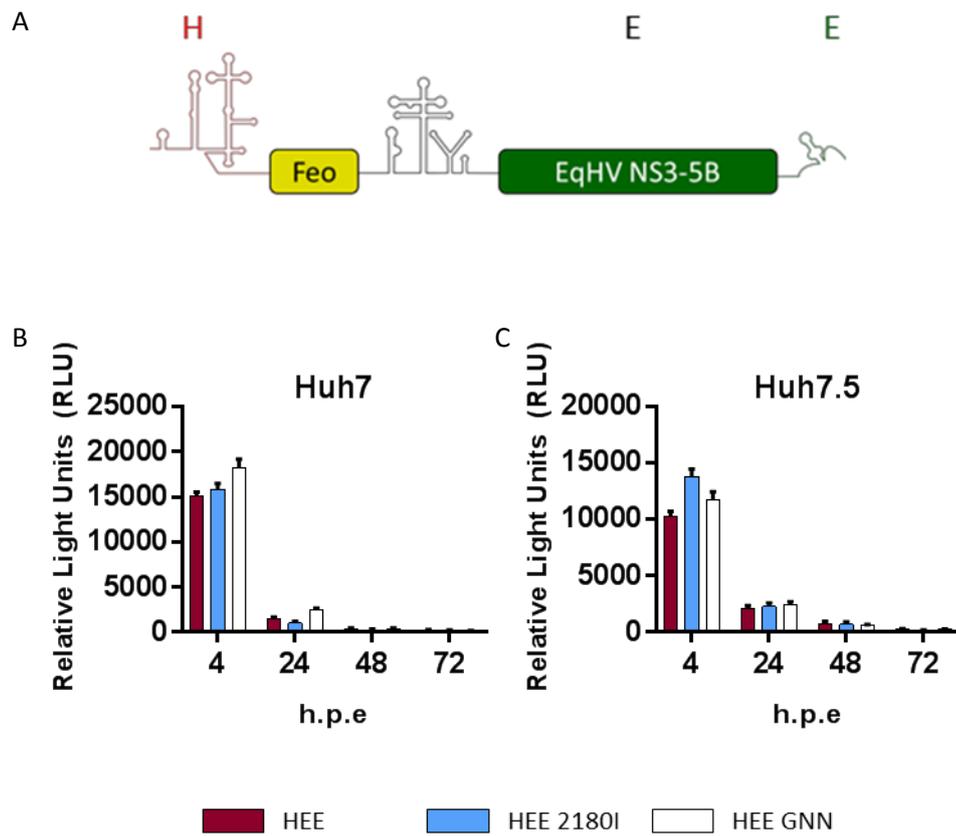


Figure 5.3: Transient replication of SGR H-E-E in human hepatoma cells

A) A schematic representation of SGR H-E-E B) SGR H-E-E, S2180, and GNN, RNA was transfected into Huh7 or C) Huh7.5 cells and assayed for firefly luciferase activity at 4, 24, 48, and 72 h.p.t.

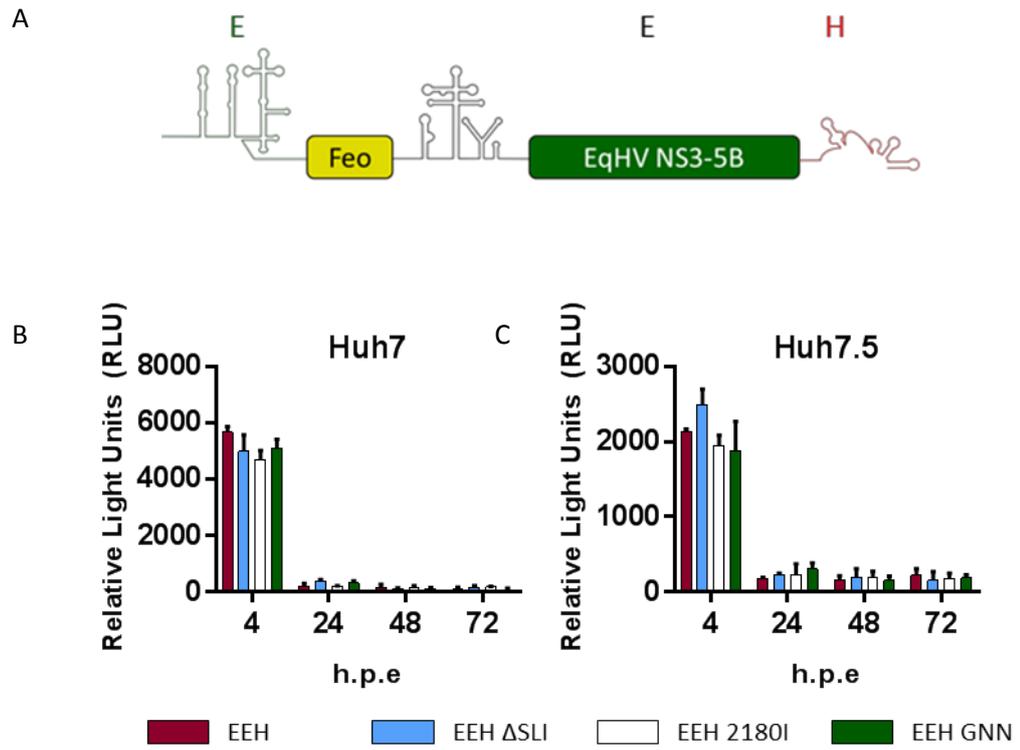


Figure 5.4: Transient replication of SGR E-E-H in human hepatoma cells

A) A schematic representation of SGR E-E-H B) SGR E-E-H, S2180, and GNN, RNA was transfected into Huh7 or C) Huh7.5 cells and assayed for firefly luciferase activity at 4, 24, 48, and 72 h.p.t.

Finally, a full EqHV SGR was created through the molecular cloning of the EqHV 3'UTR from pSGR H-E-E, into pSGR E-E-H, using the route describe above for the introduction of the 3'UTR into pSGR H-E-H (and the S2180I and GNN variations), to create an SGR which contained the EqHV 5' and 3' UTRs, and the EqHV NS3-5B coding sequence (termed SGR EqHV) (Figure 5.5A) (See Table 5.1 for a full list of constructs used in this study). Due to the complex RNA-RNA and RNA-protein interactions that take place during HCV replication the lack of replication observed in the chimeric constructs could have been due to a lack of these interactions between the different RNA sequences and structures, and proteins. Therefore it was hypothesised that a full EqHV SGR was more likely to replicate than a chimeric one.

However, upon transfection into Huh7 (Figure 5.5B), and Huh7.5 (Figure 5.5C) cells, and assay up to 72 h.p.t, SGR EqHV did not exhibit any signs of transient replication. As for the chimeric SGRs firefly luciferase activity consistently decreased from 4 to 72 h.p.t with no detectable signal above that of the replication deficient negative control GNN from 24 h.p.e onwards.

Transient replication could not be detected from any of the SGRs described here up to 72 h.p.t. It can therefore be concluded that the EqHV isolate that was here used to create either chimeric, or full, EqHV SGRs is not able to replicate transiently in either Huh7, or Huh7.5 cells.

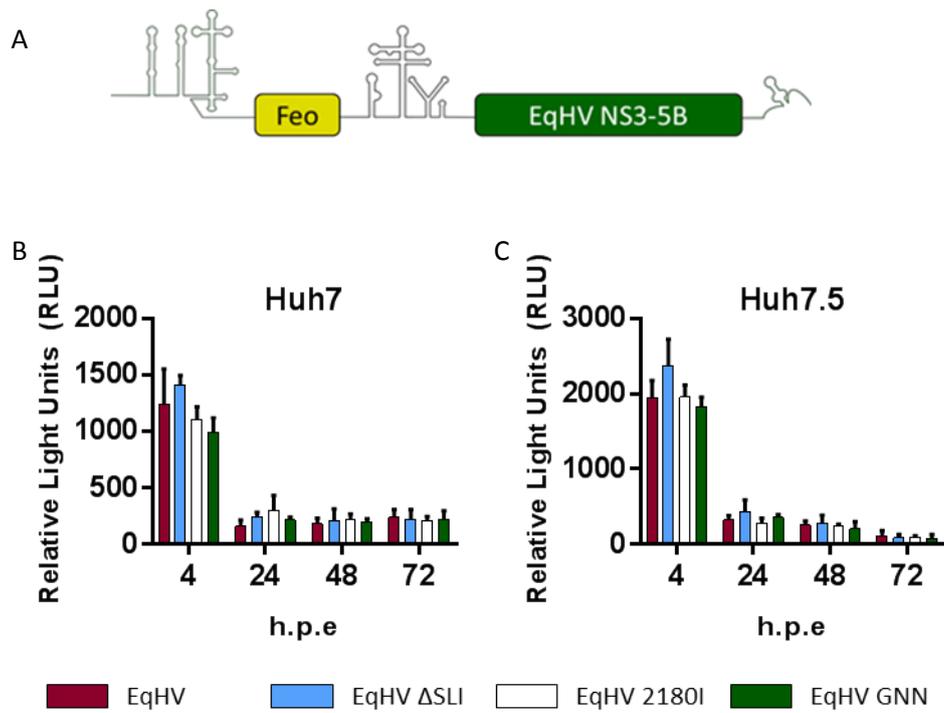


Figure 5.5: Transient replication of SGR EqsHV in human hepatoma cells

A) A schematic representation of SGR EqsHV B) SGR EqsHV, S2180, and GNN, RNA was transfected into Huh7 or C) Huh7.5 cells and assayed for firefly luciferase activity at 4, 24, 48, and 72 h.p.t.

5.2.2 Investigating alternative cell lines for SGR EqHV replication

Given the inability of EqHV to replicate in human hepatoma cells it was hypothesised that an antiviral factor may have been inhibiting replication, or a necessary proviral factor may have been lacking. The natural host of EqHV is the horse, therefore an equine cell line was sought, as well as an alternative human cell line, to assess the ability of EqHV to replicate in equine, and non-hepatoma human cells. Only SGR EqHV, and the Δ SLI, S2180I, and GNN variants, were taken forward for this investigation.

EqHV is a liver tropic virus which infects horses, therefore a liver derived equine cell line represents the ideal candidate to achieve EqHV replication in cell culture. However, to our knowledge, there exists no liver derived equine cell lines. Therefore, a non-liver equine cell line was sought as an alternative. Whilst there is a paucity of equine cell lines generally available from any organ, Foetal horse kidney (FHK) cells were provided by Dr Hiroshi Sentsui (Nihon University, Japan) (Maeda et al., 2007). This cell line has been reported to support the replication of a number of equine viruses (Oguma et al., 2013) and therefore, given the lack of equine liver cell lines, provided the next best cell line in which to achieve SGR EqHV replication. It was hypothesised that an equine derived cell line may express an equine specific host factor necessary for EqHV replication, or alternatively lack expression of an antiviral host factor which inhibits EqHV replication.

The data described in Chapter 3, and published data, demonstrated a role for miR122 in EqHV IRES mediated translation (Scheel et al., 2015), a conserved feature with HCV. miR122 has been demonstrated to be essential for HCV replication and its expression can enable HCV replication in cell lines that lack endogenous miR122 expression. Therefore the FHK 122 cell line (described in Chapter 3) was also included. It was reasoned that the expression of miR122 may cause FHK cells to become permissive to EqHV replication, if the parental cell line was not. Furthermore, expression of miR122 has been demonstrated to cause 293T cells to become permissive to HCV replication and so this cell line (as described in Chapter 3) and the parental 293T cell line were also assayed for EqHV replication. However, none of the FHK, FHK 122, 293T, or 293T 122, cell lines were able to support transient SGR EqHV replication up to 72 h.p.t (Figure 5.6).

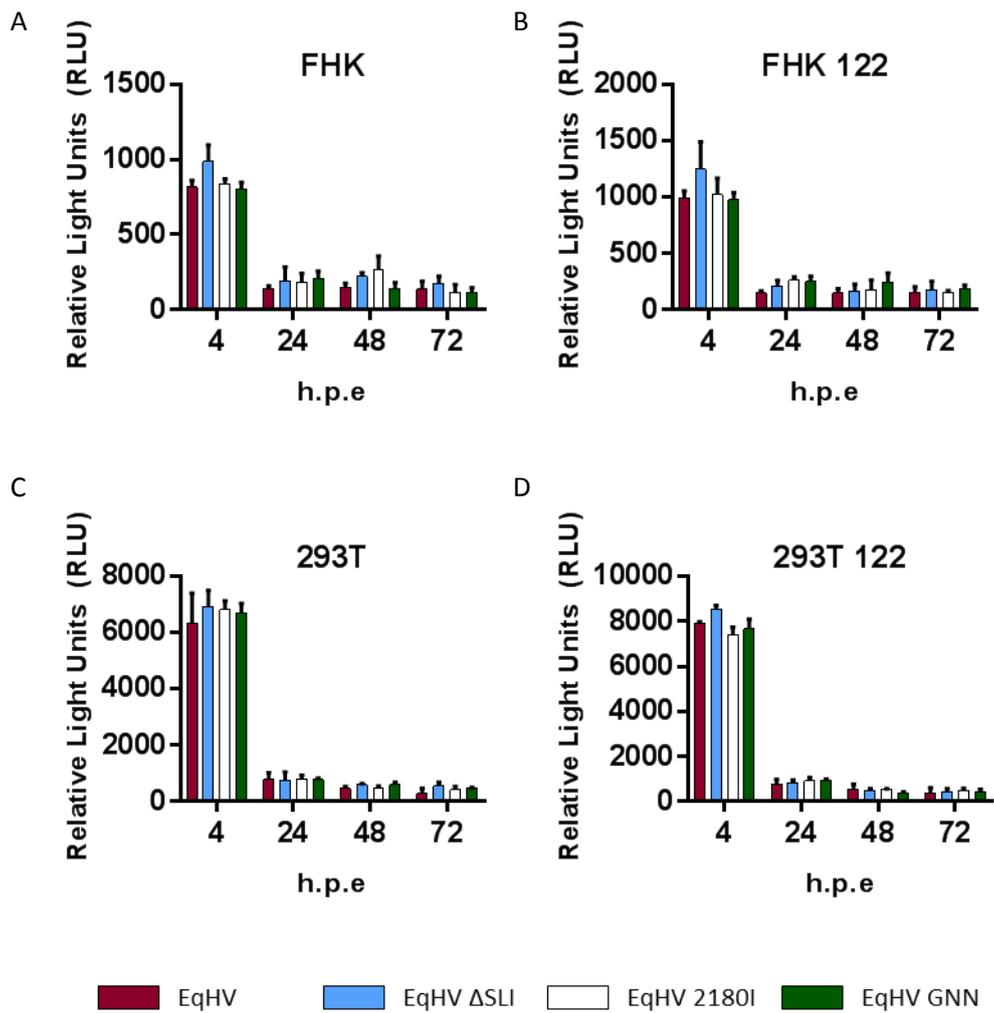


Figure 5.6: Transient replication of SGR EqHV in FHK and 293T (+/- 122) cells

SGR EqHV, Δ SLI S2180, and GNN, RNA were transfected into A) FHK B) FHK 122 C) 293T and D) 293T 122 cells and assayed for firefly luciferase activity at 4, 24, 48, and 72 h.p.t.

As transient replication could not be detected in FHK or 293T (+/- miR122) cells, other factors which have been demonstrated to enhance, or enable, HCV replication in cell culture were investigated. SEC14L2 is a human protein that has been recently demonstrated to enable RNA replication of a diverse range of HCV genotypes. Expression of this protein in Huh7.5 cells also supports HCV replication following inoculation with patient sera (Saeed et al., 2015). It was therefore hypothesised that the dramatic effects of this protein on HCV replication might also apply to EqHV, and expression of this protein may enable EqHV SGR replication. Huh7, Huh7.5, FHK, and 293T cells were therefore all transduced to stably express SEC14L2, and selected with 2.5 µg/ml puromycin to maintain expression.

The parainfluenza virus type 5 (PIV5) V protein has been demonstrated to inhibit STAT1-mediated immune activation by binding STAT1 to inhibit downstream interferon-α activation (Chambers and Takimoto, 2009) and this immune inhibition has been demonstrated to enhance HCV replication (Andrus et al., 2011). It was therefore hypothesised that the enhancing effect of the V protein on HCV replication may provide the correct cellular environment to enable replication of SGR EqHV.

Huh7, Huh7.5, FHK, and 293T cells expressing SEC14L2, and Huh7.5 cells which stably express the PIV5 V protein (provided by S.Griffin, University of Leeds), were transfected with SGR EqHV RNA and assayed for transient replication up to 72 h.p.t (Figure 5.7). Transient replication was not detectable for SGR EqHV, or 2180I, in any of the cell lines.

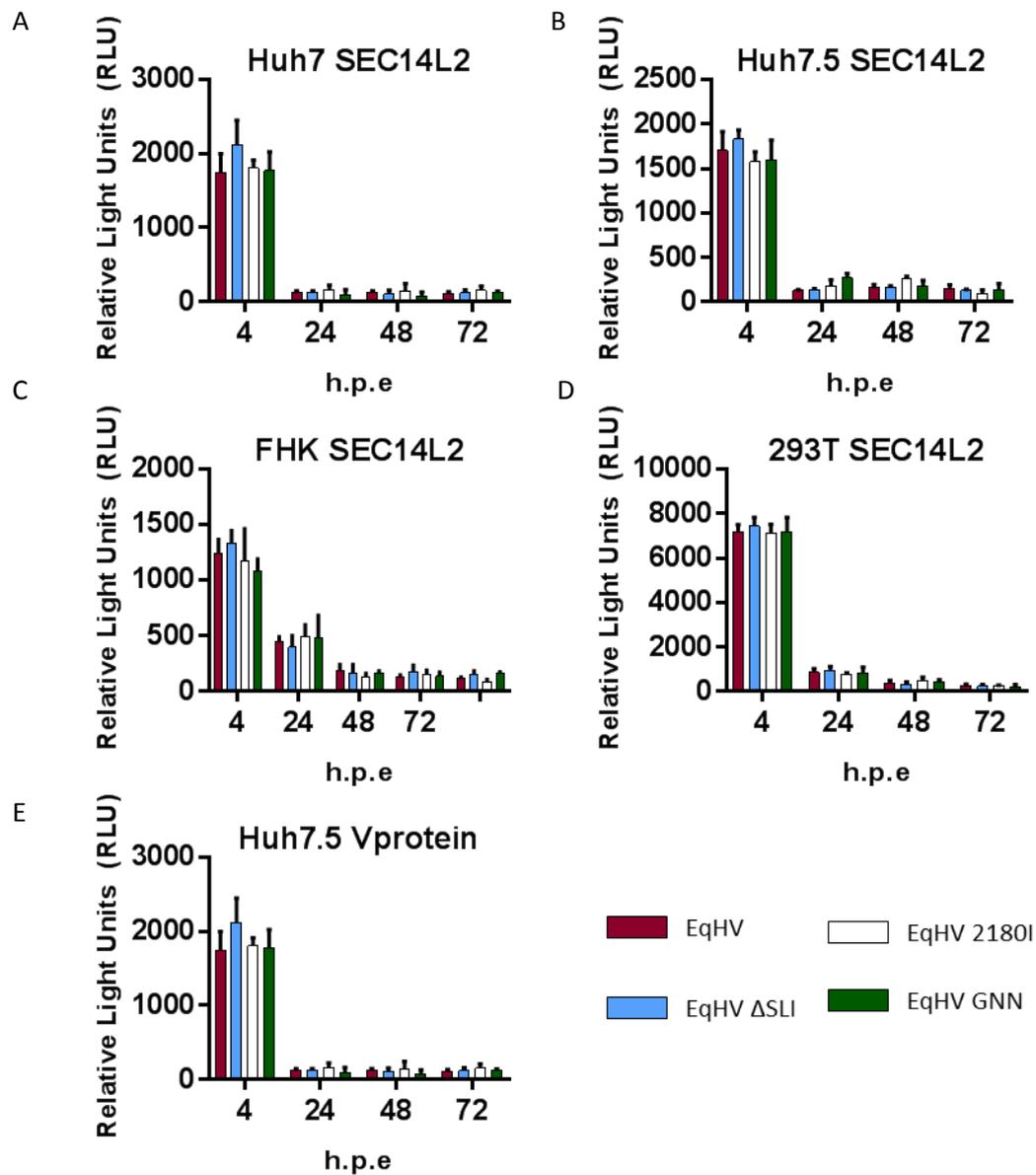


Figure 5.7: Transient replication of SGR EqHV in cell lines expressing SEC14L2 or the PIV5 V protein.

SGR EqHV, S2180, and GNN, RNA was transfected into A) Huh7 SEC14L2 B) Huh7.5 SEC14L2 C) FHK SEC14L2 D) 293T SEC14L2 and E) Huh7.5 V protein cells and assayed for firefly luciferase activity at 4, 24, 48, and 72 h.p.t.

5.2.3 Investigating stable replication of SGR EqHV

Relatively high levels of replication are required to detect transient replication. There was therefore a possibility that EqHV was replicating at low levels and was consequently undetectable in a transient assay. Therefore, the SGRs described in 5.2.1 were subject to stable selection in the cell lines described in 5.2.2. Under selection cells which do not harbour the SGR, and thus a resistance reporter gene, die due to the toxic concentration of the compound used for selection. Conversely, cells which harbour a replicating SGR express an antibiotic resistance gene and therefore the compound is not toxic. The cells therefore survive and are maintained. However upon transfection of SGR EqHV, and selection with neomycin, all cells died and therefore did not harbour an EqHV SGR.

The cell lines described in 5.2.2 had different patterns of antibiotic resistance due to the stable expression of miR122, SEC14L2, and the V protein. Furthermore FHK cells were transformed with the SV40 large T antigen under neomycin selection and so are inherently resistant to neomycin, and are therefore not suitable for selecting Feo SGRs (Table 5.2 summarises the resistance profile of each cell line). Therefore blasticidin-S deaminase (BsR) and puromycin N-acetyl transferase (PAC) reporter genes were introduced into SGR EqHV by molecular cloning to enable SGR selection in each of the cell lines described in 5.2.2. However, following transfection with the SGR, and selection with the appropriate antibiotic for the SGR resistance gene, no cells survived indicating that SGR EqHV was not able to establish replication in any of the cell lines here described (a summary of the SGR and cell line combinations assayed for stable replication can be seen in Table 5.3).

These data indicate that SGR EqHV is not able to replicate in any of the cell lines tested in this study, and it is likely that this isolate of EqHV is not able to replicate in mammalian cell culture. (See Table 5.1 for a full list of the constructs used in this study).

pSGR	Reporter	5'UTR	Δ SLI	2180I	GNN	3'UTR
HEH	Feo	HCV	N/A	×	×	HCV
HEH GNN	Feo	HCV	N/A	×	✓	HCV
HEE	Feo	HCV	N/A	×	×	EqHV
HEE 2204I	Feo	HCV	N/A	✓	×	EqHV
HEE GNN	Feo	HCV	N/A	×	✓	EqHV
EEH	Feo	EqHV	×	×	×	EqHV
EEH Δ SLI	Feo	EqHV	✓	×	×	EqHV
EEH 2204I	Feo	EqHV	×	✓	×	EqHV
EEH GNN	Feo	EqHV	×	×	✓	EqHV
EqHV	Feo	EqHV	×	×	×	EqHV
EqHV Δ SLI	Feo	EqHV	✓	×	×	EqHV
EqHV 2180I	Feo	EqHV	×	✓	×	EqHV
EqHV GNN	Feo	EqHV	×	×	✓	EqHV
EqHV GFP-PAC	GFP-PAC	EqHV	×	×	×	EqHV
EqHV BsR	BsR	EqHV	×	×	×	EqHV

Table 5.1: pSGR EqHV, and HCV chimeric, constructs used in this study

This table outlines all of the SGR contracts assayed for replication in this study. Feo is the fused FLuc/neomycin phosphotransferase gene that produces firefly luciferase and confers resistance to neomycin. GFP-PAC is a dual GFP/puromycin N-acetyl-transferase (PAC) reporter which expresses GFP and confers resistance to puromycin. BsR is the blasticidin S resistance gene and confers resistance to blasticidin. The species origin of both UTRs is noted, and the ticks and crosses represent whether a mutation was present in each given SGR.

Cell line	Resistance	
	Neomycin	Puromycin
Huh7	x	x
Huh7 SEC14L2	x	✓
Huh7.5	x	x
Huh7.5 V protein	✓	x
FHK	✓	✓
FHK 122	✓	✓
FHK SEC14L2	✓	✓
293T	x	x
293T 122	x	✓
293T SEC14 L2	x	✓

Table 5.2: Antibiotic resistance profile of mammalian cells used in this study

A list of the cell lines used in this study and the antibiotics to which they are resistant.

	H-E-H	H-E-E	E-E-H	EqHV	EqHV BsR	EqHV GFP-PAC
Huh7	✓	✓	✓	✓	✓	✓
Huh7 SEC14L2	✓	✓	✓	✓	✓	N/A
Huh7.5	✓	✓	✓	✓	✓	✓
Huh7.5 V protein	N/A	N/A	N/A	N/A	✓	✓
FHK	N/A	N/A	N/A	N/A	✓	✓
FHK 122	N/A	N/A	N/A	N/A	✓	N/A
FHK SEC14L2	N/A	N/A	N/A	N/A	✓	N/A
293T	✓	✓	✓	✓	✓	✓
293T 122	✓	✓	✓	✓	✓	N/A
293T SEC14 L2	✓	✓	✓	✓	✓	N/A

Table 5.3: Stable selection of EqHV, and chimeric, SGRs in a range of cell lines

A table of the cell lines and SGRs used in this study indicating whether they were assayed to support stable SGR replication. N/A indicates that the cell line/SGR combination was not possible due to the antibiotic resistance of the cell line (see Table 5.2)

5.2.4 A second generation, consensus sequence, EqHV SGR

A cell culture replication system for HCV was not developed until the construction of a consensus clone Con-1b (Lohmann et al., 1999). Scheel *et al* constructed a consensus clone of EqHV, NZPI, and provided the SGR (pSGR NZPI) for use in this study; they were, however, unable to achieve stable replication with this SGR in a range of cell lines (Huh-7.5, E.Derm, E.Derm/miR-122, MDBK, MDBK/miR-122, MDCK/miR-122, PK-15, and PK15/miR-122) (Scheel et al., 2015).

SGR NZPI was constructed to be stably selected using the neomycin phosphotransferase gene under control from the EqHV 5'UTR. Therefore, for use in transient replication assays, an alternative reporter gene needed to be introduced. Firefly luciferase is widely used as the reporter gene for HCV SGRs and recent discoveries have demonstrated a 100 fold increase in detection using this reporter when the codon sequence is optimised to be CpG and UpA low (Atkinson et al., 2014). This reporter (here termed CpG/UpA low FLuc) was therefore selected for this study and introduced into SGR NZPI, and GNN, by molecular cloning, to produce a construct from which transient replication could be assayed (pSGR NZCI). Further to this the S2180I mutation, and the Δ SLI 5'UTR described in 3.2.1, were also introduced into pSGR NZCI in an attempt to achieve detectable transient replication (a schematic representation of SGR NZPI/NZCI can be seen in Figure 5.8, and a summary of the SGRs in Table 5.4).

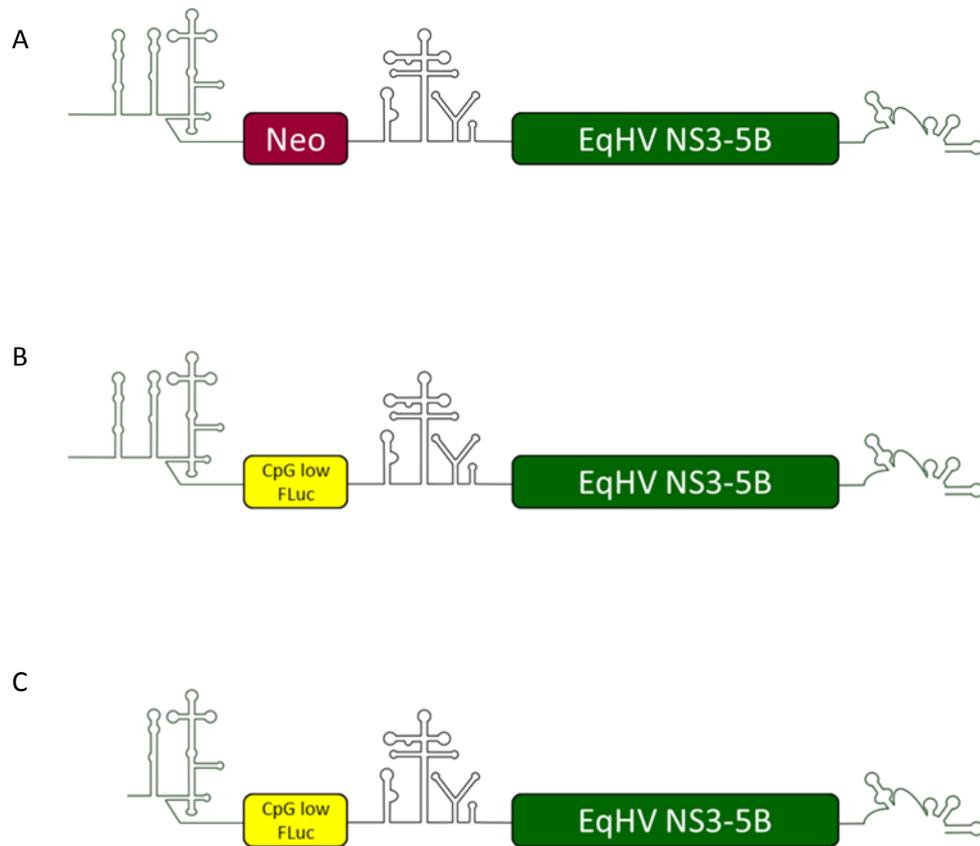


Figure 5.8: A schematic representation of SGR NZPI and SGR NZCI

A schematic representation of A) SGR NZPI B) SGR NZCI and C) SGR NZCI Δ SLI

pSGR	Reporter	Δ SLI	2180I	GNN
NZPI	Neo	×	×	×
NZPI GNN	Neo	×	×	✓
NZCI	CpG/UpA low FLuc	×	×	×
NZCI Δ SLI	CpG/UpA low FLuc	✓	×	×
NZCI 2204I	CpG/UpA low FLuc	×	✓	×
NZCI GNN	CpG/UpA low FLuc	×	×	✓

Table 5.4: A summary of pSGR NZPI/NZCI constructs

A summary of the SGR NZPI, and derivative, constructs used in this study denoting the reporter gene and mutation combinations constructed.

5.2.5 Replication assays of SGR NZPI and NZCI

To assess the replicative ability of SGR NZCI in a transient assay the cell lines as described in 5.2.2 were transfected with SGR NZCI, Δ SLI, 2180I, and GNN, and assayed for firefly luciferase activity up to 72 h.p.t. As for SGR EqHV, no replication could be detected over the GNN replication deficient control up to 72 h.p.e in any of the Huh7, or 7.5, cells or their derivatives (+SEC14L2/V protein) (Figure 5.9), FHK cells or their derivatives (+SEC14L2/miR122) (Figure 5.10), or 293T cells or their derivatives (+SEC14L2/miR122) (Figure 5.11). These data conclusively demonstrate that SGR NZCI is not able to replicate to a detectable level in a transient assay in Huh7, Huh7.5, FHK, or 293T cells, even upon introduction of a predicted culture adaptive mutation into NS5A, or in the presence of factors known to support or enhance HCV replication.

SGR NZPI was also assayed for stable replication in the cell lines described above (excluding the already neomycin resistant FHK cell lines, and Huh7.5 V protein cells). In agreement with the data presented by Scheel *et al* no colonies formed following 2-3 weeks under neomycin selection following transfection of SGR NZPI (Scheel *et al.*, 2015). These data conclusively demonstrate that SGR NZPI is unable to establish replication in Huh7, Huh7.5, or 293T cells even in the presence of factors known to support HCV replication.

Taken together these data indicate that currently available isolates, and consensus clones, of EqHV are not able to establish replication in cell culture. Host factors known to support HCV infection do not enable EqHV replication, as is also the case for a viral factor shown to enhance HCV replication. Further work to establish a replicative EqHV SGR is needed, however this may well be dependent on the isolation of new sequences, or the development of new, more physiologically relevant, cell lines.

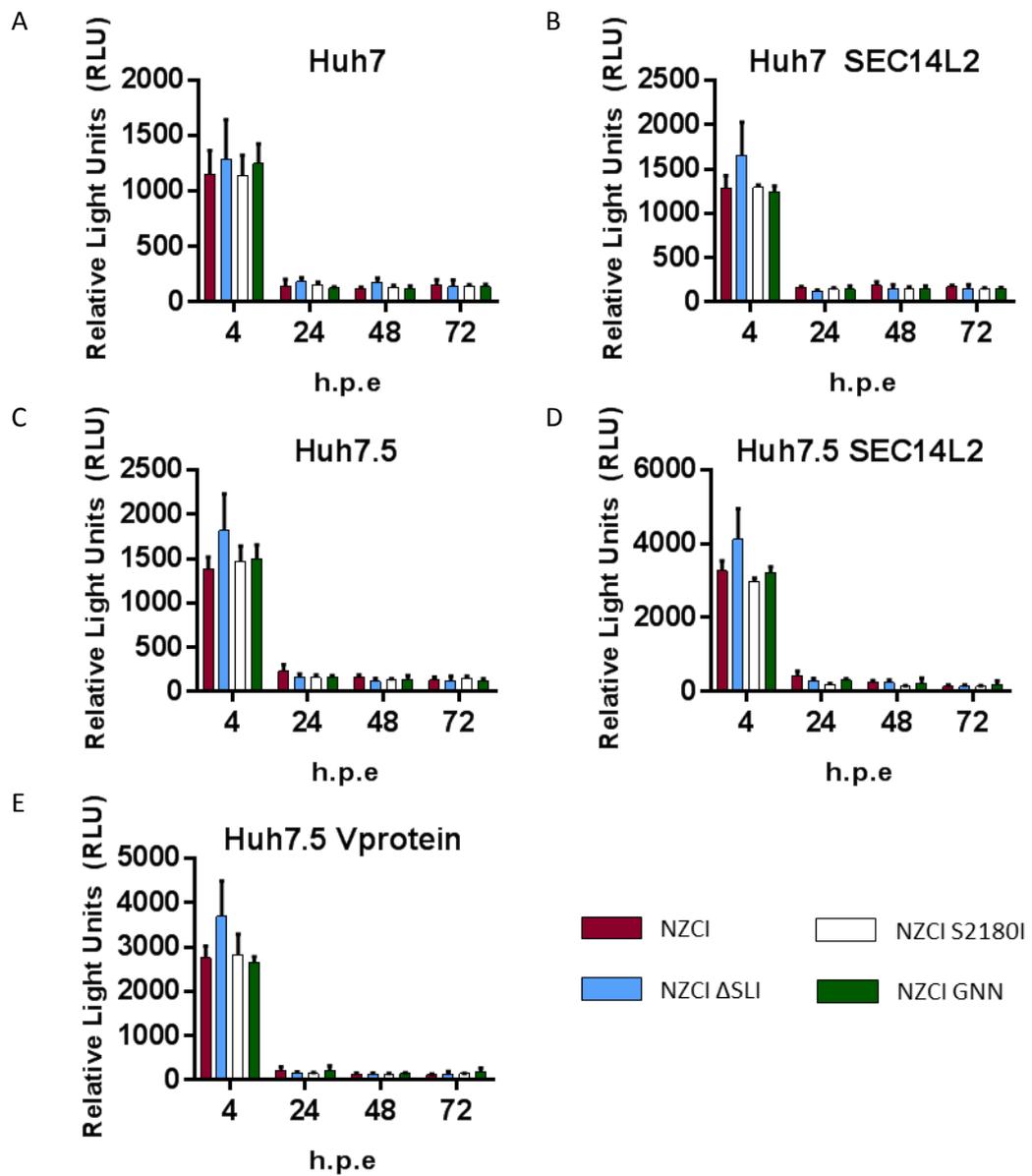


Figure 5.9: Transient replication of NZCI in human hepatoma cells

SGR NZCI, S2180, and GNN, RNA was transfected into A) Huh7 B) Huh7 SEC14L2 C) Huh7.5 D) Huh7.5 SEC14L2 and E) Huh7.5 V protein cells and assayed for firefly luciferase activity at 4, 24, 48, and 72 h.p.t.

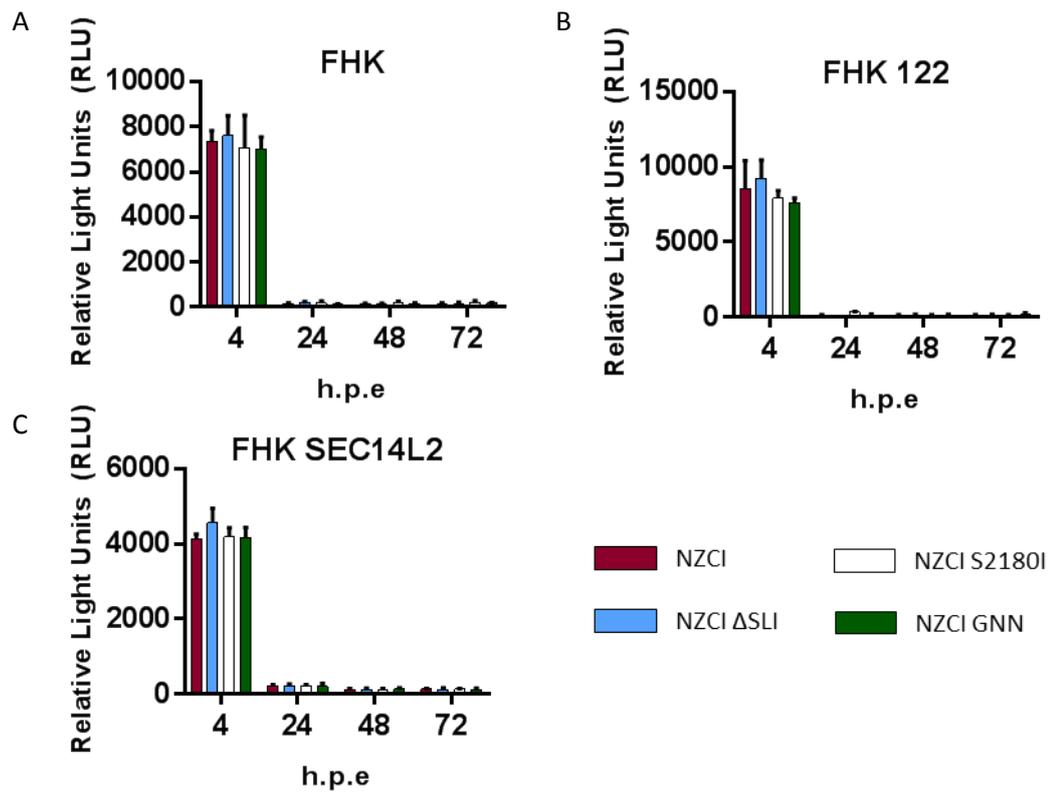


Figure 5.10: Transient replication of NZCI in equine cells

SGR NZCI, S2180, and GNN, RNA was transfected into A) FHK B) FHK 122 and C) FHK SEC14L2 cells and assayed for firefly luciferase activity at 4, 24, 48, and 72 h.p.t.

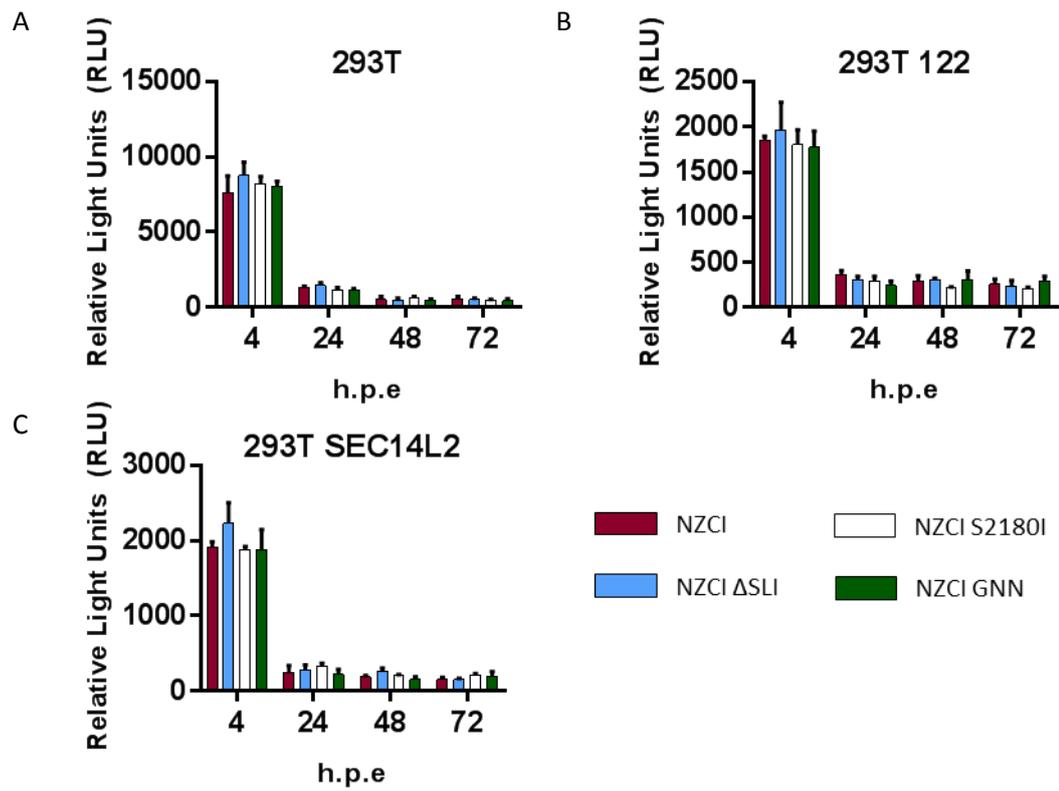


Figure 5.11: Transient replication of NZCI in human kidney cells

SGR NZCI, S2180, and GNN, RNA was transfected into A) 293T B) 293T 122 and C) 293T SEC14L2 cells and assayed for firefly luciferase activity at 4, 24, 48, and 72 h.p.t.

5.3 Discussion:

5.3.1 Construction of an EqHV SGR from the EF369_11J isolate

The construction of an EqHV SGR involved replacing HCV sequences in an SGR with the equivalent EqHV ones. Two well characterised mutations, S2204I and GDD→GNN were also introduced into the EqHV polyprotein.

HCV replication in cell cultured was not achieved until 10 years following its discovery (Lohmann et al., 1999), and even following this remarkable achievement efficient replication was not achieved until the discovery of culture adaptive mutations (Blight et al., 2000). It was therefore hypothesised that culture adaptive mutations may also be essential to establish efficient *in vitro* replication of EqHV. The HCV culture adaptive mutation 2204I significantly enhances HCV replication (Blight et al., 2000) and has been instrumental in the establishment of many HCV cell culture systems (Chung et al., 2017). S2204 is conserved between HCV and EqHV at a.a. position 2180 in EqHV (amino acid numbering according to NZPI consensus sequence polyprotein sequence GenBank accession number KP325401) and so an equivalent S2204I mutation was introduced into EqHV NS5A (S2180I) in an attempt to enhance replication in cell culture. However, this mutation impairs replication of JFH-1, which requires no culture adaptive mutations, so both the wild type and S2180I were used for all replication assays in case a similar effect was seen for EqHV. Finally, the EqHV NS5B RdRP contains the conserved GDD motif, which is well characterised as the active site of many viral RdRPs. The well characterised mutation GDD→GNN mutation was introduced into EqHV NS5B for use as a replication deficient control.

Two different 5'UTRs were used in this study, the full length 5'UTR, and a deletion mutation ΔSLI, both of which have been previously reported to exhibit IRES activity (Stewart et al., 2013). The inclusion of the full length 5'UTR was an obvious necessity in the construction of a full EqHV SGR. The ΔSLI mutant 5'UTR was included in this study based on the functional (Chapter 3), and structural data (Chapter 4), presented here. The functional analysis of the EqHV 5'UTR, in agreement with published data (Stewart et al., 2013), demonstrated that the EqHV SLI played no role in translation. It was therefore hypothesised that SLI was involved in replication of the EqHV genome and consequently may be essential for replication. The structural data obtained for ΔSLI (Figure 4.4) showed that the ΔSLI 5'UTR appeared to be structurally similar to the HCV 5'UTR (Figure 3.1B); with SLIA as the first structural domain of the 5'UTR, mimicking HCV SLI. Whilst these SLs did not appear to share any sequence homology, SLIA more closely matches

with HCV miR122 seed site 1, there was a possibility that the structure of this region, rather than sequence, was conserved between HCV and EqHV, and therefore that EqHV SLIA was a determinant of EqHV replication. In this scenario it was hypothesised that the extended, stable, structure of SLI may be a regulatory element of EqHV genome replication. If EqHV replication were achieved then the ability of the Δ SLI mutant to replicate would have provided further information into the role of the 5'UTR structural domains in the EqHV lifecycle.

Finally, the full 3'UTR from the EqHV EF369_11J isolate was synthesised as a gene block to facilitate molecular cloning of this region into the SGR. Only the full length 3'UTR of the sequence for this isolate was used.

The EqHV NS3-5B codon region was initially introduced into an HCV SGR through replacement of the HCV NS3-5B coding region, by Cheryl Walter and Hazel Stewart (University of Leeds) to create pSGR H-E-H. This construct was used as a master construct into which the NS5A, and NS5B, mutations were introduced, and from which the full EqHV SGR would be created. The 5' and 3' UTRs were sequentially introduced into the pSGR H-E-H backbone by molecular cloning to create pSGR EqHV. This process also created intermediate, chimeric, HCV/EqHV SGRs with mismatched UTRs termed pSGR E-E-H and pSGR H-E-E (Table 5.1 summarises the pSGR constructs created in this study, and the NS3-5B, and 5'UTR mutations each included).

5.3.2 pSGR NZPI and NZCI

Replication of HCV in cell culture was first achieved through the construction of a consensus clone Con-1b (Lohmann et al., 1999). Scheel *et al.* constructed an EqHV consensus sequence (NZPI) and constructed a full length molecular clone, and SGR, based off of this sequence (Scheel et al., 2015). pSGR NZPI was provided for use in this study by C. Rice (Rockefeller University). pSGR NZPI, and pSGR NZPI GNN, were both provided with the Neo reporter gene and therefore were only suitable for use in stable replication assays. Recent studies have demonstrated that a 100 fold increase in replication of HCV can be achieved through modification of the FLuc gene to produce a CpG and UpA low variant (Atkinson et al., 2014). To use SGR NZPI in transient replication assays it was necessary to introduce a reporter gene suitable for detection in this system into SGR NZPI. Due to the enhancement of detection seen for HCV using the CpG/UpA low modified FLuc this was chosen for use in this study as it is readily detectable in transient replication assays, and represented the best option to maximise the possibility of achieving and detecting replication of EqHV in cell culture. CpG/UpA low FLuc was therefore introduced into

pSGR NZPI to create a new construct termed pSGR NZCI. Furthermore, the NZPI consensus sequence contained the conserved serine at position 2180 and so the S2180I mutation (described in 5.3.1) was introduced into pSGR NZCI.

This SGR represented two major improvements over SGR EqHV (described in 5.3.1): SGR EqHV was created from a single EqHV isolate and therefore may not be representative of the viral population in nature, and could potentially harbour detrimental mutations that are only represented in this isolate which prevent EqHV replication in cell culture. Secondly, SGR NZPI includes the full length EqHV 3'UTR. The isolate used for construction of SGR EqHV had an unusually short 3'UTR. At the time of creation of SGR EqHV it was not known whether this truly represented the EqHV 3'UTR or whether it was an incomplete sequence. The work conducted by Scheel *et al.* showed that it was in fact an incomplete 3'UTR and their work demonstrated that the EqHV 3'UTR was the longest hepacivirus 3'UTR discovered to date (Appendices Figure 8.12-8.14). The full EqHV 3'UTR is 328 nucleotides long with a conserved 3'X region and an extremely long poly U tract spanning 96 nucleotides (Scheel et al., 2015). The original EqHV 3'UTR isolate used in the construction of the "first generation" SGRs matches well with the 3' portion of the HCV 3'UTR immediately downstream of the polyU region. However the EqHV polyA region did not align well with any region of the HCV 3'UTR and the 3' X region appears to be completely absent in this EqHV isolate. On the other hand, whilst the NZPI 3'UTR does exhibit an extended polyU tract, it also appears to align quite well throughout the 3' portion of the HCV 3'UTR, covering the polyU region and 3' X tail. The variable region of NZPI has limited conservation with HCV and is extended, however conservation in this region is low even between different HCV genotypes. There are also stark differences between the "original" EqHV isolate and the NZPI 3'UTRs. Whilst some regions align well the original sequence is missing the polyU tract and there are large regions where there are apparent gaps in the original sequence in comparison to NZPI.

The NZPI 3'UTR demonstrates much higher levels of conservation with the HCV 3'UTR indicating that this 3'UTR sequence is indeed the full 3'UTR. The original EqHV 3'UTR isolate demonstrates incomplete alignment with the HCV 3'UTR and upon alignment with the NZPI sequence it is clear that some portions of this 3'UTR sequence are missing and it can be said that this 3'UTR isolate is incomplete. Whilst the NZPI derived SGRs, even with the complete 3'UTR, sequence were also not able to replicate. The absence of a complete 3'UTR sequence for the first generation EqHV SGRs may have played a crucial role in their inability to replicate.

5.3.3 Cell lines for the assay of EqHV SGR replication

EqHV is an equine virus which exhibits liver tropism (Ramsay et al., 2015; Pfaender et al., 2017; Scheel et al., 2015), therefore the ideal cell line in which to assay replication of this virus would be a liver derived equine cell line. Furthermore, if EqHV were to be established as a model system with which to study HCV, the cell line in which it is assayed would ideally be widely available, easy to culture and maintain, and support HCV replication for use in comparative analysis. It was also hypothesised that cell lines capable of supporting HCV replication, and equine derived cell lines, may be more likely to support EqHV replication than others. Unfortunately, to our knowledge, there are currently no equine liver cell lines available. Human foetal liver cultures (HFLCs) have been demonstrated to support replication of HCV (Andrus et al., 2011) and Scheel *et al.* established a similar equine foetal liver culture (EFLCs), however they were unable to establish EqHV replication using this system (Scheel et al., 2015). Given this data, EFLCs were not used in this study.

Due to the lack of equine liver cell lines alternative cell culture systems needed to be used in the attempt to establish EqHV replication. Huh7, and Huh7.5, cells have long been established as the cell line of choice for HCV research, with Huh7 cells the first in which replication was established (Lohmann et al., 1999). Huh7.5 cells are a derivative of the Huh7 cell line which were cured of an HCV SGR using IFN α and have been shown to facilitate higher levels of replication (Blight et al., 2002). EqHV is a liver tropic virus and as such, in the absence of an equine derived cell line it was believed the human hepatoma Huh7, and Huh7.5, cell lines provided a viable alternative to potentially establish EqHV replication.

Whilst Huh7 cells are a liver derived line and therefore provide the liver specific factors required for HCV replication, and potentially for EqHV replication, being a human derived cell line they may lack horse specific factors that enable EqHV replication. Therefore a horse cell line was sourced. Foetal horse kidney (FHK) cells were originally established to study the replication of equine herpesviruses, but have also been demonstrated to support the replication of a range of equine viruses, and remain viable for over 100 passages (Andoh et al., 2009). These characteristics identified FHK cells as the next best alternative to an equine liver cell line and were provided for use in this study Dr Hiroshi Sentsui (Nihon University, Japan) (Maeda et al., 2007). The final cell line selected for use was the human kidney cell line 293T. This cell line has

previously been demonstrated to support replication of HCV, upon the expression of miR122, is widely used, and easy to culture and maintain.

The cell lines described above all shared the ideal characteristic of being easy to culture and maintain, and are widely available. Huh7, Huh7.5, and 293T cells have all been demonstrated to support HCV replication, and whilst FHK cells have not been demonstrated to support HCV replication they are equine derived. Therefore all of the cells selected for this study had two of the three characteristics believed to constitute an ideal cell line to study EqHV replication.

5.3.4 Alternative factors to enable EqHV replication in cell culture

As no “perfect” cell culture system was available to study EqHV it was hypothesised that the expression of protein, or RNA, shown to enable or enhance HCV replication may be required to establish EqHV replication in cell culture.

miR122 is a liver specific microRNA which has been demonstrated to function in HCV translation and replication, and the stability of the genome. The data presented in Chapter 3, in agreement with published data (Scheel et al., 2015), showed that the presence of miR122 enhances translation from the EqHV IRES and further demonstrated that this effect was dependent on a miR122 seed site within the 5'UTR. Due to the large scale conservation between an array of structural, and functional, aspects of EqHV and HCV translation described here, and the array of other conserved features between these two closely related viruses (Tanaka et al., 2014; Parera et al., 2012; Stewart et al., 2016; Pfaender et al., 2017) it was considered likely that miR122 would play a role in the EqHV lifecycle beyond translation, and may be a key determinant of replication. 293T and FHK cells, unlike the human hepatoma cell lines, do not naturally express miR122, therefore retroviral transduction was used to establish stable expression of miR122 in both 293T and FHK cells. It was hypothesised that this may enable replication of EqHV in these cells, as has previously been demonstrated for HCV in cell lines which do not endogenously express miR122.

SEC14L2 is a human protein that has recently been described to enable pan genotypic replication of HCV in cell culture, even supporting infection and replication from infected patient sera (Saeed et al., 2015). SEC14L2 is ubiquitous in human tissues and is a member of the family of cytosolic lipid-binding proteins however it cannot be readily detected in human derived cell lines, including Huh7 and Huh7.5 cells (Saeed et al., 2015). The SEC14L2 expression levels of

293T and FHK cells were not tested by Saeed *et al.* or in this study, however given the lack of expression of this protein in a range of human cell lines it was probable that SEC14L2 levels were low, or non-existent, in these lines. Therefore Huh7, Huh7.5, FHK, and 293T cells were all engineered, using a lentiviral vector, to stably express SEC14L2. EqHV displays a divergence from HCV at ~50% at the nucleotide level, with up to 65% conservation in the a.a sequence of NS3 and NS5B. HCV genotypes and subtypes are 30%–35% divergent. Although EqHV diverges from HCV to a greater extent than HCV genotypes diverge from one and other, as SEC14L2 enables pan-genotypic HCV replication in Huh7.5 cells it was hypothesised that a similar effect may be seen on EqHV replication. Expression of this protein may enable cells to support replication of the EqHV SGRs.

The PIV5 V protein is a multifunctional protein which binds STAT1, blocking downstream IFN signalling, and interacts with MDA-5, a mediator of the virally induced IFN response (Yang *et al.*, 2015). Upon transfection of HCV RNA into HFLCs stably expressing the V protein HCV replication was enhanced. It was hypothesised that expression of this viral protein may also be able to enhance replication of EqHV, or alternatively induce a cellular environment which is permissive to EqHV replication. Huh7.5 cells which stably expressed the PIV5 V protein were provided by S. Griffin (University of Leeds) for use in this study.

The addition of these cellular and viral factors which have been demonstrated to enable or enhance HCV replication, in addition to the parental cell lines discussed in 5.3.4, provided a comprehensive panel of cell lines in which to assess the replicative ability of EqHV.

5.3.5 Alternative resistance genes to study stable replication in antibiotic resistant cell lines

The construction of the panel of cell lines discussed in 5.3.4 required the stable expression of viral and host factors, and therefore required these cells to be under selection with either puromycin (miR122 and SEC14L2) or neomycin (V protein). FHK cells are also neomycin resistant due to the way in which they were created (Maeda *et al.*, 2007) (Table 5.2 summarises the antibiotic resistance profile of the cell lines). Therefore the Feo and Neo reporter genes, which confer resistance to neomycin, would not be sufficient to assay each cell line for stable replication.

Two different reporter genes were selected, PAC encoding puromycin N-acetyl transferase, and BsR encoding blasticidin-S deaminase, conferring resistance to puromycin and blasticidin. PAC

and BsR have both previously been utilised in the selection of cells harbouring stable replication of HCV SGRs, and provided resistance to antibiotics which were not included in any, or a subset, of the cell lines used in this study (Table 5.2). Therefore these genes were introduced into SGR EqHV in order to assay it for stable replication in each cell line (Table 5.3 summarises the cell line/SGR combinations assayed for stable replication).

5.3.6 Currently available EqHV SGRs are unable to replicate in cell culture

It has here been conclusively demonstrated that SGRs generated from either the EF369_11J isolate (GenBank accession no. JX948116.1) or NZPI consensus clone (GenBank accession no. KP325401) are not able to achieve detectable replication in any of the cell lines described here.

A rational design of SGR EqHV was undertaken, with the well-known HCV culture adaptive mutation S2204I (S2180I in EqHV), and a Δ SLI 5'UTR included in an attempt to facilitate SGR EqHV replication (discussed in 5.3.1). Chimeric SGRs were created as intermediates in the creation of SGR EqHV; however these were also unable to replicate. In order to assess the ability of SGR EqHV to establish stable replication alternative resistance genes (BsR and PAC), in addition to Feo/Neo, were introduced into SGR EqHV to allow selection in all cell lines. Each cell line was transfected with the relevant SGR EqHV (Table 5.3) and selected with the appropriate antibiotic, however no colonies were formed following selection and it can therefore be said that SGR EqHV was not replicating, or replicating at such a low level as to not confer resistance.

The lack of replication exhibited by SGR EqHV was most likely due to the incompleteness of the 3'UTR, an essential feature of the genome for HCV replication (Yi and Lemon, 2003). However Scheel *et al.* were successful in sequencing the entire EqHV 3'UTR, and included this in the construction of an SGR termed NZPI (Scheel *et al.*, 2015). The SGR NZPI construct was provided for use in this study by C. Rice (Rockefeller University). A reporter gene amenable to transient replication assays (CpG low FLuc) was introduced into this construct, as was the S2180I culture adaptive mutation and the Δ SLI 5'UTR. However, even with the full length 3'UTR, and consensus sequence, this construct was not able to achieve transient replication to a detectable level in any of the cell lines tested in this study. Finally, SGR NZPI was assayed for stable replication, however this construct was only assayed with the Neo reporter gene and so the FHK cell lines, and Huh7.5 V protein cells, could not be assayed for replication. As for SGR EqHV, no colonies could be detected following selection.

This chapter therefore describes a rational and comprehensive approach to the establishment of an EqHV SGR capable of replication in mammalian cell culture. The data provided here, coupled with published data, indicate that the currently available EqHV SGRs are not able to achieve replication in current cell culture systems, even with the addition of viral and host factors predicted to enable or increase replication. Factors that may be preventing the establishment of *in vitro* replication of EqHV are discussed in 5.3.8.

5.3.7 Roadblocks to establishing a cell culture system to study EqHV replication

Taken together with published data the data here clearly demonstrate that there is more work to be done if an EqHV SGR system capable of replicating in mammalian cell culture is to be established. Some major obstacles face the scientific community in achieving this goal.

There are two potentially key factors which are currently inhibiting EqHV replication *in vitro*. Firstly is the lack of a physiologically relevant cell line in which to assay EqHV replication. It is no coincidence that HCV replication was first achieved in human hepatoma cells, and that these cells are still widely used to this day. EqHV, like HCV, appears to have an extremely narrow tropism (Pfaender et al., 2017; Scheel et al., 2015; Ramsay et al., 2015) and it is predicted that if EqHV replication is to be achieved it will be done so in equine liver derived cell lines. However, such a cell line is not available for use. Therefore the main priority must be the establishment of an equine derived liver cell that is easy to maintain and culture in the laboratory, if EqHV replication is to be studied.

Secondly is the potential sequence variation within the EqHV genome. Although EqHV isolate numbers are growing, a search for “equine hepacivirus” in pubmed returns 369 sequences, whilst “hepatitis C virus” returns 1,020,438. HCV genotypes are known to have differential abilities to replicate in cell culture, and patient isolates do not readily replicate. The increasing pool of EqHV sequences has already revealed the putative existence of viral subtypes (Pronost et al., 2016). Increasing this pool may result in the identification and classification of further EqHV subtypes; and potentially different genotypes. If they mimic the biology of HCV they may prove to be more amenable to replication *in vitro*. Furthermore a consensus clone was required to establish the first HCV SGR, and whilst this has been attempted for EqHV current sequences have exhibited minimal divergence, with ~99% identity in some studies (Kapoor et al., 2011).

Increasing the number of EqHV sequences with which to establish a consensus will surely make this approach more powerful.

EqHV does not readily replicate in mammalian cell culture, and the appropriate reagents with which to study it do not currently exist. There appears to be a great deal of work that needs to be carried out before this goal will be achieved.

Chapter 6 Conclusions and future perspectives

HCV is a major global health concern, however no model system exists to study this important pathogen. Recent advances in sequencing technology have led to the discovery of a variety of animal hepaciviruses in bovines, bats, dogs, equines, rodents, an old world primate, and even the graceful catshark (Kapoor et al., 2011; Baechlein et al., 2015; Corman et al., 2015; Drexler et al., 2013; Kapoor et al., 2013; Shi et al., 2015; Quan et al., 2012; Burbelo et al., 2012). The first of these to be identified, and the closest phylogenetic relation to HCV, is equine hepacivirus (EqHV), previously known as canine hepacivirus (CHV) or non-primate hepacivirus (NPHV). This study set out to investigate the molecular biology of EqHV; focussing on 5'UTR IRES function, how RNA secondary structure is linked to this function, how the 5'UTR interacts with the host cell translational machinery, and finally to establish an EqHV SGR capable of replicating in mammalian cell culture.

Following determination of the EqHV 5'UTR secondary structure, deletions of 5'UTR structural regions (Δ SLI, Δ SLI+II, Δ SLIII, Δ SLIIIb, and Δ SLIIId) were made to investigate how the structure of the 5'UTR was linked to IRES function. These deletion mutations were assayed in the context of an RLuc/FLuc bicistronic reporter construct, and an SGR, to assess IRES activity. This investigation revealed that SLI plays no role in IRES activity and that SLII plays a non-essential role. SLIII, SLIIIb and SLIIId were found to be essential. The observations made on the role of these structural regions in EqHV were similar to those that have been previously published for HCV, and began to provide evidence for the existence of a conserved mechanism of translation initiation from these two closely related viruses.

This study has provided, to our knowledge, the first experimental description of the secondary structure of the EqHV 5'UTR. The EqHV 5'UTR contains three major structural domains: SLI, SLII, and SLIII(a, b, c, d, e, and f). Evidence for the existence of a pseudoknot was also observed, but this structure was not experimentally confirmed. The secondary structure of the EqHV 5'UTR exhibited three major differences from HCV: a large and extended SLI, an intermediate SLIIA, and does not contain SLIV. However, SLII, and SLIII are extremely well structurally conserved and exhibit localised sequence conservation across the apical loops of SLIIIa, SLIIIc, and SLIIId, and the predicted pseudoknot. Secondary structure analysis also confirmed that the effects of structural deletions in the 5'UTR were specific to the deleted region, and thus confirmed the hypothesis that arose from the functional analysis regarding the roles of these structures. The conservation of structure, coupled with the functional observations, provided further evidence for a conserved mechanism of translation initiation.

Following the observations made regarding how the structure of the EqHV 5'UTR links to its function, a more detailed investigation of the roles of the apical loops of SLIIIb and SLIIId were carried out. This investigation revealed that the sequence of the SLIIIb apical loop plays a role in IRES activity, but that the structure of SLIIIb appears to be more important. To investigate the mechanisms influencing the diminished ability of sequence mutants to initiate translation SHAPE footprinting analysis was carried out. This demonstrated that eIF3 specifically interacts with SLIIIb, and that mutation of the apical loop abolishes this interaction. However the ability of the SLIIIb apical loop mutant to initiate translation in the absence of an interaction with eIF3 raises some interesting questions regarding the role of eIF3 in EqHV IRES mediated translation. There may be structural features that also mediate an IRES-eIF3 interaction. The differences observed between the sequence, and structural, mutations of SLIIIb indicate that the structure of SLIIIb, and potentially the conservation of the SLIIIa,b,c junction, are more important than the apical loop in translation.

A double mutation in the apical loop of SLIIId abolished translation. SHAPE footprinting revealed that the apical loop interacts with the 40S ribosomal subunit, and that a double mutation in the GGG motif disrupts this. The GGG motif in the apical loop of EqHV SLIIId plays an essential role in translation through its interaction with the 40S ribosomal subunit, presumably acting to recruit 40s and correctly position it on the AUG start codon. The observations made regarding the roles of the apical loops of SLIIIb and SLIIId provided further evidence toward the hypothesis that there exists a conserved mechanism of translation initiation between EqHV and HCV.

Further evidence was provided using the EqHV 5'UTR as bait in a pull-down assay, and identifying the interacting partners by TMT MS. This analysis revealed that the EqHV 5'UTR interacts with all of the eIFs required for HCV translation initiation, as well as the 40S and 60S ribosomal subunits required for the assembly of the 80S ribosome. Taken together these data provide evidence for a conserved mechanism of translation between EqHV and HCV, and potentially indicates a conserved mechanism of hepaciviral translation.

Finally, it was set out to establish an EqHV SGR capable of replicating in mammalian cell culture. SGR EqHV was constructed using sequences derived from the first EqHV isolate. SGR NZPI was provided for use by C. Rice (Rockefeller University), and a CpG/UpA low FLuc introduced into this construct to create SGR NZCI. A range of cell lines expressing cellular and viral factors known to enable, or enhance, HCV replication were assayed. Furthermore the well-known S2204I

(S2180I in EqHV) culture adaptive mutation was introduced into EqHV NS5A in an attempt to achieve *in vitro* replication. Different transient reporter, and antibiotic resistance genes, were also utilised to maximise the possibility of achieving EqHV replication. Neither transient nor stable replication was observed in this study. The currently available EqHV sequences, and cell lines in which to assay them, may therefore lack some fundamental factors or characteristics that are required for *in vitro* EqHV replication.

The natural progression of this study would be to further investigate EqHV translation initiation. The exact role of SLIIIb, and how its interaction with eIF3 affects translation, was not determined in this study. Further mutational analysis on the stem of SLIIIb, and the SLIIIa,b,c junction, in combination with SHAPE footprinting, may well provide the evidence needed to determine this. A more detailed mutational analysis of EqHV SLII would also be required to understand the role that this domain plays in EqHV translation initiation, and whether this is determined by sequence or structure. *In vitro* translation experiments using purified initiation factors to investigate the individual steps that occur towards initiation from the EqHV IRES would also be needed to confirm the exact method of translation initiation. Sequential addition, or depletion, of initiation factors, and sucrose density centrifugation, could be utilised to investigate the order of assembly of eIFs and determine the minimal requirement of eIFs for EqHV IRES mediated translation.

There is also a large body of work that needs to be carried out if replication of EqHV is to be achieved *in vitro*. The establishment of an equine derived liver cell line which is easy to culture and maintain is an essential next step towards *in vitro* EqHV replication. In the absence of full EqHV SGR replication, however, parts of the viral lifecycle may still be amenable to study by constructing chimeric HCV/EqHV SGRs or viruses. Furthermore SGRs constructed from newly isolated EqHV sequences from different genotypes, if they are discovered, and updated consensus sequences, will need to be constructed. However this work will depend upon the continued isolation of EqHV sequences by the scientific community. Furthermore the construction of SGRs from the other hepaciviral species which have been, and are still being, discovered may provide a replicative model system to study HCV, as well as providing a means by which to establish whether there is indeed a conserved mechanism of hepaciviral translation.

A recent report has described a mouse model for acute, and chronic, infection with NrHV (hepacivirus G) (Billerbeck et al., 2017) and provides a shining example of the potential

usefulness of the newly discovered animal hepaciviruses. However, given the genetic heterogeneity within the genus, and the observed differences that have been reported with regards to the molecular biology and pathogenicity of these viruses, further investigation is still required. Establishing a complete picture with regards to the molecular biology, lifecycle, and pathogenicity of this model virus is absolutely crucial. If the animal hepaciviruses demonstrate fundamental differences from HCV these will need to be known, and taken into account, when using them as models.

Furthermore, the 5'UTR of NrHV has not yet been described, and the sequence has not yet been determined. Isolation of this sequence, and the characterisation of structural features and their functions as described here, provides an interesting avenue of investigation to begin to understand the molecular biology of this model virus. In addition to this, rational design of an NrHV SGR could potentially provide an *in vitro* model to study the viral lifecycle. A number of mouse derived liver cell lines are available and listed on the American Type Culture Collection (ATCC) website for purchase, potentially providing the ideal cell line in which to assay NrHV replication. Construction of a replicative SGR would provide an additional method with which to study this virus; current HCV DAAs could be assayed for effect against NrHV in cell culture, reducing the need for animal test subjects. Conversely, if currently available DAAs could be demonstrated to have an effect against NrHV, an SGR could provide an excellent model with which to test putative new HCV DAAs without the need for animals. Therefore rational construction of an NrHV SGR, as has been described here for EqHV, could prove an exciting and valuable tool for the future of HCV research.

Chapter 7 References

- Agnello, V., Abel, G., Elfahal, M., Knight, G.B. and Zhang, Q.X. 1999. Hepatitis C virus and other flaviviridae viruses enter cells via low density lipoprotein receptor. *Proceedings of the National Academy of Sciences of the United States of America*. [Online]. **96**(22),pp.12766–12771.
- Ali, I.K., McKendrick, L., Morley, S.J. and Jackson, R.J. 2001. Activity of the hepatitis A virus IRES requires association between the cap-binding translation initiation factor (eIF4E) and eIF4G. *Journal of virology*. **75**(17),pp.7854–63.
- Aminev, A.G., Amineva, S.P. and Palmenberg, A.C. 2003. Encephalomyocarditis viral protein 2A localizes to nucleoli and inhibits cap-dependent mRNA translation. *Virus Research*. [Online]. **95**(1–2),pp.45–57.
- Andersen, C.B.F. 2006. Structure of the Exon Junction Core Complex with a Trapped DEAD-Box ATPase Bound to RNA. *Science*. **313**(5795),pp.1968–1972.
- Andino, R., Rieckhof, G.E., Achacoso, P.L. and Baltimore, D. 1993. Poliovirus RNA synthesis utilizes an RNP complex formed around the 5'-end of viral RNA. *The EMBO Journal*. **12**(9),pp.3587–98.
- Andino, R., Rieckhof, G.E. and Baltimore, D. 1990. A functional ribonucleoprotein complex forms around the 5' end of poliovirus RNA. *Cell*. [Online]. **63**(2),pp.369–380.
- Andoh, K., Kai, K., Matsumura, T. and Maeda, K. 2009. Further Development of an Equine Cell Line that can be Propagated over 100 Times. *Journal of equine science*. **20**(2),pp.11–4.
- Andrus, L., Marukian, S., Jones, C.T., Catanese, M.T., Sheahan, T.P., Schoggins, J.W., Barry, W.T., Dustin, L.B., Trehan, K., Ploss, A., Bhatia, S.N. and Rice, C.M. 2011. Expression of paramyxovirus V proteins promotes replication and spread of hepatitis C virus in cultures of primary human fetal liver cells. *Hepatology*. [Online]. **54**(6),pp.1901–1912.
- Angulo, J., Ulryck, N., Deforges, J., Chamond, N., Lopez-Lastra, M., Masquida, B. and Sargueil, B. 2016. LOOP IIIId of the HCV IRES is essential for the structural rearrangement of the 40S-HCV IRES complex. *Nucleic Acids Research*. **44**(3),pp.1309–25.
- Asnani, M., Kumar, P. and Hellen, C.U.T. 2015. Widespread distribution and structural diversity of Type IV IRESs in members of Picornaviridae. *Virology*. [Online]. **478**,pp.61–74.
- Atkinson, N.J., Witteveldt, J., Evans, D.J. and 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*. [Online]. **42**(7),pp.4527–4545.
- Auperin, D.D., Romanowski, V., Galinski, M. and Bishop, D.H. 1984. Sequencing studies of pichinde arenavirus S RNA indicate a novel coding strategy, an ambisense viral S RNA. *Journal of virology*. **52**(3),pp.897–904.
- Aviran, S., Trapnell, C., Lucks, J.B., Mortimer, S.A., Luo, S., Schroth, G.P., Doudna, J.A., Arkin, A.P. and Pachter, L. 2011. Modeling and automation of sequencing-based characterization of RNA structure. *Proceedings of the National Academy of Sciences*. **108**(27),pp.11069–11074.
- Baechlein, C., Fischer, N., Grundhoff, A., Alawi, M., Indenbirken, D., Postel, A., Baron, A.L., Offinger, J., Becker, K., Beineke, A., Rehage, J. and Becher, P. 2015. Identification of a

- Novel Hepacivirus in Domestic Cattle from Germany. *Journal of virology*. **89**(14),pp.7007–15.
- Barría, M.I., González, A., Vera-Otarola, J., León, U., Vollrath, V., Marsac, D., Monasterio, O., Pérez-Acle, T., Soza, A. and López-Lastra, M. 2009. Analysis of natural variants of the hepatitis C virus internal ribosome entry site reveals that primary sequence plays a key role in cap-independent translation. *Nucleic Acids Research*. [Online]. **37**(3),pp.957–971.
- Bartenschlager, R., Ahlborn-Laake, L., Mous, J. and Jacobsen, H. 1994. Kinetic and structural analyses of hepatitis C virus polyprotein processing. *Journal of Virology*. [Online]. **68**(8),pp.5045–5055.
- Bartenschlager, R., Lohmann, V., Wilkinson, T. and Koch, J.O. 1995. Complex Formation between the NS3 Serine-Type Proteinase of the Hepatitis C Virus and NS4A and Its Importance for Polyprotein Maturation. *Journ of Virology* [Online]. **69**(12),pp.7519–7528.
- Bartenschlager, R., Penin, F., Lohmann, V. and André, P. 2011. Assembly of infectious hepatitis C virus particles. *Trends Microbiol*. **19**(2),pp.95–103.
- Belsham, G.J., McInerney, G.M. and Ross-Smith, N. 2000. Foot-and-mouth disease virus 3C protease induces cleavage of translation initiation factors eIF4A and eIF4G within infected cells. *Journal of virology*. **74**(1),pp.272–80.
- Bentham, M.J., Foster, T.L., McCormick, C. and Griffin, S. 2013. Mutations in hepatitis C virus p7 reduce both the egress and infectivity of assembled particles via impaired proton channel function. *Journal of General Virology*. [Online]. **94**(PART10),pp.2236–2248.
- Berry, K.E., Waghray, S. and Doudna, J. a 2010. The HCV IRES pseudoknot positions the initiation codon on the 40S ribosomal subunit. *RNA (New York, N.Y.)*. [Online]. **16**(8),pp.1559–1569.
- Bexfield, N.H., Watson, P.J., Heaney, J., Heaney, J.L. and Tiley, L. 2014. Canine hepacivirus is not associated with chronic liver disease in dogs. *Journal of viral hepatitis*. **21**(3),pp.223–8.
- Billerbeck, E., Wolfisberg, R., Fahnøe, U., Xiao, J.W., Quirk, C., Luna, J.M., Cullen, J.M., Hartlage, A.S., Chiriboga, L., Ghoshal, K., Lipkin, W.I., Bukh, J., Scheel, T.K.H., Kapoor, A. and Rice, C.M. 2017. Mouse models of acute and chronic hepacivirus infection. *Science*. **357**(6347).
- Blight, K.J., Kolykhalov, a a and Rice, C.M. 2000. Efficient initiation of HCV RNA replication in cell culture. *Science*. [Online]. **290**(5498),pp.1972–1974.
- Blight, K.J., McKeating, J. a, Marcotrigiano, J. and Rice, C.M. 2003. Efficient replication of hepatitis C virus genotype 1a RNAs in cell culture. *J Virol*. **77**(5),pp.3181–3190.
- Blight, K.J., Mckeating, J. a and Rice, C.M. 2002. Highly Permissive Cell Lines for Subgenomic and Genomic Hepatitis C Virus RNA Replication Highly Permissive Cell Lines for Subgenomic and Genomic Hepatitis C Virus RNA Replication. *Journal of Virology*. [Online]. **76**(24),pp.13001–13014.
- Borman, a M. and Kean, K.M. 1997. Intact eukaryotic initiation factor 4G is required for hepatitis A virus internal initiation of translation. *Virology*. **237**(1),pp.129–36.
- Bradrick, S.S., Walters, R.W. and Gromeier, M. 2006. The hepatitis C virus 3'-untranslated

- region or a poly(A) tract promote efficient translation subsequent to the initiation phase. *Nucleic Acids Research*. [Online]. **34**(4),pp.1293–1303.
- de Breyne, S., Yu, Y., Unbehaun, A., Pestova, T. V and Hellen, C.U.T. 2009. Direct functional interaction of initiation factor eIF4G with type 1 internal ribosomal entry sites. *Proceedings of the National Academy of Sciences of the United States of America*. [Online]. **106**(23),pp.9197–9202.
- Brown, E. a, Zajac, a J. and Lemon, S.M. 1994. In vitro characterization of an internal ribosomal entry site (IRES) present within the 5' nontranslated region of hepatitis A virus RNA: comparison with the IRES of encephalomyocarditis virus. *Journal of virology*. [Online]. **68**(2),pp.1066–1074.
- Burbelo, P.D., Dubovi, E.J., Simmonds, P., Medina, J.L., Henriquez, J.A., Mishra, N., Wagner, J., Tokarz, R., Cullen, J.M., Iadarola, M.J., Rice, C.M., Lipkin, W.I. and Kapoor, A. 2012. Serology-Enabled Discovery of Genetically Diverse Hepaciviruses in a New Host. *Journal of Virology*. **86**(11),pp.6171–6178.
- Catanese, M.T., Uryu, K., Kopp, M., Edwards, T.J., Andrus, L., Rice, W.J., Silvestry, M., Kuhn, R.J. and Rice, C.M. 2013. Ultrastructural analysis of hepatitis C virus particles. *Proceedings of the National Academy of Sciences of the United States of America*. **110**(23),pp.9505–10.
- Chambers, R. and Takimoto, T. 2009. Antagonism of innate immunity by Paramyxovirus accessory proteins. *Viruses*. [Online]. **1**(3),pp.574–593.
- Chen, S.L. and Morgan, T.R. 2006. The natural history of hepatitis C virus (HCV) infection. *International Journal of Medical Sciences*. [Online]. **3**(2),pp.47–52.
- Cheung, Y.N., Maag, D., Mitchell, S.F., Fekete, C.A., Algire, M.A., Takacs, J.E., Shirokikh, N., Pestova, T., Lorsch, J.R. and Hinnebusch, A.G. 2007. Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection in vivo. *Genes and Development*. [Online]. **21**(10),pp.1217–1230.
- Choo, Q.L., Kuo, G., Weiner, A.J., Overby, L.R., Bradley, D.W. and Houghton, M. 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science (New York, N.Y.)*. **244**(4902),pp.359–62.
- Choo, Q.L., Weiner, a J., Overby, L.R., Kuo, G., Houghton, M. and Bradley, D.W. 1990. Hepatitis C virus: the major causative agent of viral non-A, non-B hepatitis. *British Medical Bulletin*. [Online]. **46**(2),pp.423–441.
- Chung, A., Jin, B., Han, K.-H., Ahn, S.H. and Kim, S. 2017. Cell culture-adaptive mutations of NS5A affect replication of hepatitis C virus differentially depending on the viral genotypes. *Journal of Medical Virology*. **89**(1),pp.146–152.
- Clever, J., Sasseti, C. and Parslow, T.G. 1995. RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1. *Journal of virology*. [Online]. **69**(4),pp.2101–2109.
- Corman, V.M., Grundhoff, A., Baechlein, C., Fischer, N., Gmyl, A., Wollny, R., Dei, D., Ritz, D., Binger, T., Adankwah, E., Marfo, K.S., Annison, L., Annan, A., Adu-Sarkodie, Y., Oppong, S., Becher, P., Drosten, C. and Drexler, J.F. 2015. Highly divergent hepaciviruses from African cattle. *Journal of Virology*. **89**(March),p.JVI.00393-15.

- Counihan, N.A., Rawlinson, S.M. and Lindenbach, B.D. 2011. Trafficking of hepatitis C virus core protein during virus particle assembly. *PLoS Pathogens*. [Online]. **7**(10).
- Deigan, K.E., Li, T.W., Mathews, D.H. and Weeks, K.M. 2009. Accurate SHAPE-directed RNA structure determination. *Proceedings of the National Academy of Sciences*. **106**(1),pp.97–102.
- Drexler, J.F., Corman, V.M., Müller, M.A., Lukashev, A.N., Gmyl, A., Coutard, B., Adam, A., Ritz, D., Leijten, L.M., van Riel, D., Kallies, R., Klose, S.M., Gloza-Rausch, F., Binger, T., Annan, A., Adu-Sarkodie, Y., Oppong, S., Bourgarel, M., Rupp, D., Hoffmann, B., Schlegel, M., Kümmerer, B.M., Krüger, D.H., Schmidt-Chanasit, J., Setién, A.A., Cottontail, V.M., Hemachudha, T., Wacharapluesadee, S., Osterrieder, K., Bartenschlager, R., Matthee, S., Beer, M., Kuiken, T., Reusken, C., Leroy, E.M., Ulrich, R.G. and Drosten, C. 2013. Evidence for Novel Hepaciviruses in Rodents. *PLoS Pathogens*. [Online]. **9**(6).
- Egger, D., Wölk, B., Gosert, R., Bianchi, L., Blum, H.E., Moradpour, D. and Bienz, K. 2002. Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. *Journal of virology*. **76**(12),pp.5974–84.
- El-Attar, L.M.R., Mitchell, J.A., Brooks Brownlie, H., Priestnall, S.L. and Brownlie, J. 2015. Detection of non-primate hepaciviruses in UK dogs. *Virology*. [Online]. **484**,pp.93–102.
- Elia, G., Lanave, G., Lorusso, E., Parisi, A., Cavaliere, N., Patruno, G., Terregino, C., Decaro, N., Martella, V. and Buonavoglia, C. 2017. Identification and genetic characterization of equine hepaciviruses in Italy. *Veterinary Microbiology*. **207**,pp.239–247.
- Elia, G., Lanave, G., Lorusso, E., Parisi, A., Trotta, A., Buono, R., Martella, V., Decaro, N. and Buonavoglia, C. 2017. Equine hepacivirus persistent infection in a horse with chronic wasting. *Transboundary and Emerging Diseases*.
- Evans, M.J., von Hahn, T., Tscherne, D.M., Syder, A.J., Panis, M., Wölk, B., Hatzioannou, T., McKeating, J. a, Bieniasz, P.D. and Rice, C.M. 2007. Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. *Nature*. [Online]. **446**(7137),pp.801–805.
- Farquhar, M.J., Hu, K., Harris, H.J., Davis, C., Brimacombe, C.L., Fletcher, S.J., Baumert, T.F., Rappoport, J.Z., Balfe, P. and McKeating, J.A. 2012. Hepatitis C virus induces CD81 and claudin-1 endocytosis. *Journal of virology*. **86**(8),pp.4305–16.
- Figueiredo, A.S., Lampe, E., do Esp??rito-Santo, M.P., do Amaral Mello, F.C., de Almeida, F.Q., de Lemos, E.R.S., Godoi, T.L.O.S., Dimache, L.A.G., dos Santos, D.R.L. and Villar, L.M. 2015. Identification of two phylogenetic lineages of equine hepacivirus and high prevalence in Brazil. *Veterinary Journal*. [Online]. **206**(3),pp.414–416.
- Filbin, M.E. and Kieft, J.S. 2011. HCV IRES domain IIb affects the configuration of coding RNA in the 40S subunit's decoding groove. *RNA (New York, N.Y.)*. **17**(7),pp.1258–73.
- Filbin, M.E., Vollmar, B.S., Shi, D., Gonen, T. and Kieft, J.S. 2013. HCV IRES manipulates the ribosome to promote the switch from translation initiation to elongation. *Nature structural & molecular biology*. **20**(2),pp.150–8.
- Firth, C., Bhat, M., Firth, M.A., Williams, S.H., Frye, M.J., Simmonds, P., Conte, J.M., Ng, J., Garcia, J., Bhuvu, N.P., Lee, B., Che, X., Quan, P.L. and Ian Lipkin, W. 2014. Detection of zoonotic pathogens and characterization of novel viruses carried by commensal rattus

- norvegicus in New York city. *mBio*. [Online]. **5**(5).
- Frank, C., Mohamed, M.K., Strickland, G.T., Lavanchy, D., Arthur, R.R., Magder, L.S., El Khoby, T., Abdel-Wahab, Y., Aly Ohn, E.S., Anwar, W. and Sallam, I. 2000. The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *Lancet*. [Online]. **355**(9207),pp.887–891.
- Fraser, C.S., Berry, K.E., Hershey, J.W.B. and Doudna, J.A. 2007. eIF3j Is Located in the Decoding Center of the Human 40S Ribosomal Subunit. *Molecular Cell*. [Online]. **26**(6),pp.811–819.
- Fraser, C.S. and Doudna, J.A. 2007. Structural and mechanistic insights into hepatitis C viral translation initiation. *Nature Reviews Microbiology*. **5**(1),pp.29–38.
- Fraser, C.S., Hershey, J.W.B. and Doudna, J.A. 2009. The pathway of hepatitis C virus mRNA recruitment to the human ribosome. *Nat Struct Mol Biol*. **16**(4),pp.397–404.
- Fricke, M., Dünnes, N., Zayas, M., Bartenschlager, R., Niepmann, M. and Marz, M. 2015. Conserved RNA secondary structures and long-range interactions in hepatitis C viruses. *RNA (New York, N.Y.)*,p.rna.049338.114-.
- Friebe, P., Lohmann, V., Krieger, N. and Bartenschlager, R. 2001. Sequences in the 5' nontranslated region of hepatitis C virus required for RNA replication. *Journal of virology*. **75**(24),pp.12047–57.
- Fukuhara, T., Kambara, H., Shiokawa, M., Ono, C., Katoh, H., Morita, E., Okuzaki, D., Maehara, Y., Koike, K. and Matsuura, Y. 2012. Expression of microRNA miR-122 facilitates an efficient replication in nonhepatic cells upon infection with hepatitis C virus. *Journal of virology*. **86**(15),pp.7918–33.
- Gao, Y., Sun, S.-Q. and Guo, H.-C. 2016. Biological function of Foot-and-mouth disease virus non-structural proteins and non-coding elements. *Virology journal*. **13**,p.107.
- Gastaminza, P., Cheng, G., Wieland, S., Zhong, J., Liao, W. and Chisari, F. V 2008. Cellular determinants of hepatitis C virus assembly, maturation, degradation, and secretion. *Journal of virology*. **82**(5),pp.2120–9.
- Gastaminza, P., Dryden, K. a, Boyd, B., Wood, M.R., Law, M., Yeager, M. and Chisari, F. V 2010. Ultrastructural and biophysical characterization of hepatitis C virus particles produced in cell culture. *Journal of virology*. [Online]. **84**(21),pp.10999–11009.
- Gather, T., Walter, S., Todt, D., Pfaender, S., Brown, R.J.P., Postel, A., Becher, P., Moritz, A., Hansmann, F., Baumgaertner, W., Feige, K., Steinmann, E. and Cavalleri, J.M. V 2016. Vertical transmission of hepatitis C virus-like nonprimate hepacivirus in horses. *Journal of General Virology*. **97**(10),pp.2540–2551.
- Germi, R., Crance, J.-M., Garin, D., Guimet, J., Lortat-Jacob, H., Ruigrok, R.W.H., Zarski, J.-P. and Drouet, E. 2002. Cellular glycosaminoglycans and low density lipoprotein receptor are involved in hepatitis C virus adsorption. *Journal of medical virology*. **68**(2),pp.206–15.
- Gingras, A.C., Svitkin, Y., Belsham, G.J., Pause, A. and Sonenberg, N. 1996. Activation of the translational suppressor 4E-BP1 following infection with encephalomyocarditis virus and poliovirus. *Proceedings of the National Academy of Sciences of the United States of America*. **93**(11),pp.5578–83.

- Gradi, A., Svitkin, Y. V, Imataka, H. and Sonenberg, N. 1998. Proteolysis of human eukaryotic translation initiation factor eIF4GII, but not eIF4GI, coincides with the shutoff of host protein synthesis after poliovirus infection. *Proceedings of the National Academy of Sciences of the United States of America*. **95**(19),pp.11089–94.
- Griffin, S., Clarke, D., McCormick, C., Rowlands, D. and Harris, M. 2005. Signal peptide cleavage and internal targeting signals direct the hepatitis C virus p7 protein to distinct intracellular membranes. *J Virol*. **79**(24),pp.15525–15536.
- von der Haar, T., Gross, J.D., Wagner, G. and McCarthy, J.E.G. 2004. The mRNA cap-binding protein eIF4E in post-transcriptional gene expression. *Nature structural & molecular biology*. [Online]. **11**(6),pp.503–511.
- Hajdin, C.E., Bellaousov, S., Huggins, W., Leonard, C.W., Mathews, D.H. and Weeks, K.M. 2013. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proceedings of the National Academy of Sciences of the United States of America*. **110**(14),pp.5498–503.
- Hartlage, A.S., Cullen, J.M. and Kapoor, A. 2016. The Strange, Expanding World of Animal Hepaciviruses. *Annual Review of Virology*. **3**(1),pp.53–75.
- Hashem, Y., Des Georges, A., Dhote, V., Langlois, R., Liao, H.Y., Grassucci, R.A., Pestova, T. V, Hellen, C.U.T. and Frank, J. 2013. Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit. *Nature*. **503**.
- Hertz, M.I. and Thompson, S.R. 2011. Mechanism of translation initiation by Dicistroviridae IGR IRESs. *Virology*. [Online]. **411**(2),pp.355–361.
- Hijikata, M., Mizushima, H., Akagi, T., Mori, S., Kakiuchi, N., Kato, N., Tanaka, T., Kimura, K. and Shimotohno, K. 1993. Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus. *Journal of virology*. [Online]. **67**(8),pp.4665–4675.
- Honda, M., Ping, L.-H., Rijnbrand, R.C.A., Amphlett, E., Clarke, B., Rowlands, D. and Lemon, S.M. 1996. Structural Requirements for Initiation of Translation by Internal Ribosome Entry within Genome-Length Hepatitis C Virus RNA. *Virology*. **222**(1),pp.31–42.
- Jaafar, Z.A., Oguro, A., Nakamura, Y. and Kieft, J.S. 2016. Translation initiation by the hepatitis C virus IRES requires eIF1A and ribosomal complex remodeling. *eLife*. [Online]. **5**(DECEMBER2016).
- Jackson, R.J., Hellen, C.U.T. and Pestova, T. V 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*. **11**(2),pp.113–127.
- Jang, S.K., Kräusslich, H.G., Nicklin, M.J., Duke, G.M., Palmenberg, A.C. and Wimmer, E. 1988. A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *Journal of virology*. **62**(8),pp.2636–43.
- Jangra, R.K., Yi, M. and Lemon, S.M. 2010. Regulation of hepatitis C virus translation and infectious virus production by the microRNA miR-122. *Journal of virology*. **84**(13),pp.6615–25.

- Ji, H. 2004. Coordinated assembly of human translation initiation complexes by the hepatitis C virus internal ribosome entry site RNA. *Proceedings of the National Academy of Sciences of the United States of America*. **101**(49),pp.16990–5.
- Jopling, C.L., Yi, M., Lancaster, A.M., Lemon, S.M. and Sarnow, P. 2005. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science (New York, N.Y.)*. **309**(5740),pp.1577–81.
- Jubin, R., Vantuno, N.E., Kieft, J.S., Murray, M.G., Doudna, J.A., Lau, J.Y. and Baroudy, B.M. 2000. Hepatitis C virus internal ribosome entry site (IRES) stem loop IIIId contains a phylogenetically conserved GGG triplet essential for translation and IRES folding. *Journal of virology*. **74**(22),pp.10430–7.
- Jubin, R., Vantuno, N.E., Kieft, J.S., Murray, M.G., Doudna, J., Lau, J.Y. and Baroudy, B.M. 2000. Hepatitis C virus internal ribosome entry site (IRES) stem loop IIIId contains a phylogenetically conserved GGG triplet essential for translation and IRES folding. *Journal of Virology*. [Online]. **74**(22),pp.10430–10437.
- Kalliampakou, K.I., Psaridi-Linardaki, L. and Mavromara, P. 2002. Mutational analysis of the apical region of domain II of the HCV IRES. *FEBS Letters*. [Online]. **511**(1–3),pp.79–84.
- Kambara, H., Fukuhara, T., Shiokawa, M., Ono, C., Ohara, Y., Kamitani, W. and Matsuura, Y. 2012. Establishment of a Novel Permissive Cell Line for the Propagation of Hepatitis C Virus by Expression of MicroRNA miR122. *Journal of Virology*. [Online]. **86**,pp.1382–1393.
- Kamoshita, N., Tsukiyama, K.K., Kohara, M. and Nomoto, A. 1997. Genetic analysis of internal ribosomal entry site on hepatitis C virus RNA: implication for involvement of the highly ordered structure and cell type-specific transacting factors. *Virology*. [Online]. **233**(1),pp.9–18.
- Kapoor, A., Simmonds, P., Gerold, G., Qaisar, N., Jain, K., Henriquez, J.A., Firth, C., Hirschberg, D.L., Rice, C.M., Shields, S. and Lipkin, W.I. 2011. Characterization of a canine homolog of hepatitis C virus. *Proceedings of the National Academy of Sciences*. **108**(28),pp.11608–11613.
- Kapoor, A., Simmonds, P., Scheel, T.K.H., Hjelle, B., Cullen, J.M., Burbelo, P.D., Chauhan, L. V., Duraisamy, R., Sanchez Leon, M., Jain, K., Vandegrift, K.J., Calisher, C.H., Rice, C.M. and Lipkin, W.I. 2013. Identification of rodent homologs of hepatitis C virus and pegiviruses. *mBio*. [Online]. **4**(2).
- Kapoor, Simmonds, P., Gerold, G., Qaisar, N., Jain, K., Henriquez, J. a, Firth, C., Hirschberg, D.L., Rice, C.M., Shields, S. and Lipkin, W.I. 2011. Characterization of a canine homolog of hepatitis C virus. *Proceedings of the National Academy of Sciences of the United States of America*. [Online]. **108**(28),pp.11608–11613.
- Karabiber, F., McGinnis, J.L., Favorov, O. V and Weeks, K.M. 2013. QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA (New York, N.Y.)*. **19**(1),pp.63–73.
- Kato, T., Date, T., Miyamoto, M., Furusaka, A., Tokushige, K., Mizokami, M. and Wakita, T. 2003. Efficient Replication of the Genotype 2a Hepatitis C Virus Subgenomic Replicon. *Gastroenterology*. [Online]. **125**(6),pp.1808–1817.

- Kato, T., Date, T., Miyamoto, M., Zhao, Z., Mizokami, M. and Wakita, T. 2005. Nonhepatic cell lines HeLa and 293 support efficient replication of the hepatitis C virus genotype 2a subgenomic replicon. *Journal of virology*. [Online]. **79**(1),pp.592–596.
- Kenyon, J., Prestwood, L. and Lever, A. 2014. Current perspectives on RNA secondary structure probing. *Biochemical Society transactions*. **42**(4),pp.1251–5.
- Khawaja, A., Vopalensky, V. and Pospisek, M. 2015. Understanding the potential of hepatitis C virus internal ribosome entry site domains to modulate translation initiation via their structure and function. *Wiley Interdisciplinary Reviews: RNA*. **6**(2),pp.211–224.
- Kieft, J.S., Zhou, K., Jubin, R. and Doudna, J.A. 2001. Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA (New York, N.Y.)*. **7**(2),pp.194–206.
- Kim, D.W., Gwack, Y., Han, J.H. and Choe, J. 1995. C-terminal domain of the hepatitis C virus NS3 protein contains an RNA helicase activity. *Biochemical and biophysical research communications*. [Online]. **215**(1),pp.160–6.
- Kim, H.-S., Moon, H.-W., Sung, H.W. and Kwon, H.M. 2017. First identification and phylogenetic analysis of equine hepacivirus in Korea. *Infection, Genetics and Evolution*. **49**,pp.268–272.
- Kinge, C.N.W., Espiritu, C., Prabdial-Sing, N., Sithebe, N.P., Saeed, M. and Rice, C.M. 2014. Hepatitis C virus genotype 5a subgenomic replicons for evaluation of direct-acting antiviral agents. *Antimicrobial Agents and Chemotherapy*. [Online]. **58**(9),pp.5386–5394.
- Kolupaeva, V.G., Lomakin, I.B., Pestova, T. V and Hellen, C.U.T. 2003. Eukaryotic initiation factors 4G and 4A mediate conformational changes downstream of the initiation codon of the encephalomyocarditis virus internal ribosomal entry site. *Molecular and cellular biology*. [Online]. **23**(2),pp.687–698.
- Kolupaeva, V.G., Pestova, T. V and Hellen, C.U. 2000. An enzymatic footprinting analysis of the interaction of 40S ribosomal subunits with the internal ribosomal entry site of hepatitis C virus. *Journal of virology*. **74**(14),pp.6242–50.
- Kozak, M. 1991. Structural features in eukaryotic mRNAs that modulate the initiation of translation. *Journal of Biological Chemistry*. [Online]. **266**(30),pp.19867–19870.
- Van Der Laan, L.J.W., De Ruiter, P.E., Van Gils, I.M., Fieten, H., Spee, B., Pan, Q., Rothuizen, J. and Penning, L.C. 2014. Canine hepacivirus and idiopathic hepatitis in dogs from a Dutch cohort. *Journal of Viral Hepatitis*. [Online]. **21**(12),pp.894–896.
- Lauck, M., Sibley, S.D., Lara, J., Purdy, M.A., Khudyakov, Y., Hyeroba, D., Tumukunde, A., Weny, G., Switzer, W.M., Chapman, C.A., Hughes, A.L., Friedrich, T.C., O'Connor, D.H. and Goldberg, T.L. 2013. A novel hepacivirus with an unusually long and intrinsically disordered NS5A protein in a wild Old World primate. *Journal of virology*. **87**(16),pp.8971–81.
- Lauer, G.M. and Walker, B.D. 2001. Hepatitis C virus infection. *The New England journal of medicine*. **345**(1),pp.41–52.
- Lee, J.H., Pestova, T. V, Shin, B.-S., Cao, C., Choi, S.K. and Dever, T.E. 2002. Initiation factor eIF5B catalyzes second GTP-dependent step in eukaryotic translation initiation. *Proceedings of the National Academy of Sciences of the United States of America*. [Online]. **99**(26),pp.16689–16694.

- Lee, K.M., Chen, C.J. and Shih, S.R. 2017. Regulation Mechanisms of Viral IRES-Driven Translation. *Trends in Microbiology*. [Online]. **25**(7),pp.546–561.
- Lefebvre, A.K., Korneeva, N.L., Trutschl, M., Cvek, U., Duzan, R.D., Bradley, C.A., Hershey, J.W.B. and Rhoads, R.E. 2006. Translation initiation factor eIF4G-1 binds to eIF3 through the eIF3e subunit. *Journal of Biological Chemistry*. [Online]. **281**(32),pp.22917–22932.
- Li, K., Foy, E., Ferreon, J.C., Nakamura, M., Ferreon, A.C.M., Ikeda, M., Ray, S.C., Gale, M. and Lemon, S.M. 2005. Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. *Proceedings of the National Academy of Sciences of the United States of America*. [Online]. **102**(8),pp.2992–7.
- Lin, L.-T., Noyce, R.S., Pham, T.N.Q., Wilson, J. a, Sisson, G.R., Michalak, T.I., Mossman, K.L. and Richardson, C.D. 2010. Replication of subgenomic hepatitis C virus replicons in mouse fibroblasts is facilitated by deletion of interferon regulatory factor 3 and expression of liver-specific microRNA 122. *Journal of virology*. [Online]. **84**(18),pp.9170–9180.
- Lindenbach, B.D. and Rice, C.M. 2013. The ins and outs of hepatitis C virus entry and assembly. *Nature reviews. Microbiology*. **11**(10),pp.688–700.
- Locker, N., Easton, L.E. and Lukavsky, P.J. 2007. HCV and CSFV IRES domain II mediate eIF2 release during 80S ribosome assembly. *The EMBO journal*. **26**(3),pp.795–805.
- Lohmann, V. and Bartenschlager, R. 2014. On the history of hepatitis C virus cell culture systems. *Journal of Medicinal Chemistry*. [Online]. **57**(5),pp.1627–1642.
- Lohmann, V., Körner, F., Koch, J., Herian, U., Theilmann, L. and Bartenschlager, R. 1999. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. *Science (New York, N.Y.)*. [Online]. **285**(5424),pp.110–113.
- Lomakin, I.B., Kolupaeva, V.G., Marintchev, A., Wagner, G. and Pestova, T. V. 2003. Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. *Genes and Development*. [Online]. **17**(22),pp.2786–2797.
- Lorenz, R., Luntzer, D., Hofacker, I.L., Stadler, P.F. and Wolfinger, M.T. 2015. SHAPE directed RNA folding. *Bioinformatics*. [Online]. **32**(1),pp.145–147.
- Lozano, G. and Martínez-Salas, E. 2015. Structural insights into viral IRES-dependent translation mechanisms. *Current Opinion in Virology*. [Online]. **12**,pp.113–120.
- Lu, G., Sun, L., Xu, T., He, D., Wang, Z., Ou, S., Jia, K., Yuan, L. and Li, S. 2016. First description of hepacivirus and pegivirus infection in domestic Horses in China: A study in guangdong province, heilongjiang province and Hong Kong district. *PLoS ONE*. **11**(5).
- Lukavsky, P.J. 2009. Structure and function of HCV IRES domains. *Virus research*. **139**(2),pp.166–71.
- Luo, G., Xin, S. and Cai, Z. 2003. Role of the 5'-proximal stem-loop structure of the 5' untranslated region in replication and translation of hepatitis C virus RNA. *Journal of virology*. [Online]. **77**(5),pp.3312–3318.
- Lyons, S., Kapoor, A., Schneider, B.S., Wolfe, N.D., Culshaw, G., Corcoran, B., Durham, A.E., Burden, F., McGorum, B.C. and Simmonds, P. 2014. Viraemic frequencies and seroprevalence of non-primate hepacivirus and equine pegiviruses in horses and other

- mammalian species. *Journal of General Virology*. [Online]. **95**(PART 8),pp.1701–1711.
- Lyons, S., Kapoor, A., Sharp, C., Schneider, B.S., Wolfe, N.D., Culshaw, G., Corcoran, B., McGorum, B.C. and Simmonds, P. 2012. Nonprimate hepaciviruses in domestic horses, United Kingdom. *Emerging Infectious Diseases*. [Online]. **18**(12),pp.1976–1982.
- Maag, D., Algire, M.A. and Lorsch, J.R. 2006. Communication between eukaryotic translation initiation factors 5 and 1A within the ribosomal pre-initiation complex plays a role in start site selection. *Journal of Molecular Biology*. [Online]. **356**(3),pp.724–737.
- Macdonald, A. and Harris, M. 2004. Hepatitis C virus NS5A: Tales of a promiscuous protein. *Journal of General Virology*. [Online]. **85**(9),pp.2485–2502.
- Mackenzie, J. 2005. Wrapping things up about virus RNA replication. *Traffic*. [Online]. **6**(11),pp.967–977.
- Maeda, K., Yasumoto, S., Tsuruda, A., Andoh, K., Kai, K., Otoi, T. and Matsumura, T. 2007. Establishment of a novel equine cell line for isolation and propagation of equine herpesviruses. *The Journal of veterinary medical science / the Japanese Society of Veterinary Science*. [Online]. **69**(9),pp.989–991.
- Malygin, A.A., Kossinova, O.A., Shatsky, I.N. and Karpova, G.G. 2013. HCV IRES interacts with the 18S rRNA to activate the 40S ribosome for subsequent steps of translation initiation. *Nucleic acids research*. **41**(18),pp.8706–14.
- Marintchev, A., Edmonds, K.A., Marintcheva, B., Hendrickson, E., Oberer, M., Suzuki, C., Herdy, B., Sonenberg, N. and Wagner, G. 2009. Topology and Regulation of the Human eIF4A/4G/4H Helicase Complex in Translation Initiation. *Cell*. **136**(3),pp.447–460.
- Marintchev, A., Kolupaeva, V.G., Pestova, T. V and Wagner, G. 2003. Mapping the binding interface between human eukaryotic initiation factors 1A and 5B: a new interaction between old partners. *Proceedings of the National Academy of Sciences of the United States of America*. [Online]. **100**,pp.1535–1540.
- Martin, D.N. and Uprichard, S.L. 2013. Identification of transferrin receptor 1 as a hepatitis C virus entry factor. *Proceedings of the National Academy of Sciences of the United States of America*. **110**(26),pp.10777–82.
- Martínez-Salas, E., Piñeiro, D. and Fernández, N. 2012. Alternative mechanisms to initiate translation in eukaryotic mRNAs. *Comparative and Functional Genomics*. [Online]. **2012**.
- Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M. and Turner, D.H. 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences of the United States of America*. **101**(19),pp.7287–92.
- Matsuda, D. and Mauro, V.P. 2014. Base pairing between hepatitis C virus RNA and 18S rRNA is required for IRES-dependent translation initiation in vivo. *Proceedings of the National Academy of Sciences of the United States of America*. **111**(43),pp.15385–9.
- Matsuu, A., Hobo, S., Ando, K., Sanekata, T., Sato, F., Endo, Y., Amaya, T., Osaki, T., Horie, M., Masatani, T., Ozawa, M. and Tsukiyama-Kohara, K. 2015. Genetic and serological surveillance for non-primate hepacivirus in horses in Japan. *Veterinary Microbiology*. [Online]. **179**(3–4),pp.219–227.

- McGinnis, J.L., Dunkle, J.A., Cate, J.H.D. and Weeks, K.M. 2012. The mechanisms of RNA SHAPE chemistry. *Journal of the American Chemical Society*. [Online]. **134**(15),pp.6617–6624.
- McLauchlan, J. 2000. Properties of the hepatitis C virus core protein: a structural protein that modulates cellular processes. *Journal of viral hepatitis*. [Online]. **7**(1),pp.2–14.
- Merino, E.J., Wilkinson, K.A., Coughlan, J.L. and Weeks, K.M. 2005. RNA structure analysis at single nucleotide resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *Journal of the American Chemical Society*. [Online]. **127**(12),pp.4223–4231.
- Meylan, E., Curran, J., Hofmann, K., Moradpour, D., Binder, M., Bartenschlager, R. and Tschopp, J. 2005. Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. *Nature*. [Online]. **437**(7062),pp.1167–1172.
- Mitra, S., Shcherbakova, I. V., Altman, R.B., Brenowitz, M. and Laederach, A. 2008. High-throughput single-nucleotide structural mapping by capillary automated footprinting analysis. *Nucleic Acids Research*. [Online]. **36**(11).
- Monazahian, M., Böhme, I., Bonk, S., Koch, a, Scholz, C., Grethe, S. and Thomssen, R. 1999. Low density lipoprotein receptor as a candidate receptor for hepatitis C virus. *Journal of medical virology*. [Online]. **57**(3),pp.223–229.
- Moradpour, D., Brass, V., Bieck, E., Friebe, P., Gosert, R., Blum, H.E., Bartenschlager, R., Penin, F. and Lohmann, V. 2004. Membrane association of the RNA-dependent RNA polymerase is essential for hepatitis C virus RNA replication. *J Virol*. [Online]. **78**,pp.13278–13284.
- Nakabayashi, H., Taketa, K., Miyano, K., Yamane, T. and Sato, J. 1982. Growth of human hepatoma cells lines with differentiated functions in chemically defined medium. *Cancer research*. **42**(9),pp.3858–3863.
- Nilsen, T.W. 2013. RNA structure determination using nuclease digestion. *Cold Spring Harbor Protocols*. [Online]. **8**(4),pp.379–382.
- Odreman-Macchioli, F.E., Tisminetzky, S.G., Zotti, M., Baralle, F.E. and Buratti, E. 2000. Influence of correct secondary and tertiary RNA folding on the binding of cellular factors to the HCV IRES. *Nucleic acids research*. **28**(4),pp.875–85.
- Oguma, K., Ishida, M., Maeda, K. and Sentsui, H. 2013. Efficient Propagation of Equine Viruses in a Newly Established Equine Cell Line, FHK-Tcl3.1 Cells. *Journal of Veterinary Medical Science*. **75**(9),pp.1223–1225.
- Ohto, H., Terazawa, S., Sasaki, N., Hino, K., Ishiwata, C., Kako, M., Ujiie, N., Endo, C. and Matsui, A. 1994. Transmission of hepatitis C virus from mothers to infants. The Vertical Transmission of Hepatitis C Virus Collaborative Study Group. *The New England journal of medicine*. **330**(11),pp.744–50.
- Otto, G.A. and Puglisi, J.D. 2004. The pathway of HCV IRES-mediated translation initiation. *Cell*. [Online]. **119**(3),pp.369–380.
- Pang, P.S., Elazar, M., Pham, E.A. and Glenn, J.S. 2011. Simplified RNA secondary structure mapping by automation of SHAPE data analysis. *Nucleic Acids Research*. [Online]. **39**(22).
- Parera, M., Martrus, G., Franco, S., Clotet, B. and Martinez, M.A. 2012. Canine hepatic virus NS3 serine protease can cleave the human adaptor proteins MAVS and TRIF. *PLoS ONE*.

[Online]. **7**(8).

- Passmore, L.A., Schmeing, T.M., Maag, D., Applefield, D.J., Acker, M.G., Algire, M.A., Lorsch, J.R. and Ramakrishnan, V. 2007. The Eukaryotic Translation Initiation Factors eIF1 and eIF1A Induce an Open Conformation of the 40S Ribosome. *Molecular Cell*. **26**(1),pp.41–50.
- Pelletier, J. and Sonenberg, N. 1988. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature*. [Online]. **334**(6180),pp.320–325.
- Pestova, T. V. and Kolupaeva, V.G. 2002. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes and Development*. [Online]. **16**(22),pp.2906–2922.
- Pestova, T. V., Shatsky, I.N., Fletcher, S.P., Jackson, R.J. and Hellen, C.U.T. 1998. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes and Development*. [Online]. **12**(1),pp.67–83.
- Pestova, T. V, Borukhov, S.I. and Hellen, C.U. 1998. Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature*. [Online]. **394**(6696),pp.854–859.
- Pestova, T. V, Shatsky, I.N., Fletcher, S.P., Jackson, R.J. and Hellen, C.U. 1998. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev*. **12**,pp.67–83.
- Pestova, T. V, Shatsky, I.N. and Hellen, C.U. 1996. Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes. *Molecular and cellular biology*. **16**(12),pp.6870–6878.
- Pfaender, S., Cavalleri, J.M. V, Walter, S., Doerrbecker, J., Campana, B., Brown, R.J.P., Burbelo, P.D., Postel, A., Hahn, K., Anggakusuma, Riebesehl, N., Baumgärtner, W., Becher, P., Heim, M.H., Pietschmann, T., Feige, K. and Steinmann, E. 2015. Clinical course of infection and viral tissue tropism of hepatitis C virus-like nonprimate hepaciviruses in horses. *Hepatology*. [Online]. **61**(2),pp.447–459.
- Pfaender, S., Walter, S., Grabski, E., Todt, D., Bruening, J., Romero-Brey, I., Gather, T., Brown, R.J.P., Hahn, K., Puff, C., Pfankuche, V.M., Hansmann, F., Postel, A., Becher, P., Thiel, V., Kalinke, U., Wagner, B., Bartenschlager, R., Baumgärtner, W., Feige, K., Pietschmann, T., Cavalleri, J.M. V and Steinmann, E. 2017. Immune protection against reinfection with nonprimate hepacivirus. *Proceedings of the National Academy of Sciences of the United States of America*. **114**(12),pp.E2430–E2439.
- Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A.J., Houghton, M., Rosa, D., Grandi, G. and Abrignani, S. 1998. Binding of hepatitis C virus to CD81. *Science (New York, N.Y.)*. **282**(5390),pp.938–41.
- Pilipenko, E. V, Gmyl, A.P., Maslova, S. V, Svitkin, Y. V, Sinyakov, A.N. and Agol, V.I. 1992. Prokaryotic-like cis elements in the cap-independent internal initiation of translation on picornavirus RNA. *Cell*. **68**(1),pp.119–131.

- Pisarev, A. V., Hellen, C.U.T. and Pestova, T. V. 2007. Recycling of Eukaryotic Posttermination Ribosomal Complexes. *Cell*. [Online]. **131**(2),pp.286–299.
- Pisarev, A. V., Kolupaeva, V.G., Pisareva, V.P., Merrick, W.C., Hellen, C.U.T. and Pestova, T. V. 2006. Specific functional interactions of nucleotides at key-3 and+4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes and Development*. [Online]. **20**(5),pp.624–636.
- Ploss, A., Evans, M.J., Gaysinskaya, V.A., Panis, M., You, H., de Jong, Y.P. and Rice, C.M. 2009. Human occludin is a hepatitis C virus entry factor required for infection of mouse cells. *Nature*. **457**(7231),pp.882–6.
- Postel, A., Cavalleri, J.-M. V., Pfaender, S., Walter, S., Steinmann, E., Fischer, N., Feige, K., Haas, L. and Becher, P. 2016. Frequent presence of hepaciviruses in commercial equine serum pools. *Veterinary Microbiology*. **182**,pp.8–14.
- Pronost, S., Hue, E., Fortier, C., Foursin, M., Fortier, G., Desbrosse, F., Rey, F.A., Pitel, P.H., Richard, E. and Saunier, B. 2016. Prevalence of Equine Hepacivirus Infections in France and Evidence for Two Viral Subtypes Circulating Worldwide. *Transboundary and Emerging Diseases*. [Online].
- Quan, P., Firth, C., Conte, J.M., Williams, S.H. and Zambrana-torrel, C.M. 2012. Bats are a major natural reservoir for hepaciviruses and pegiviruses. *Proceedings of the National Academy of Sciences*. [Online]. **110**,pp.3–8.
- Ramsay, J.D., Evanoff, R., Wilkinson, T.E., Divers, T.J., Knowles, D.P. and Mealey, R.H. 2015. Experimental transmission of equine hepacivirus in horses as a model for hepatitis C virus. *Hepatology*. [Online]. **61**(5),pp.1533–1546.
- Ray, P.S. and Das, S. 2004. Inhibition of hepatitis C virus IRES-mediated translation by small RNAs analogous to stem-loop structures of the 5'-untranslated region. *Nucleic Acids Research*. [Online]. **32**(5),pp.1678–1687.
- Redondo, N., Sanz, M.A., Steinberger, J., Skern, T., Kusov, Y. and Carrasco, L. 2012. Translation Directed by Hepatitis A Virus IRES in the Absence of Active eIF4F Complex and eIF2. *PLoS ONE*. [Online]. **7**(12).
- Reuter, G., Maza, N., Pankovics, P. and Boros, A. 2014. Non-primate hepacivirus infection with apparent hepatitis in a horse - Short communication. *Acta veterinaria Hungarica*. **62**(3),pp.422–7.
- Reuter, J.S. and Mathews, D.H. 2010. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*. **11**(1),p.129.
- Reynolds, J.E., Kaminski, A., Carroll, R., Clarke, B.E., Rowlands, D.J. and Jackson, R.J. 1996. Internal initiation of translation of hepatitis C virus RNA: the ribosome entry site is at the authentic initiation codon. *RNA*. **2**,pp.867–78.
- Rice, G.M., Leonard, C.W. and Weeks, K.M. 2014. RNA secondary structure modeling at consistent high accuracy using differential SHAPE. *RNA (New York, N.Y.)*. **20**(6),pp.846–54.
- Rogers, G.W., Richter, N.J., Lima, W.F. and Merrick, W.C. 2001. Modulation of the Helicase Activity of eIF4A by eIF4B, eIF4H, and eIF4F. *Journal of Biological Chemistry*. [Online]. **276**(33),pp.30914–30922.

- Ross-Thriepland, D. and Harris, M. 2015. Hepatitis C virus NS5A: Enigmatic but still promiscuous 10 years on! *Journal of General Virology*. [Online]. **96**(4),pp.727–738.
- Saeed, M., Andreo, U., Chung, H.-Y., Espiritu, C., Branch, A.D., Silva, J.M. and Rice, C.M. 2015. SEC14L2 enables pan-genotype HCV replication in cell culture. *Nature*. **524**(7566),pp.471–5.
- Saeed, M., Scheel, T.K.H., Gottwein, J.M., Marukian, S., Dustin, L.B., Bukh, J. and Rice, C.M. 2012. Efficient replication of genotype 3a and 4a hepatitis C virus replicons in human hepatoma cells. *Antimicrobial Agents and Chemotherapy*. [Online]. **56**(10),pp.5365–5373.
- Sainz, B.J., Barretto1, N., Martin, D.N., Hiraga, N., Imamura, M., Hussain, S., Marsh, K.A., Yu, X., Chayama, K., Alrefai, W.A. and Uprichard, S.L. 2012. Identification of the Niemann-Pick C1-like 1 cholesterol absorption receptor as a new hepatitis C virus entry factor Bruno. *Nature Medicine*. [Online]. **18**(2),pp.281–285.
- Salonen, a, Ahola, T. and Kääriäinen, L. 2005. Viral RNA replication in association with cellular membranes. *Current topics in microbiology and immunology*. [Online]. **285**,pp.139–173.
- Sasaki, J. and Nakashima, N. 1999. Translation initiation at the CUU codon is mediated by the internal ribosome entry site of an insect picorna-like virus in vitro. *Journal of virology*. [Online]. **73**(2),pp.1219–1226.
- Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R.M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R. and Vitelli, A. 2002. The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. *The EMBO journal*. [Online]. **21**(19),pp.5017–25.
- Scheel, T.K.H., Kapoor, A., Nishiuchi, E., Brock, K. V., Yu, Y., Andrus, L., Gu, M., Renshaw, R.W., Dubovi, E.J., McDonough, S.P., Van de Walle, G.R., Lipkin, W.I., Divers, T.J., Tennant, B.C. and Rice, C.M. 2015. Characterization of nonprimate hepacivirus and construction of a functional molecular clone. *Proceedings of the National Academy of Sciences of the United States of America*. **112**(7),pp.2192–2197.
- Schroeder, S.J. 2009. MINIREVIEW Advances in RNA Structure Prediction from Sequence: New Tools for Generating Hypotheses about Viral RNA Structure-Function Relationships □ WHAT IS RNA STRUCTURE PREDICTION? *J. Virol.* [Online]. **83**(13),pp.6326–6334.
- Sheppard, T. 2005. Getting RNA into SHAPE. *Nature Chemical Biology, Published online: 08 April 2005; | doi:10.1038/nchembio010*.
- Shi, M., Lin, X.-D., Vasilakis, N., Tian, J.-H., Li, C.-X., Chen, L.-J., Eastwood, G., Diao, X.-N., Chen, M.-H., Chen, X., Qin, X.-C., Widen, S.G., Wood, T.G., Tesh, R.B., Xu, J., Holmes, E.C. and Zhang, Y.-Z. 2015. Divergent viruses discovered in arthropods and vertebrates revise the evolutionary history of the *Flaviviridae* and related viruses. *Journal of Virology*. **90**(2),p.JVI.02036-15.
- Shi, S.T. and Lai, M.M.C. 2006. HCV 5' UTR and 3' UTR : When Translation Meets Replication. *Hepatitis C Viruses: Genomes and Molecular Biology*. [Online],pp.49–87.
- Shiokawa, M., Fukuhara, T., Ono, C., Yamamoto, S., Okamoto, T., Watanabe, N., Wakita, T. and Matsuura, Y. 2014. Novel permissive cell lines for complete propagation of hepatitis C virus. *Journal of virology*. **88**(10),pp.5578–94.

- Simmonds, P., Tuplin, A. and Evans, D.J. 2004. Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence. *RNA (New York, N.Y.)*. [Online]. **10**(9),pp.1337–1351.
- Siridechadilok, B. 2005. Structural Roles for Human Translation Factor eIF3 in Initiation of Protein Synthesis. *Science*. **310**(5753),pp.1513–1515.
- Sizova, D. V., Kolupaeva, V.G., Pestova, T. V., Shatsky, I.N. and Hellen, C.U.T. 1998. Specific Interaction of Eukaryotic Translation Initiation Factor 3 with the 5' Nontranslated Regions of Hepatitis C Virus and Classical Swine Fever Virus RNAs. *Journal of Virology*. **72**(6),pp.4775–4782.
- Smith, D.B., Becher, P., Bukh, J., Gould, E.A., Meyers, G., Monath, T., Muerhoff, A.S., Pletnev, A., Rico-Hesse, R., Stapleton, J.T. and Simmonds, P. 2016. Proposed update to the taxonomy of the genera Hepacivirus and Pegivirus within the Flaviviridae family. *Journal of General Virology*. [Online]. **97**(11),pp.2894–2907.
- Song, Yutong, Friebe, P., Tzima, E., Jünemann, C., Bartenschlager, R. and Niepmann, M. 2006. The Hepatitis C Virus RNA 3'-Untranslated Region Strongly Enhances Translation Directed by the Internal Ribosome Entry Site. *Journal of Virology*. [Online]. **80**(23),pp.11579–11588.
- Spahn, C.M.T., Kieft, J.S., Grassucci, R.A., Penczek, P.A., Zhou, K., Doudna, J.A. and Frank, J. 2001. Hepatitis C Virus IRES RNA-Induced Changes in the Conformation of the 40S Ribosomal Subunit. *Science*. **291**(5510),pp.1959–1962.
- Stewart, H., Bingham, R.J., White, S.J., Dykeman, E.C., Zothner, C., Tuplin, A.K., Stockley, P.G., Twarock, R. and Harris, M. 2016. Identification of novel RNA secondary structures within the hepatitis C virus genome reveals a cooperative involvement in genome packaging. *Scientific Reports*. **6**,p.22952.
- Stewart, H., Walter, C., Jones, D., Lyons, S., Simmonds, P. and Harris, M. 2013. The non-primate hepacivirus 5' untranslated region possesses internal ribosomal entry site activity. *Journal of General Virology*. [Online]. **94**(PART 12),pp.2657–2663.
- Sun, C., Querol-Audí, J., Mortimer, S.A., Arias-Palomo, E., Doudna, J.A., Nogales, E. and Cate, J.H.D. 2013. Two RNA-binding motifs in eIF3 direct HCV IRES-dependent translation. *Nucleic Acids Research*. [Online]. **41**(15),pp.7512–7521.
- Sweeney, T.R., Abaeva, I.S., Pestova, T. V. and Hellen, C.U.T. 2014. The mechanism of translation initiation on type 1 picornavirus IRESs. *EMBO Journal*. [Online]. **33**(1),pp.76–92.
- Sweeney, T.R., Dhote, V., Yu, Y. and Hellen, C.U.T. 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**(3),pp.1468–86.
- Tai, C.L., Chi, W.K., Chen, D.S. and Hwang, L.H. 1996. The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3). *Journal of virology*. **70**(12),pp.8477–84.
- Tanaka, T., Kasai, H., Yamashita, A., Okuyama-Dobashi, K., Yasumoto, J., Maekawa, S., Enomoto, N., Okamoto, T., Matsuura, Y., Morimatsu, M., Manabe, N., Ochiai, K., Yamashita, K. and Moriishi, K. 2014. Hallmarks of hepatitis C virus in equine hepacivirus.

Journal of virology. **88**(22),pp.13352–13366.

- Targett-Adams, P. and McLauchlan, J. 2005. Development and characterization of a transient-replication assay for the genotype 2a hepatitis C virus subgenomic replicon. *Journal of General Virology*. [Online]. **86**(11),pp.3075–3080.
- Terenin, I.M., Smirnova, V. V., Andreev, D.E., Dmitriev, S.E. and Shatsky, I.N. 2017. A researcher's guide to the galaxy of IRESs. *Cellular and Molecular Life Sciences*. [Online]. **74**(8),pp.1431–1455.
- Thibault, P.A., Huys, A., Amador-Cañizares, Y., Gailius, J.E., Pinel, D.E. and Wilson, J.A. 2015. Regulation of Hepatitis C Virus Genome Replication by Xrn1 and MicroRNA-122 Binding to Individual Sites in the 5' Untranslated Region. *Journal of virology*. **89**(12),pp.6294–311.
- Thomas, D.L., Villano, S. a, Riester, K. a, Hershov, R., Mofenson, L.M., Landesman, S.H., Hollinger, F.B., Davenney, K., Riley, L., Diaz, C., Tang, H.B. and Quinn, T.C. 1998. Perinatal transmission of hepatitis C virus from human immunodeficiency virus type 1-infected mothers. Women and Infants Transmission Study. *The Journal of infectious diseases*. [Online]. **177**(6),pp.1480–1488.
- Thompson, A., Sch??fer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T. and Hamon, C. 2003. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical Chemistry*. [Online]. **75**(8),pp.1895–1904.
- Thompson, S.R., Gulyas, K.D. and Sarnow, P. 2001. Internal initiation in *Saccharomyces cerevisiae* mediated by an initiator tRNA/eIF2-independent internal ribosome entry site element. *Proceedings of the National Academy of Sciences*. **98**(23),pp.12972–12977.
- Unbehaun, A., Borukhov, S.I., Hellen, C.U.T. and Pestova, T. V. 2004. Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes and Development*. [Online]. **18**(24),pp.3078–3093.
- Vasa, S.M., Guex, N., Wilkinson, K.A., Weeks, K.M. and Giddings, M.C. 2008. ShapeFinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis. *RNA*. **14**(10),pp.1979–1990.
- Vieyres, G., Thomas, X., Descamps, V., Duverlie, G., Patel, A.H. and Dubuisson, J. 2010. Characterization of the envelope glycoproteins associated with infectious hepatitis C virus. *Journal of virology*. **84**(19),pp.10159–68.
- Walters, B. and Thompson, S.R. 2016. Cap-Independent Translational Control of Carcinogenesis. *Frontiers in Oncology*. **6**(May),p.128.
- Wang, C., Sarnow, P. and Siddiqui, A. 1994. A conserved helical element is essential for internal initiation of translation of hepatitis C virus RNA. *Journal of virology*. **68**(11),pp.7301–7.
- WHO 2017. *Global hepatitis report, 2017* [Online].
<http://www.who.int/hepatitis/publications/global-hepatitis-report2017/en/>
- Wilkinson, K.A., Gorelick, R.J., Vasa, S.M., Guex, N., Rein, A., Mathews, D.H., Giddings, M.C. and Weeks, K.M. 2008. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biology*. [Online].

6(4),pp.883–899.

- Wilkinson, K.A., Vasa, S.M., Deigan, K.E., Mortimer, S.A., Giddings, M.C. and Weeks, K.M. 2009. Influence of nucleotide identity on ribose 2'-hydroxyl reactivity in RNA. *Rna*. **15**(7),pp.1314–1321.
- Wilson, J.E., Powell, M.J., Hoover, S.E. and Sarnow, P. 2000. Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. *Molecular and cellular biology*. [Online]. **20**(14),pp.4990–4999.
- Witteveldt, J., Blundell, R., Maarleveld, J.J., McFadden, N., Evans, D.J. and Simmonds, P. 2013. The influence of viral RNA secondary structure on interactions with innate host cell defences. *Nucleic acids research*.,pp.1–16.
- Wozniak, A.L., Griffin, S., Rowlands, D., Harris, M., Yi, M., Lemon, S.M. and Weinman, S.A. 2010. Intracellular Proton Conductance of the Hepatitis C Virus p7 Protein and Its Contribution to Infectious Virus Production C. M. Rice, ed. *PLoS Pathogens*. **6**(9),p.e1001087.
- Wyld, R., Robertson, J.R., Brettell, R.P., Mellor, J., Prescott, L. and Simmonds, P. 1997. Absence of hepatitis C virus transmission but frequent transmission of HIV-1 from sexual contact with doubly-infected individuals. *Journal of Infection*. [Online]. **35**(2),pp.163–166.
- Yamamoto, H., Unbehaun, A., Loerke, J., Behrmann, E., Collier, M., Bürger, J., Mielke, T. and Spahn, C.M.T. 2014. Structure of the mammalian 80S initiation complex with initiation factor 5B on HCV-IRES RNA. *Nature Structural & Molecular Biology*. **21**(8),pp.721–7.
- Yang, Y., Zengel, J., Sun, M., Sleeman, K., Timani, K.A., Aligo, J., Rota, P., Wu, J. and He, B. 2015. Regulation of Viral RNA Synthesis by the V Protein of Parainfluenza Virus 5. *Journal of Virology*. **89**(23),pp.11845–11857.
- Yi, M. and Lemon, S.M. 2003. 3' nontranslated RNA signals required for replication of hepatitis C virus RNA. *Journal of virology*. **77**(6),pp.3557–68.
- Yoon, S., Kim, J., Hum, J., Kim, H., Park, S., Kladwang, W. and Das, R. 2011. HiTRACE: high-throughput robust analysis for capillary electrophoresis. *Bioinformatics (Oxford, England)*. **27**(13),pp.1798–805.
- Yu, M., Peng, B., Chan, K., Gong, R., Yang, H., Delaney, W. and Cheng, G. 2014. Robust and persistent replication of the genotype 6a hepatitis c virus replicon in cell culture. *Antimicrobial Agents and Chemotherapy*. [Online]. **58**(5),pp.2638–2646.
- Zhang, J., Nguyen, D. and Hu, K.-Q. 2016. Chronic Hepatitis C Virus Infection: A Review of Current Direct-Acting Antiviral Treatment Strategies. *North American journal of medicine & science*. **9**(2),pp.47–54.
- Ziehler, W.A. and Engelke, D.R. 2001. Probing RNA structure with chemical reagents and enzymes. *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.]*. **Chapter 6**,p.Unit 6.1.
- Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*. [Online]. **31**(13),pp.3406–3415.

Chapter 8 Appendices

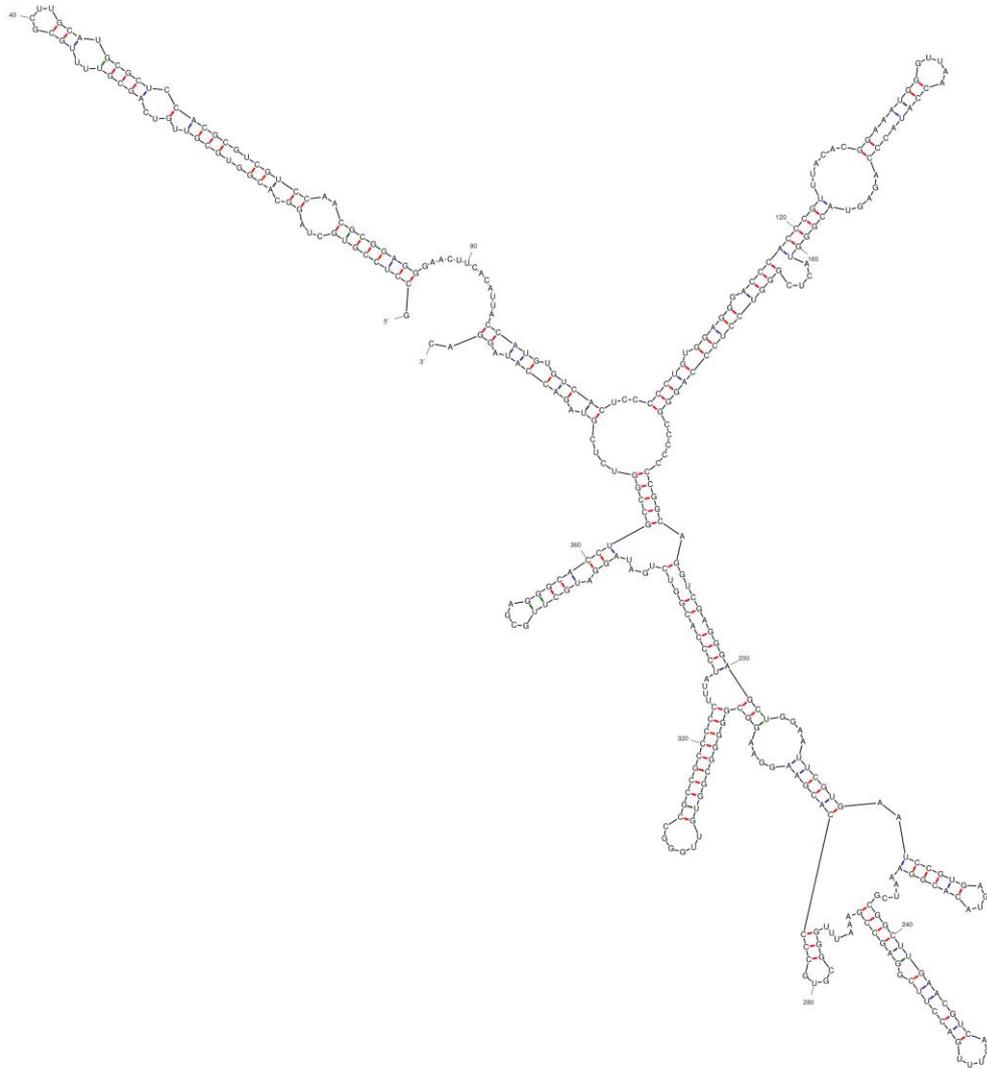


Figure 8.1: Mfold predicted structure of the EqHV 5'UTR

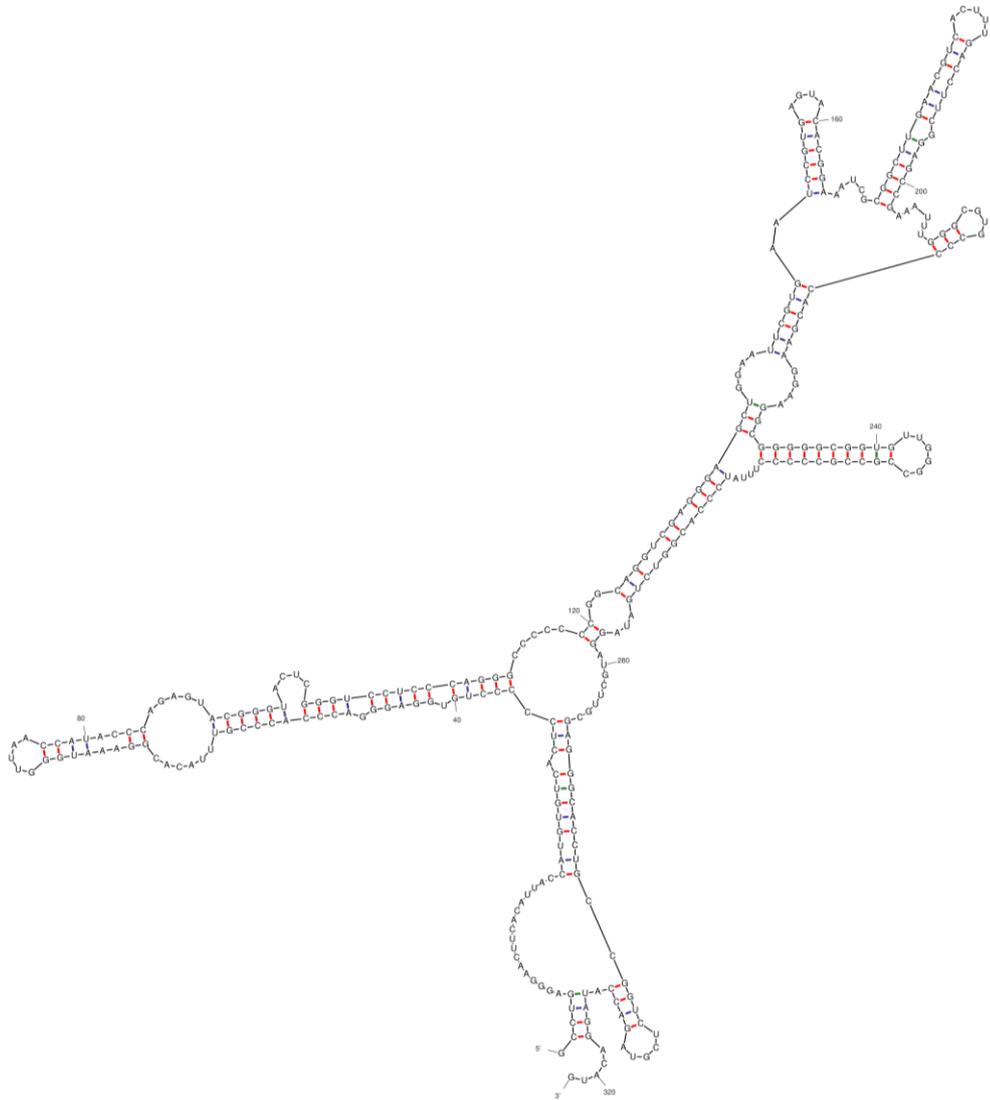


Figure 8.2: Mfold predicted structure of the EqHV Δ SLI 5'UTR

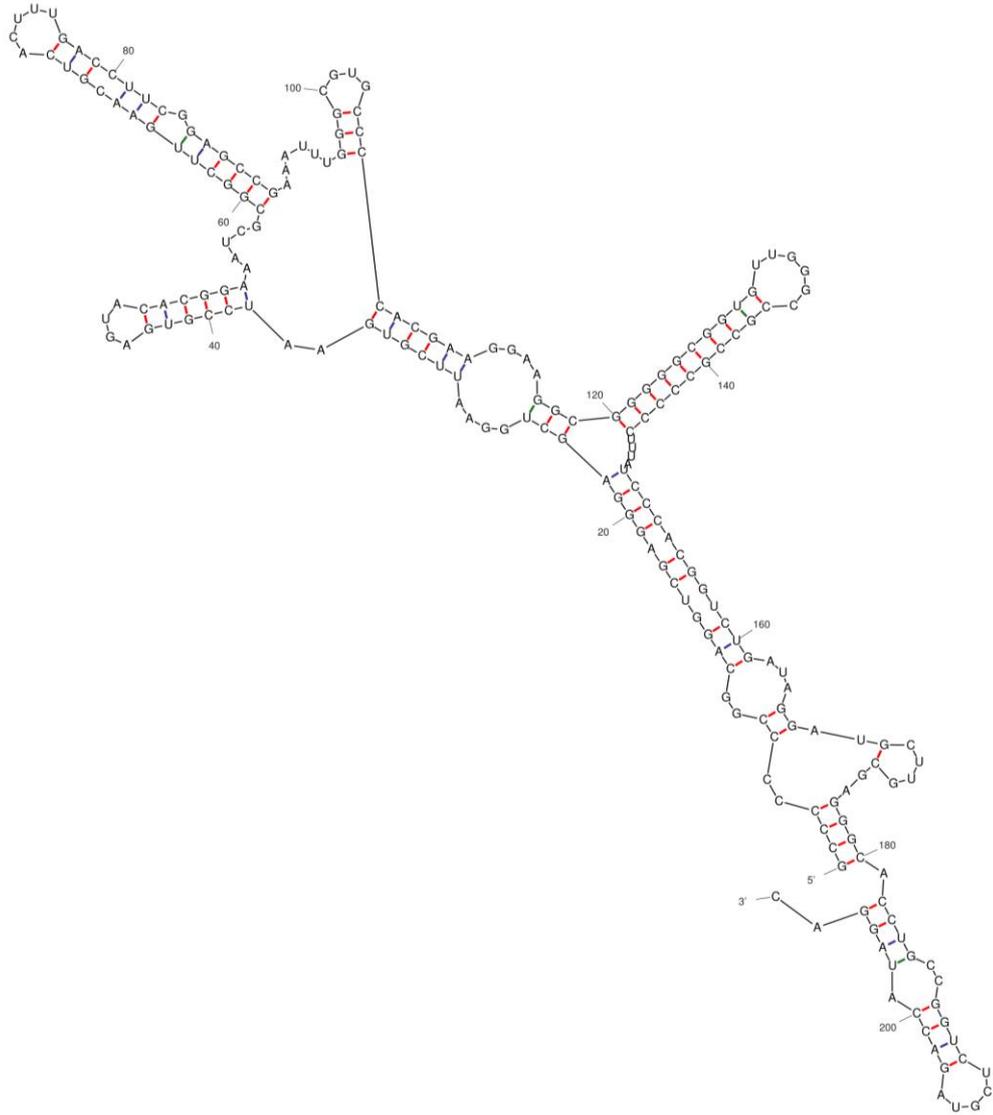


Figure 8.3: Mfold predicted structure of the EqHV Δ SLI+II 5'UTR

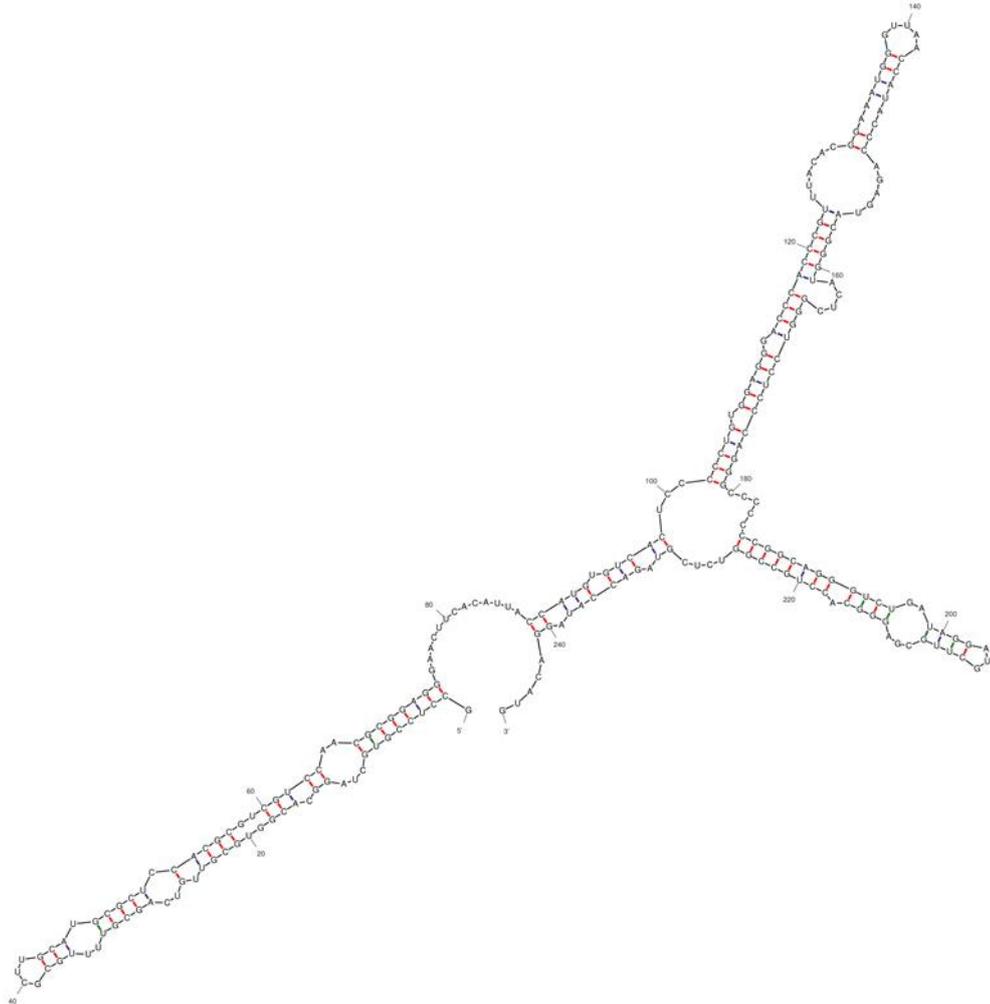


Figure 8.4: Mfold predicted structure of the EqHV ΔSLIII 5'UTR

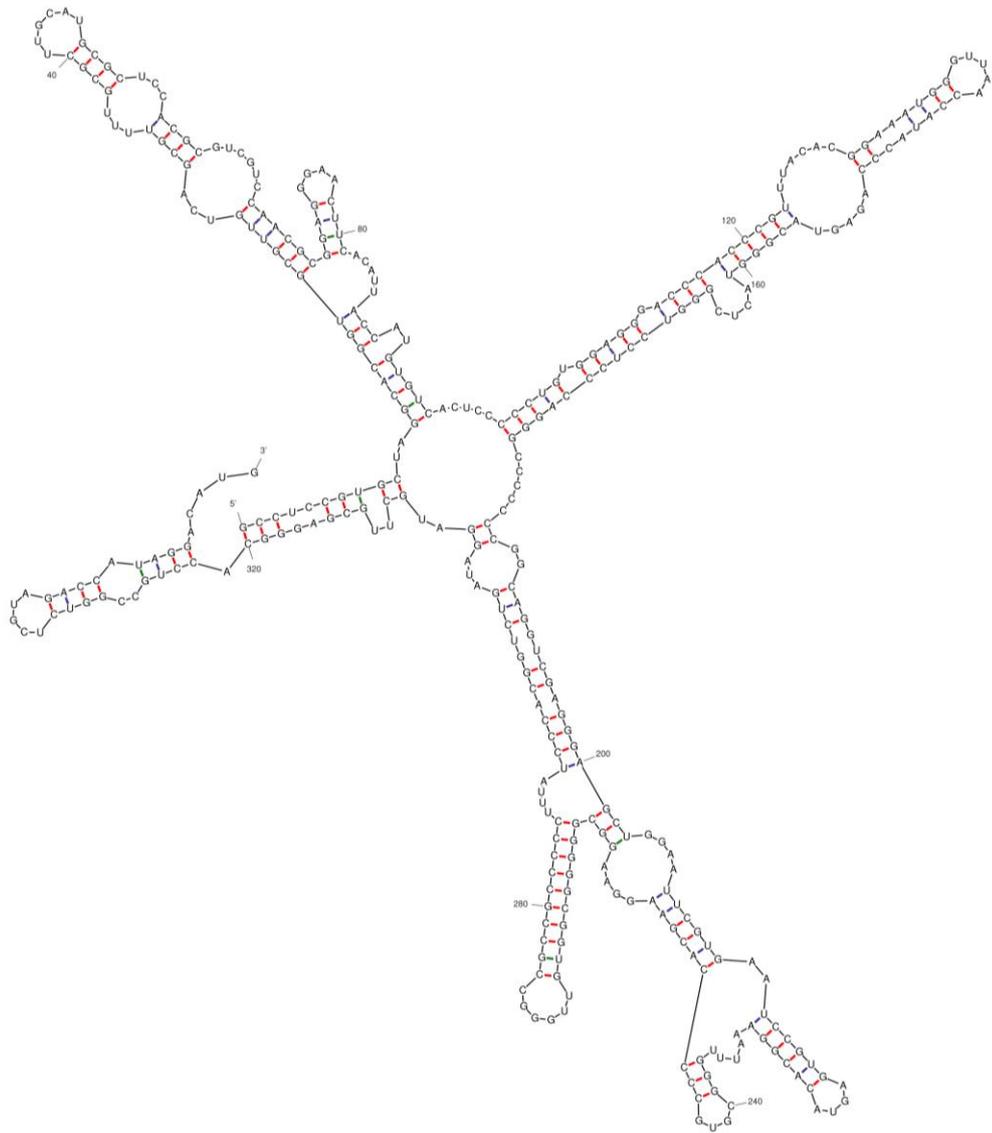


Figure 8.5: Mfold predicted structure of the EqHV Δ SLIIIb 5'UTR

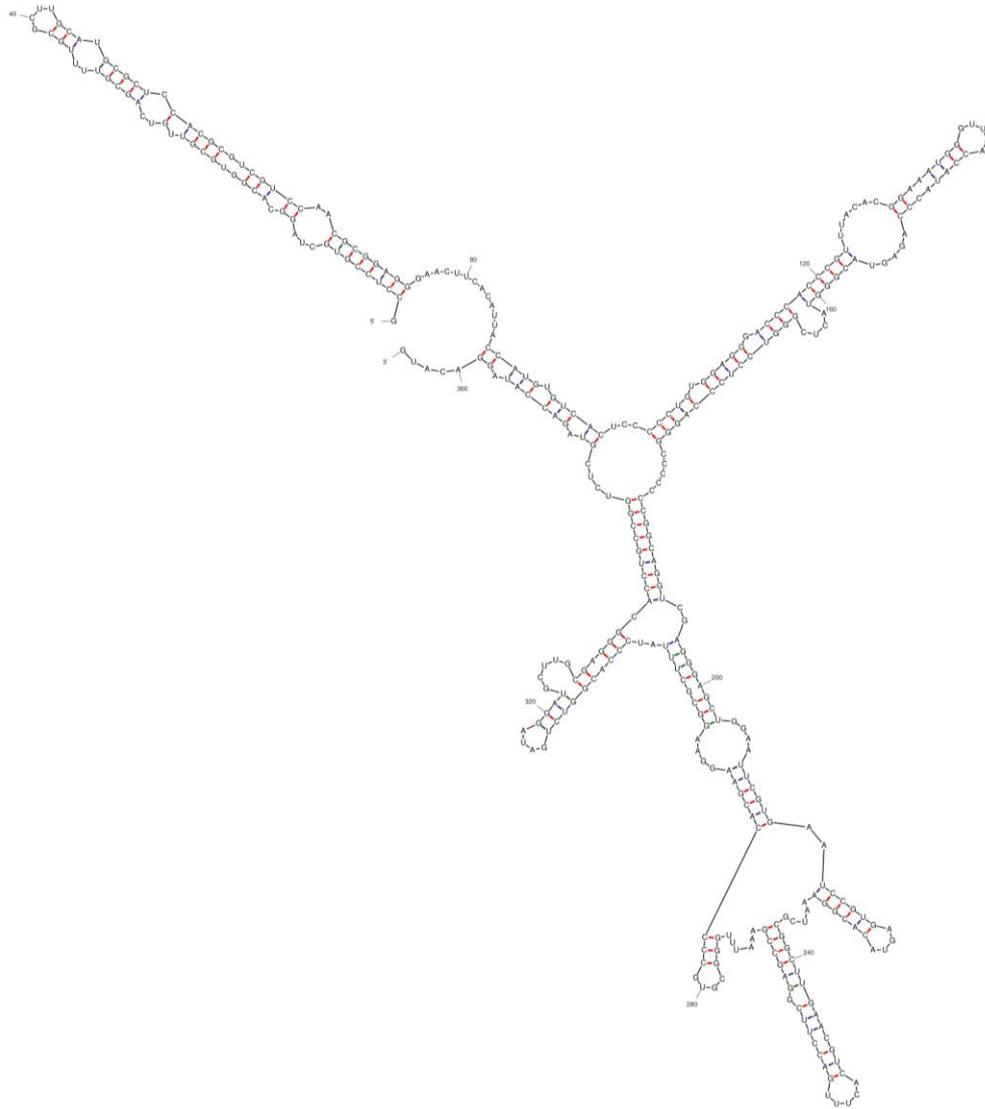


Figure 8.6: Mfold predicted structure of the EqHV Δ SLIId 5'UTR

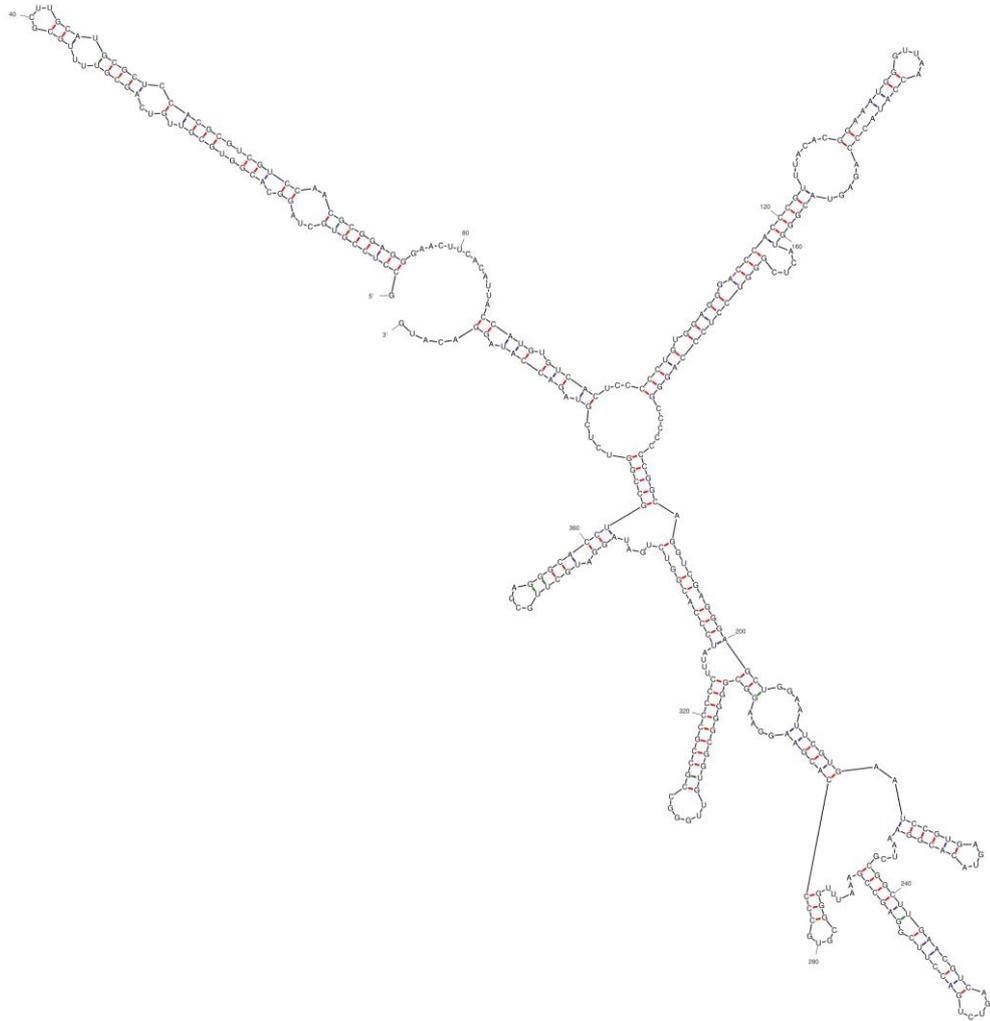


Figure 8.7: Mfold predicted structure of the EqHV GUC 5'UTR



Figure 8.8: Mfold predicted structure of the EqHV AGU 5'UTR

Bicistronic: CTTATGCAGTTGCTCTCCAG
SGR NZCI: GTTCCATCCTCCAGAGGATAGAAT

Figure 8.9: SHAPE primer sequences

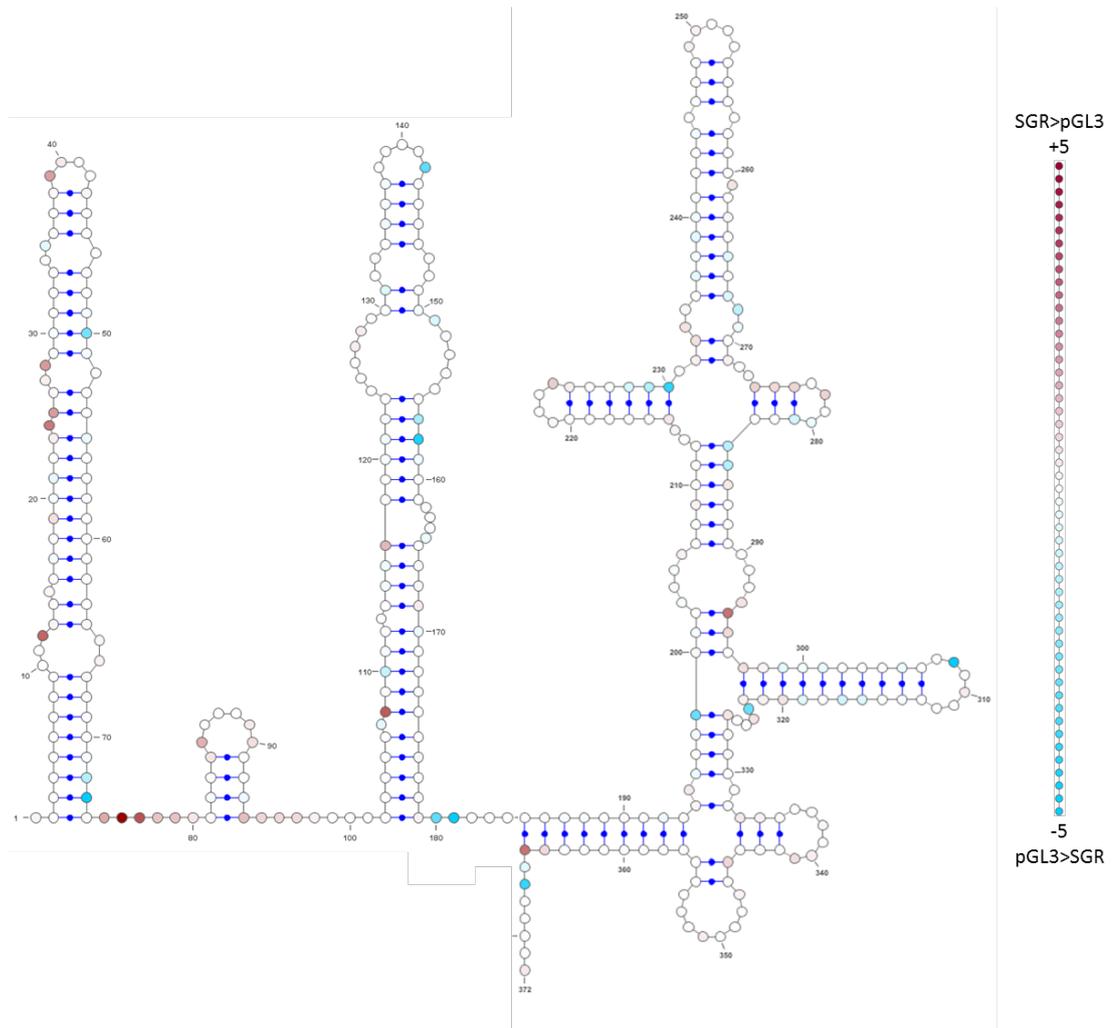


Figure 8.10: Comparison of SHAPE reactivity data between SGR NZCI and pGL3 NPHV

SHAPE reactivity data for pGL3 WT was subtracted from that of SGR NZCI to create a representative reactivity profile which displays the differences in reactivity between the NPHV 5'UTR in the context of the replicon, or bicistronic construct. Red represents those nucleotides where the SGR exhibited higher reactivity and blue where reactivity was higher in the bicistronic construct.

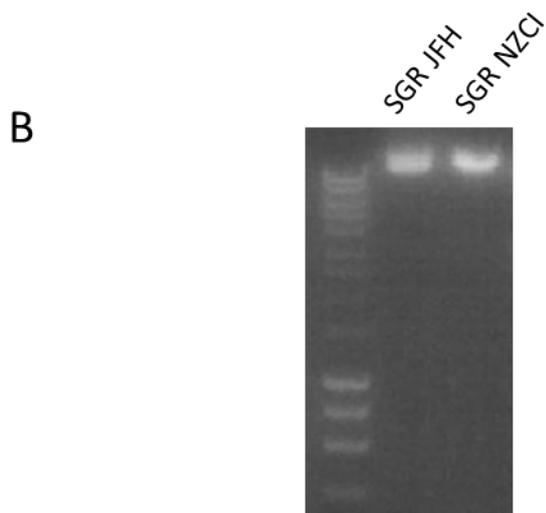
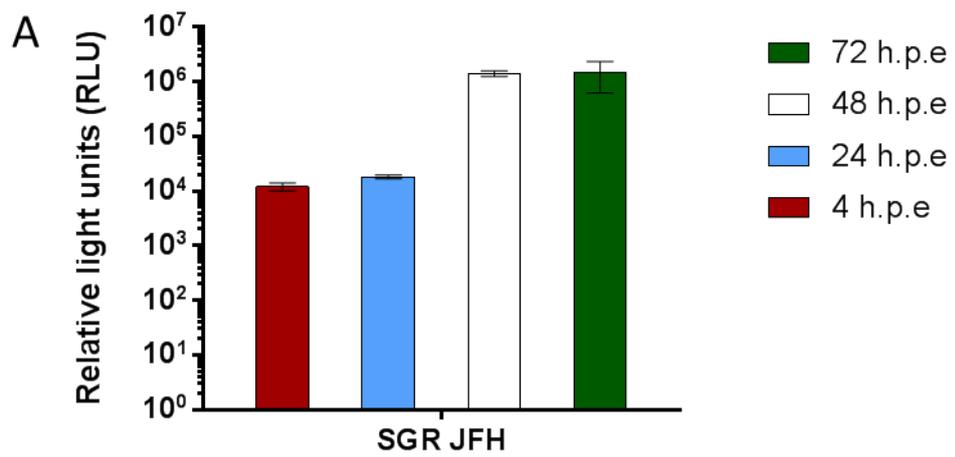


Figure 8.11 A representative SGR JFH Replication assay and RNA gel

A) SGR JFH (2 μ g) was electroporated into Huh7s and assayed for luciferase activity up to 72 h.p.e. B) A representative MOPS gel showing 0.5 μ g of IVT SGR JFH and SGR NZCI

```

HCV          1 TGAGTGGTATACACTCCAATTCTGTTTTTTTTTTTTTTTTTTTTTTTTT 50
SMKL         1 ----- 0
HCV          51 TTTTTTTTTTTTTTTTTTTTTTTTCTTCTTAACCTTCTATCATCTT 100
SMKL         1 ----- 0
HCV          101 ATATCCTTCTT-----AAGGTGGCTCCATCT 126
           |                               ||..||..||..||
SMKL         1 -----TAAAAAAAAAAAAAAAAAAAAAAAAAACATGTCGCCGCCT 40
HCV          127 TAGCCCTAGTCACGGCTAGCTGTGAAAGGTCCG--TG-----AGC 164
           ..||| .||..|| ||..||..|| ||| || |||
SMKL         41 CGGCC---GCATGACT-GCGGAGAGAGG-CCGCATGACTGCAGAGAAGC 84
HCV          165 CGCTTGACTGCAGAGA-GTGCTGTA--ACTG-----GCCTCTCTGC 202
           |||..|||||..||||| ||..||| ||| ||| ||..|||
SMKL         85 CGCATGACTGCAGAGAGGGGCGTATGACTGCAGAGAGCCGCATGACTGC 134
HCV          203 AGATCAAGT----- 211
           |||..|||
SMKL         135 AGAGAGAGGCCGCATGACTGCAGAGAGGGGCTGCATGACTGCAGAGAGCC 184
HCV          212 ----- 211
SMKL         185 GCATGACTGCAGAAAGAGGCCGCATGACTGCAGAGAGCTTGGAAAGCGGAG 234
HCV          212 ----- 211
SMKL         235 GCCGCATGACTGCAGAGAGAGTCTCCGCTTCGGGCTTGGAA 276

```

Figure 8.12: Alignment of the EqHV 3'UTR (accession number JX948116.1 [denoted SMKL]) and the HCV 3'UTR (accession number AF011753.1)

```

HCV          1 ----- 0
NZPI         1 TAAAAAAGAAAAATAATTAGCTCCTAATTCATTCTTCTTCTCCCT 50
HCV          1 -----TG 2
           ||
NZPI         51 CTTTATTTCTTTATTGTTACTTCTATGGAAGAACAGGAGGGTGGGTG 100
HCV          3 AGTGGTA-----TACACTCAA-----TT 21
           | |||. | |..|. |. |
NZPI         101 A-TGGGAGCCCTGTTCGCCCTATGGGGCGAAAATGTTTTTTTTTTTT 149
HCV          22 CTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 71
           .|.|||||..|||||..|||||..|||||..|||||..|||||..
NZPI         150 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 199
HCV          72 TTTTCTTCTTAACCTTCTATCATCTTATATCCTTCTTAA-----GG 115
           |||||..|||..||..||..||..||..||..||..||..||
NZPI         200 TTTTTTTTTTTTTTTTTTTTTTTTTTTTCTTCTTCTATTGATGGG 249
HCV          116 TGGCTCCATCTTAGCCCTAGTCACGGCTAGC-TGTGAAAGGTCCGTGAGC 164
           ||||| |..|||||..|||||..|||||..|| | |||..|||||
NZPI         250 TGGCTCC-CCTTAGCTCTAGTCACGGCTAGCTTCTG-AAGGCCCGTGAGC 297
HCV          165 CGCTTGACTGCAGAGAGTGTGTAAGTGGCCTCTCTGCAGATCAAGT 211
           |||..| |..|. | |..|. |. |
NZPI         298 CGCATG-GTCCCGGGA----TATCCCGGACTATGT----- 328

```

Figure 8.13: Alignment of the EqHV NZPI 3'UTR (accession number KP325401) and the HCV 3'UTR (accession number AF011753.1)

```

SMKL      1 TAAAAAAAAAAAAAAAAAAAAA-----AAAACA---TGTCGCCGC      38
           .|||||.|||||.||      ||..||  |.|||.||
NZPI      1 -----TAAAAAAGAAAAATAATTAGCTCCTAATTCATTCTTCTCTTC      44

SMKL     39 CTCGGCCGCATGACTGC-----GGAGAGAGGCCGCATGAC-----TG      75
           .||  |||.|||.|||.||      ||      |.||  ||
NZPI     45 TTC--CCTCTTATTTCTTTATTGG-----TTACTTCCTATG      80

SMKL     76 CAGAGAAGCCGCATGACTGCAGAGAGGGGCGTATGACTGCAGAGAG---      122
           .|  ||||      |||  |||||      ||..|||.|||.|||
NZPI     81 GA-AGAA-----CAG-GAGGG-----TGGGTGATGGGAGCCC      110

SMKL    123 ----CCGC----ATGACTGCAGAGAGAGGCCGCA-----          148
           ||||  |||      |||.|||
NZPI    111 TGTTCCGCCCTATG-----GGCGAAAAATGTTTTTTTTTTTTT      148

SMKL    149 -----          148

NZPI    149 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT      198

SMKL    149 -----TGACTGCAGAGAGGG          163
           |..|||.|||.|||
NZPI    199 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTTCTCTATTGATGG      248

SMKL    164 GCTGCATGACTGCAGAGAGCCGCATGACTGCAGAAAAGAGGCCGCATGACT      213
           |  ||.||  |||.|||.|||.||      |.||.
NZPI    249 G----TGGCT-----CCCCTTAGCTCTAG-----TCACG      273

SMKL    214 GCAGAGAGCTT-GGAAG---CG-GAGGCCGCATGACTGCAGAGAGAGTCC      258
           ||  .||||| .||||  ||  |||  |||||  ||||
NZPI    274 GC---TAGCTTCTGAAGGCCCGTGA-GCCGCATG-----GTCC      307

SMKL    259 -----TCCGCTTCGGG-CT-TGGAA      276
           |||  ||||  |||.
NZPI    308 CGGGATATCC-----CGGGACTATGT--      328

```

Figure 8.14: Alignment of the “original” EqHV 3’UTR (accession number JX948116.1 [denoted SMK]) and the consensus EqHV NZPI 3’UTR (accession number KP325401)