

A structural and functional analysis of  
opal stop codon read-through during  
Chikungunya virus replication

Raymond Li

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

The University of Leeds  
Faculty of Biological Sciences  
School of Molecular and Cellular Biology

September 2019

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.

The right of Raymond Li to be identified as Author of this work has been asserted by Raymond Li in accordance with the Copyright, Designs and Patents Act 1998.

## **Acknowledgments**

I would like to thank my supervisors Prof. Mark Harris and Dr. Andrew Tuplin for their excellent supervision and support throughout my research project. Their patience and advice inspired me to work to the best of my ability throughout my time at Leeds. I want to thank the past and present members of the Garstang 8.61 lab and the Virology research group for making me feel part of a community of talented and passionate individuals.

I especially want to thank Christian, Ollie, Amelia and Mari for their friendship and support, their kindness was a beacon during the most stressful times during my project. I also want to thank Carsten for keeping the lab running even at the busiest of times. I feel very lucky to have started my PhD alongside Yanni Gao who was always a joy to work with in the lab. I thank my family and amazing partner who have encouraged, motivated and supported me during my PhD.

I thank the Microbiology Society for giving me the chance to present and discuss my work with experts from the scientific community. I am very grateful for the University Research Scholarship awarded to me by University of Leeds to fund this project. I want to thank the Faculty of Biological Sciences for helping me throughout my time at Leeds.

## **Abstract**

Chikungunya virus (CHIKV) is a positive-sense single-stranded RNA virus belonging to the alphavirus genus within the *Togaviridae* family. It is transmitted by *Aedes* species mosquitoes. Whereas CHIKV infection in mosquitoes is persistent and symptomless, the virus causes an acute onset of muscle and joint pain, fever, nausea and headache in humans. Most infected individuals suffer from persistent chronic joint and muscle pain after the initial infection is cleared. CHIKV has been reported in over 50 countries worldwide, as yet there are no specific antiviral agents or vaccines available.

The CHIKV genome contains two open reading frames. The 5' of these encodes a polyprotein that is cleaved to 4 non-structural proteins (nsP1-4). The third of these (nsP3) is essential for virus replication but the exact function is unknown. Some strains of CHIKV contain an in-frame opal stop codon near the C-terminus of nsP3. Translational termination at this position would result in expression of a C-terminally truncated nsP3 and a lack of nsP4 (the viral RNA dependant RNA polymerase). Programmed read-through of this stop codon is necessary to express all four nsPs.

The advantages conferred by possession of the opal stop codon are unknown. To test this a panel of CHIKV subgenomic replicons and infectious clones with either the opal, ochre or amber stop codon in place of the naturally occurring arginine codon were generated. The presence of a stop codon reduced CHIKV replication in mammalian cells but enhanced replication in mosquito cells. SHAPE was used to determine the RNA structure surrounding the stop codon and revealed a large 3' stem-loop that may play a role in translational read-through. The data suggests that the stop codon has species-specific functionality, in particular that the stop codon read-through signal may function more efficiently in the invertebrate host.



## Contents

Acknowledgments.....	2
Abstract.....	3
Contents .....	4
List of Figures .....	9
List of tables.....	11
Abbreviations .....	12
Chapter 1: Introduction .....	16
1.1 Introduction to alphaviruses.....	16
1.2 Chikungunya virus .....	16
1.2.1 Introduction to CHIKV.....	16
1.2.1 Pathology of CHIKV.....	17
1.2.3 Epidemiology of CHIKV .....	18
1.3 Molecular biology of CHIKV.....	20
1.3.1. Genome organisation .....	20
1.3.2 Non-structural proteins .....	21
1.3.2.1 nsP1 .....	21
1.3.2.2 nsP2 .....	22
1.3.2.3 nsP3 .....	22
1.3.2.4 nsP4 .....	24
1.3.3 Structural proteins .....	25
1.3.3.1 Capsid .....	25
1.3.3.2 E1 .....	25
1.3.3.3 E2.....	25
1.3.3.4 E3.....	26
1.3.3.5 6K.....	26
1.3.4 Viral lifecycle .....	26

1.4 Differences in insect and mammalian host responses to CHIKV infection .....	29
1.5 Translation.....	30
1.5.1 Mechanism of eukaryotic translation .....	30
1.5.2 Stop codon read-through.....	37
1.5.3 Opal stop codon read-through in the CHIKV nsP3 .....	38
1.5.4 Potential mechanism of stop codon read-through .....	41
1.5.5 Pharmacological modulation of stop codon read-through .....	44
1.5.6 Other virus mechanisms of recoding translation.....	45
1.6 Aims and objectives of the research.....	45
Chapter 2: Materials and Methods.....	47
2.1 DNA and RNA manipulation .....	47
2.1.1 Plasmid and virus constructs.....	47
2.1.2 Cloning strategy .....	47
2.1.3 Replication of plasmids in bacteria.....	48
2.1.4 DNA and RNA Quantification .....	49
2.1.5 DNA and RNA agarose gel electrophoresis .....	49
2.1.6 Linearisation and purification of DNA .....	50
2.1.7 <i>In vitro</i> transcription of RNA.....	51
2.1.8 Site directed mutagenesis .....	51
2.1.9 Sub-cloning and ligation .....	52
2.2 Tissue culture .....	52
2.2.1 Cell lines.....	52
2.2.2 Passaging cell lines.....	53
2.2.3 Transfection using lipofectamine 2000 .....	53
2.2.4 Electroporation .....	54
2.2.5 Generating and maintaining a cell line stably expressing the non-cytotoxic CHIKV replicon.....	54

2.2.6 Protein extraction .....	55
2.3 Molecular Biology .....	55
2.3.1 Dual luciferase reporter assay .....	55
2.3.2 DNA sequencing .....	56
2.3.3 RNA extraction .....	56
2.3.4 Reverse transcription PCR .....	56
2.3.5 Quantification of protein concentration .....	57
2.3.6 Streptactin column affinity purification .....	57
2.3.7 Streptactin batch purification .....	58
2.3.8 Protein separation using SDS-PAGE gels .....	58
2.3.9 Western blot .....	59
2.3.10 Genome alignment .....	59
2.4 Infectious clone of CHIKV experiments .....	60
2.4.1 Generating virus stocks by electroporation .....	60
2.4.2 Generating virus stocks using Lipofectamine 2000 transfection .....	60
2.4.3 Plaque assay .....	60
2.4.4 Infectious centre assay (ICA) .....	61
2.4.5 Virus infection for protein extraction .....	61
2.4.6 One step growth curve .....	62
2.4.7 RNA extraction from transfected cells .....	62
2.5 Nucleic acid structure determination .....	63
2.5.1 Mfold prediction of RNA structure .....	63
2.5.2 <i>In vitro</i> transcription of uncapped ICRES CHIKV RNA .....	63
2.5.2 QuSHAPE .....	63
2.5.3 RNAstructure program prediction .....	65
Chapter 3: Effect of opal stop codon read-through on CHIKV replication .....	66
3.1 Introduction .....	66
3.2 Results .....	69

3.2.1 The effect of stop codon read-through on CHIKV SGR replication in non-human mammalian cells .....	69
3.2.2 The effects of stop codon read-through in CHIKV replication in human cell types .....	72
3.2.3 The effects of stop codon read-through on CHIKV replication in C6/36 (mosquito) cells .....	77
3.2.4 The effect of stop codon read-through on CHIKV infectivity in BHK cells .....	79
3.2.5 The effect of stop codon read-through on CHIKV infectivity in C6/36 cells .....	85
3.2.6 Sequencing of stop codon ICRES CHIKV produced in BHK and C6/36 cells .....	87
3.2.7 One step growth curve of RD and BHK cells infected with stop codon infectious clone CHIKV.....	89
3.3 Discussion .....	91
Chapter 4: The role of RNA structure on opal stop codon read-through in the context of CHIKV replication .....	95
4.1 Introduction.....	95
4.2 Results.....	102
4.2.1 The read-through signal RNA structure was detected in the full-length CHIKV genome RNA when folded at 37°C.....	102
4.2.2 The stop codon read-through signal RNA structure forms when the full-length CHIKV genome is folded at 28°C .....	108
4.2.3 Disrupting the stop codon read-through recoding RNA structure disrupts the production of CHIKV in BHK cells.....	116
4.2.4 Disrupting the stop codon read-through recoding RNA structure does not affect the production of CHIKV in the C6/36 mosquito cells .....	122
4.3 Discussion .....	126
Chapter 5: Developing an affinity tagged nsP3 pulldown system .....	129
5.1 Introduction.....	129

5.2 Results.....	131
5.2.1 Tagged nsP3 was successfully expressed using CHIKV SGR.....	131
5.2.2 Untagged and TST nsP3 was expressed using a non-cytotoxic replicon in Huh7 cells.....	134
5.2.3 The TST was successfully incorporated into the reading frame of nsP3 of an infectious clone of CHIKV.....	141
5.3 Discussion .....	145
Chapter 6: Conclusion and future perspectives .....	146
6.1 Replication of stop codon containing CHIKV mutants displayed a host/ vector specific phenotype.....	146
6.2 The alphavirus stop codon recoding RNA structure folds at temperatures corresponding with the host and vector body temperature.....	149
6.3 An infectious clone of CHIKV possessing an affinity tagged TST-nsP3 was successfully generated.....	152
6.4 Conclusion and future prospects .....	153
References.....	155
Appendix .....	168

## List of Figures

<b>FIGURE 1: REPORTED CASES OF CHIKV BY COUNTRIES OR TERRITORIES AS OF 29.05.18.</b>	19
<b>FIGURE 2: THE CHIKV GENOME IS TRANSLATED IN TWO POLYPROTEINS, THE FIRST EXPRESSES THE NSPS AND THE SECOND EXPRESSES THE STRUCTURAL PROTEINS.</b>	20
<b>FIGURE 3: A DIAGRAM VISUALLY DEPICTING THE KNOWN FUNCTIONS OF THE CHIKV NSP1 TO NSP4.</b>	24
<b>FIGURE 4: A DIAGRAM OF THE CHIKV LIFECYCLE.</b>	28
<b>FIGURE 5: EUKARYOTIC TRANSLATION INITIATION.</b>	31
<b>FIGURE 6: EUKARYOTIC TRANSLATION ELONGATION.</b>	33
<b>FIGURE 7: EUKARYOTIC TRANSLATION TERMINATION.</b>	36
<b>FIGURE 8: POTENTIAL STOP CODON READ-THROUGH MECHANISM.</b>	43
<b>FIGURE 9: SEQUENCES OF CHIKV ISOLATES AND RELATED ALPHAVIRUSES WERE ALIGNED AGAINST THE ICRES ECSA STRAIN OF CHIKV.</b>	67
<b>FIGURE 10: STOP CODONS WERE SUBSTITUTED INTO THE ECSA DERIVED DLUC CHIKV SGR CONSTRUCT.</b>	68
<b>FIGURE 11: EFFECT OF STOP CODON IN NSP3 ON CHIKV GENOME REPLICATION IN MURINE MOUSE MUSCLE CELLS (C2C12).</b>	70
<b>FIGURE 12: EFFECT OF STOP CODON IN NSP3 ON CHIKV GENOME REPLICATION IN BHK CELLS.</b>	71
<b>FIGURE 13: EFFECT OF STOP CODON IN NSP3 ON CHIKV GENOME REPLICATION IN HUMAN LIVER CELLS (HUH7).</b>	74
<b>FIGURE 14: EFFECT OF STOP CODON IN NSP3 ON CHIKV GENOME REPLICATION HUMAN GLIAL CELLS (SVG-A).</b>	75
<b>FIGURE 15: EFFECT OF STOP CODON IN NSP3 ON CHIKV GENOME REPLICATION IN HUMAN MUSCLE CELLS (RD).</b>	76
<b>FIGURE 16: EFFECT OF STOP CODON IN NSP3 ON CHIKV GENOME REPLICATION IN AE. ALBOPICTUS CELLS (C6/36).</b>	78
<b>FIGURE 17: STOP CODON MUTANTS OF INFECTIOUS CLONE CHIKV (ICRES) AND THE FLOW DIAGRAMS DEPICTING THE PLAQUE ASSAY AND ICA PROTOCOLS.</b>	81
<b>FIGURE 18: EFFECT OF STOP CODON READ-THROUGH ON ICRES CHIKV VIRUS PRODUCTION AND RNA INFECTIVITY IN BHK CELLS.</b>	83
<b>FIGURE 19: EFFECT OF STOP CODON READ-THROUGH ON CHIKV PRODUCTION IN C6/36 CELLS AND NSP1 AND NSP3 EXPRESSION IN BHK AND RD CELLS.</b>	86
<b>FIGURE 20: STOP CODON MUTATIONS WERE MAINTAINED IN THE BHK AND C6/36 EXTRACTED RNA SAMPLES.</b>	88
<b>FIGURE 21: THE EFFECT OF STOP CODONS ON CHIKV GROWTH IN RD AND BHK CELLS OVERTIME.</b>	90
<b>FIGURE 22: THREE TYPES OF READ-THROUGH SIGNALS HAVE BEEN IDENTIFIED.</b>	96
<b>FIGURE 23: ALPHAVIRUS STOP CODON READ-THROUGH SIGNALS.</b>	97

<b>FIGURE 24: THE BOTTOM STEM OF THE RNA STRUCTURE READ-THROUGH SIGNAL IDENTIFIED IN VEEV AND SINV WAS IDENTIFIED IN CHIKV SEQUENCES.</b>	<b>98</b>
<b>FIGURE 25: THE QUSHAPE EXPERIMENTAL PROCEDURE.</b>	<b>101</b>
<b>FIGURE 26: THE READ-THROUGH SIGNAL RNA STRUCTURE WAS PREDICTED TO FOLD WHEN SIMULATED USING MFOLD (37°C).</b>	<b>103</b>
<b>FIGURE 27: THE PREDICTED RNA STRUCTURE DOWNSTREAM OF THE OPAL STOP CODON WAS VALIDATED USING QUSHAPE.</b>	<b>105</b>
<b>FIGURE 28: SHAPE REACTIVITY ADJUSTED STRUCTURAL PREDICTION OF THE READ-THROUGH RNA STRUCTURE IN THE CHIKV GENOME RNA FOLDED AT 37°C.</b>	<b>107</b>
<b>FIGURE 29: THE READ-THROUGH SIGNAL RNA STRUCTURE WAS SHOWN TO FOLD IN THE CHIKV SEQUENCE WHEN FOLDED AT 28°C.</b>	<b>109</b>
<b>FIGURE 30: THE RNA STRUCTURE PREDICTED TO FOLD AT 28°C WAS VALIDATED USING QUSHAPE.</b>	<b>111</b>
<b>FIGURE 31: AN IMPROVED MODEL USING THE SHAPE DATA FOR THE 28°C FOLDED CHIKV RNA STRUCTURE WAS GENERATED.</b>	<b>113</b>
<b>FIGURE 32: COMPARING NMIA REACTIVITIES OF THE 37°C AND 28°C SHAPE REACTION REVEALED SHARED REGIONS OF NMIA REACTIVITY.</b>	<b>115</b>
<b>FIGURE 33: MFOLD PREDICTED STRUCTURES OF DISRUPTED AND COMPENSATED RNA STRUCTURES IN THE CONTEXT OF THE WILDTYPE ICRES CHIKV MUTANTS FOLDED AT 37°C.</b>	<b>118</b>
<b>FIGURE 34: DISRUPTING THE RNA STRUCTURE DECREASES THE RNA INFECTIVITY AND PRODUCTION OF VIRUS FOR THE STOP CODON ICRES CHIKV MUTANTS.</b>	<b>120</b>
<b>FIGURE 35: MFOLD PREDICTED STRUCTURES OF DISRUPTED AND COMPENSATED RNA STEM LOOP IN THE CONTEXT OF THE WILDTYPE ICRES CHIKV WHEN FOLDED AT 28°C.</b>	<b>123</b>
<b>FIGURE 36: DISRUPTING THE RNA STRUCTURE DID NOT AFFECT CHIKV REPLICATION IN C6/36 CELLS.</b>	<b>125</b>
<b>FIGURE 37: AFFINITY TAGS WERE INSERTED INTO THE SG GLUC CHIKV SGR USING THE UNIQUE SPEI RESTRICTION SITE IN THE NSP3 HVD.</b>	<b>131</b>
<b>FIGURE 38: TAGGED NSP3 WAS SUCCESSFULLY EXPRESSED WHEN INCORPORATED INTO THE CODING SEQUENCE OF NSP3 IN THE SG CHIKV GLUC REPLICON.</b>	<b>133</b>
<b>FIGURE 39: STABLE CELL LINE EXPRESSED CHIKV REPLICONS IN HUH7 CELLS PRODUCE LARGE AMOUNTS OF NSP3.</b>	<b>136</b>
<b>FIGURE 40: TST TAGGED NSP3 CAN BE AFFINITY PURIFIED FROM CELL LYSATES.</b>	<b>138</b>
<b>FIGURE 41: TST NSP3 WAS SUCCESSFULLY PURIFIED USING THE BATCH PURIFICATION PROTOCOL.</b>	<b>140</b>
<b>FIGURE 42: AN ICRES CHIKV EXPRESSING A TST NSP3 WAS SUCCESSFULLY GENERATED.</b>	<b>142</b>
<b>FIGURE 43: TST NSP3 EXTRACTED FROM INFECTED HUH7 CELLS WAS PURIFIED BY BATCH PURIFICATION USING STREPTACTIN RESIN.</b>	<b>144</b>

## List of tables

<b>TABLE 1: LIST OF PRIMERS USED DURING THIS PROJECT.....</b>	<b>168</b>
<b>TABLE 2: LIST OF NMIA REACTIVITIES OF THE 37°C FOLDED RNA SHAPE REACTIONS. .....</b>	<b>171</b>
<b>TABLE 3: LIST OF NMIA REACTIVITIES OF THE 28°C FOLDED RNA SHAPE REACTIONS. .....</b>	<b>175</b>
<b>TABLE 4: MFOLD P-NUM VALUES. ....</b>	<b>179</b>
<b>TABLE 5: LIST OF PRIMARY ANTIBODIES.....</b>	<b>180</b>



## Abbreviations

A	Alanine (Ala)
aa	Amino acid
ABCE1	ATP-binding cassette sub-family E member 1
ADP	Adenosine diphosphate
AUD	Alphavirus unique domain
BCA	Bicinchoninic acid
BHK	Baby hamster kidney
BIN1	Amphiphysin 2
bp	Base pairs
BSA	Bovine serum albumin
C	Cysteine (Cys)
cDNA	Complementary DNA
CDC	Centres for disease control and prevention
CIP	Calf intestinal phosphatase
CPE	Cytopathic effect
CPV-I	Cytopathic vacuolar structure type 1
CHIKV	Chikungunya virus
CSE	Conserved sequence element
CTD	C-terminal domain
D	Aspartic acid (Asp)
DEPC	Diethyl pyrocarbonate
Dluc	Dual luciferase
DMD	Duchenne Muscular Dystrophy
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide
ddNTP	Dideoxynucleotide
ds	Double stranded
E (1 – 3)	Envelope protein (1 – 3)
ECSA	East/Central and South African
EDTA	Ethylenediamine tetra-acetic acid
EEEV	Eastern Equine Encephalitis virus
eIF	Eukaryotic initiation factor
eEF	Eukaryotic elongation factor

eRF	Eukaryotic release factor
ER	Endoplasmic reticulum
F	Phenylalanine (Phe)
FBS	Fetal bovine serum
FHS	FLAG-His-SBP tag
F3His	Triple FLAG His tag
FGDF	Phe-Gly-Asp-Phe
Fluc	Firefly luciferase
FXR	Fragile X-related Family
G	Glycine (Gly)
g	Gravitational force
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
G3BP	Ras GTPase-activating protein binding protein
G418	Geneticin
GDP	Guanosine diphosphate
GLB	Glasgow lysis buffer
Gluc	Gaussia luciferase
GTP	Guanosine triphosphate
H	Histidine (His)
HA	Haemagglutinin
HCV	Hepatitis C virus
His-tag	Histidine tag
h.p.i	Hours post infection
h.p.t	Hours post transfection
Huh7	Human hepatocellular carcinoma cell line-7
HVD	Hypervariable domain
I	Isoleucine (Ile)
ICA	Infectious centre assay
ICRES	Integration of Chikungunya RESEARCH
IFN	Interferon
IgM	Immunoglobulin M
kDa	Kilodalton
L	Leucine (L)
LB	Luria Bertani
LiCl	Lithium chloride

M	Middle domain
Mw	Molecular weight
MC	Methyl-cellulose
MAR	Mono-ADP ribose
MeOH	Methanol
Mg <sup>2+</sup>	Magnesium
MOI	Multiplicity of infection
mRNA	Messenger RNA
MuLV	Moloney murine leukaemia virus
NaOAc	Acetic acid
NaOH	Sodium hydroxide
NCT	Non-cytotoxic
Nc-tRNA	Near-cognate tRNA
NEAA	Non-essential amino acids
NMIA	N-methylisatoic anhydride
NsP (1 – 4)	Non-structural protein (1 to 4)
nt	Nucleotide
NTD	N-terminal domain
NTPase	Nucleoside triphosphotase
ONNV	O' Nyong Nyong virus
ORF	Open reading frame
PAGE	Polyacylamide gel electrophoresis
PABP	Poly-(A) binding protein
PARP	Poly (ADP-ribose) polymerase
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PFU	Plaque forming units
PLB	Passive lysis buffer
Pol	Polymerase
PTC	Peptidyl transferase centre
Q	Glutamine (Glu)
R	Arginine (Arg)
RdRp	RNA dependent RNA polymerase
RLU	Relative light units

Rluc	Renilla luciferase
RNA	Ribonucleic acid
RNAi	RNA interference
RRV	Ross River virus
RT-PCR	Reverse transcription PCR
S	Serine (Ser)
SDS	Sodium dodecyl sulphate
S.E.M	Standard error of the mean
SINV	Sindbis virus
SHAPE	Selective 2' hydroxyl acylation and primer extension
SFV	Semliki Forest virus
SG	Subgenomic
SBP	Streptavidin binding protein
SH3	Src homology 3
TAE	Tris Acetate EDTA buffer
TBS	Tris buffered saline
TC	Ternary complex
TEMED	Tetramethylethylenediamine
TF	Transframe
TMV	Tobacco mosaic virus
tRNA	Transfer RNA
TST	Twin-strep tag
UTR	Untranslated region
V	Valine (Val)
v/v	Volume by volume
VEEV	Venezuelan Equine Encephalitis virus
viRNA	Virus derived small interfering RNAs
W	Tryptophan (Trp)
w/v	Weight by volume
WEEV	Western Equine Encephalitis virus

## **Chapter 1: Introduction**

### **1.1 Introduction to alphaviruses**

Alphaviruses are single stranded RNA viruses in the Togaviridae family which typically infect both mosquitoes and mammalian hosts. The only other genus in the Togaviridae family is the rubiviruses, the sole member of this group is Rubella virus (Lambert et al., 2015). Alphavirus infections in mosquitoes are life-long and symptomless but infection in mammalian hosts causes acute infection. The alphaviruses can be classified in two groups: The Old World and New World, the prototype alphavirus is the Old World Sindbis virus (SINV). Alphavirus virions are icosahedral nucleocapsids which contain the approximately 11.7 kilobase positive sense RNA genome, the virion is surrounded by a lipid bilayer envelope. The genome encodes for a polyprotein which expresses the four non-structural proteins (nsPs), a second open reading frame controlled by a subgenomic (SG) promoter expresses the structural polyprotein containing the E1, E2, E3, 6K and capsid (Strauss and Strauss, 1994). Many alphaviruses cause serious problems for the medical services and economies of many countries worldwide. These include the Old World Chikungunya virus (CHIKV), O' Nyong Nyong virus (ONNV) and Ross River virus (RRV) which are found in Africa, Asia, Australia and the Americas. The New World alphaviruses are prevalent in the Americas, these include the Venezuelan Equine Encephalitis Virus (VEEV), Western Equine Encephalitis Virus and Eastern Equine Encephalitis Virus (EEEV). The New World alphaviruses infect horses and humans. New World virus infections are characterised by virally induced encephalitis while Old World alphaviruses are characterised by arthralgia symptoms (Atkins, 2013). The main vectors for alphavirus infection are the *Aedes aegypti* and *Aedes albopictus* mosquitoes, the global distribution of alphaviruses has been attributed to the increased population density and increased movement of people such as tourism and migration (Braack et al., 2018)

### **1.2 Chikungunya virus**

#### **1.2.1 Introduction to CHIKV**

CHIKV is an Old World alphavirus which was first isolated in Tanzania in 1952. There are three strains of the virus, these are the East/Central and South African (ECSA), the West African and the Asian strains. Infection with CHIKV causes a non-fatal febrile illness in infected humans, CHIKV infection can cause persistent

debilitating joint pain that lasts for months after the initial infection (Powers and Logue, 2007). The acute phase of CHIKV infection in humans occurs after 2 to 4 days, this is characterised by a sudden onset of fever, headache, nausea, muscle pain, joint pain and maculopapular rash. The viremia of CHIKV reaches  $10^8$  particles per ml of blood during the acute stage, this stage of CHIKV infection typically lasts a few weeks (Solignat et al., 2009). CHIKV infection progresses on to chronic infection in 40% of cases, this is where the individuals suffer from long term painful persistent joint and muscle pain (Schwartz and Albert, 2010).

Neonates and children are predisposed to suffer from rare cases of severe CHIKV, they suffer from neurological symptoms, encephalitis and seizures. In elderly individuals CHIKV infection can result in fatality due to neurological and pulmonary deterioration and multiple organ failure (Tanabe et al., 2018). There are no licensed vaccines or specific antivirals available for CHIKV, current treatment consists of nonsteroidal anti-inflammatory drugs to provide symptom relief (Parashar and Cherian, 2014).

### **1.2.1 Pathology of CHIKV**

CHIKV is spread to humans by infected *Aedes* mosquito bites, during blood meals the virus is spread to the susceptible endothelial, macrophage and fibroblast cells in the skin. The virus replicates locally in the skin until it is disseminated to the local lymphatic system and major organs. The target organs which replicate CHIKV are the liver, remote lymph nodes, muscles and joints (Dupuis-Maguiraga et al., 2012). CHIKV can spread to uninfected cells via apoptosis, the apoptotic blebs evade host responses to the virus and are taken in by neighbouring cells to propagate the infection (Lum and Ng, 2015).

The type I interferon (IFN) innate immunity was shown to control viral replication during early infection however  $CD4^+$  T cells specific for CHIKV were shown to possess a pathogenic role as they amplified joint inflammation but did not reduce virus dissemination and replication (Teo et al., 2013). The host immune response has been implicated in the development of chronic stage of CHIKV symptoms. CHIKV infected macrophages isolated from infected macaques three months after initial inoculation were shown to function as key cellular reservoirs of infection, this result coincided with the detection of RNA genome and infectious virus in homogenised muscle, liver and spleen tissue up to 3 months after the acute stage of infection was detected (Labadie et al., 2010). It has also been

shown that CHIKV developed persistent sites of replication in mice joint tissues, low levels of CHIKV replication at these sites was detected up to 16 weeks post infection. The mouse T and or B cell mediated adaptive immune response was shown to reduce the level of CHIKV replication over time but did not completely remove CHIKV in the joint tissues, the presence of low levels of CHIKV and therefore RNA in the joints after long periods of time post infection was hypothesised to cause joint inflammation and pain (Hawman et al., 2013). Neutralising IgM antibodies are typically found in patients suffering from chronic stage symptoms of CHIKV infection which is indicative of the persistence of CHIKV virus and antigens (Hawman et al., 2013).

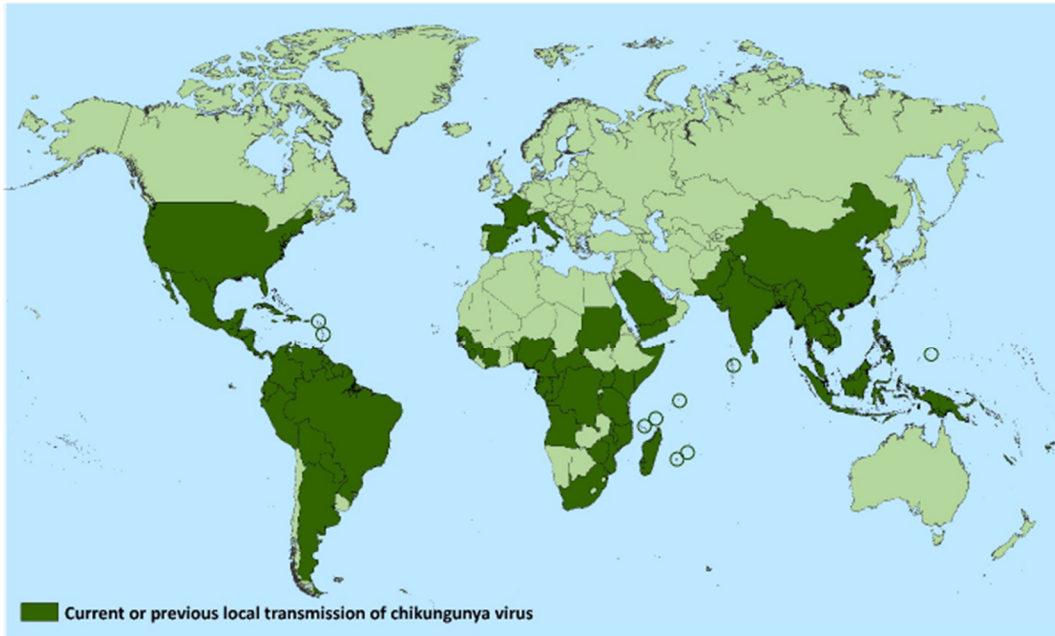
These results suggest that chronic stage of CHIKV infection could be caused by persistent low levels of CHIKV surviving in cellular reservoirs. The small amount of CHIKV could either cause the joint pain directly by causing cell death in the target joint or muscle tissues or the host immune response may cause indirect damage to these areas in response to the persistence.

### **1.2.3 Epidemiology of CHIKV**

Enzootic transmission cycles between non-human primates and mosquitoes were discovered soon after initially identifying CHIKV in East Africa in 1952. In 2004 an ECSA CHIKV strain had re-emerged in Kenya before spreading to the Indian ocean, this strain of CHIKV went on to cause outbreaks in India and Southeast Asia (Coffey et al., 2014). CHIKV strains are typically transmitted by the main vector *Ae. aegypti*, however adaptive mutations in the CHIKV genome facilitated a vector jump to the urban dwelling *Ae. albopictus*. The *Ae. albopictus* mosquitoes were first shown to transmit CHIKV in 2006 on the island of La Réunion in the Indian ocean, in this outbreak there were 244,000 cases of CHIKV. A single alanine to valine point mutation in the E1 glycoprotein at position 226 conferred increased fitness in *Ae. albopictus*. This mutation is hypothesised to facilitate a cholesterol dependent lipid membrane recognition which is important for replication in *Ae. albopictus* but not *Ae. aegypti* mosquitoes (Tsetsarkin et al., 2007).

The *Aedes* mosquitoes are distributed all over the world, this highlights the potential for CHIKV to spread to different countries. Recent outbreaks in the Western hemisphere have been recorded, CHIKV infections have been detected in Italy and France. An Asian strain of CHIKV caused outbreaks in South America

before spreading to the Caribbean and Central America (Weaver and Forrester, 2015). The potential for the virus to cause serious economic and social burden is exemplified by the number of cases of CHIKV reported in India in 2006 which reached up to 1.38 million (Naresh Kumar and Sai Gopal, 2010). A diagram depicting the worldwide spread of different CHIKV strains as of 29.05.18 published by the (CDC, 2018) can be seen below (**Figure 1**).



\*Does not include countries or territories where only imported cases have been documented.

**Figure 1: Reported cases of CHIKV by countries or territories as of 29.05.18.**

The dark green regions indicate countries or territories which have reported endemic cases of CHIKV. Areas which have reported cases of CHIKV which were imported from abroad are not highlighted above. The diagram was taken from CDC for the report on the distribution of CHIKV as of 29<sup>th</sup> of May 2018.

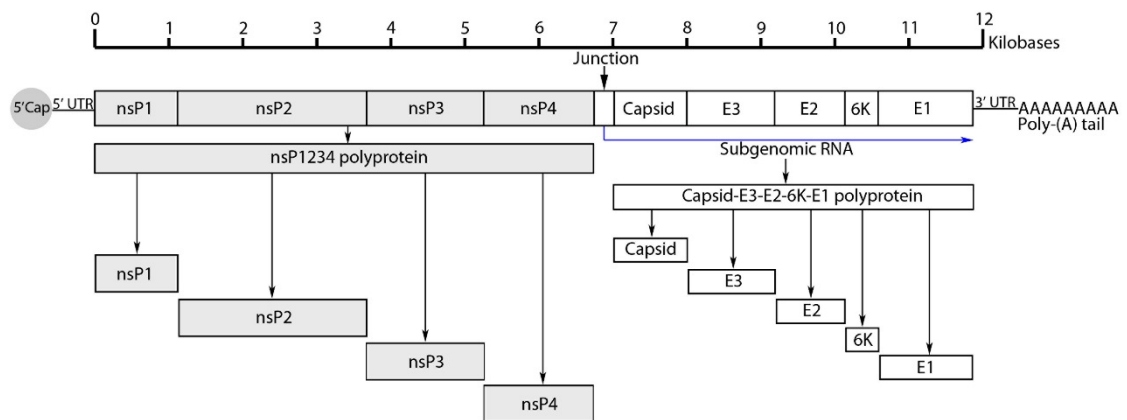
CHIKV poses a significant threat to currently unaffected countries as naïve populations are at high risk of epidemic. This threat is amplified by the increased number of infected travellers returning from countries which have endemic cases of CHIKV, these individuals risk transferring the virus to endogenous population of mosquitoes which could establish the spread of CHIKV. The *Ae. albopictus* mosquitoes can survive cold winters and in temperate climates such as Europe and North America which has increased the distribution of the virus. Interestingly the strain of CHIKV circulating in Southern America is an Asian strain which does not possess the A226V mutation which permits *Ae. albopictus* infection, this corresponds with the observed higher prevalence of *Ae. aegypti* in this region (Weaver, 2014).



### 1.3 Molecular biology of CHIKV

#### 1.3.1. Genome organisation

The CHIKV genome is organised as a capped 11.8 kilobase single stranded RNA with a 3' poly-(A) tail, the viral proteins are expressed by host translation machinery as two polyproteins. The first polyprotein contains nsP1 to nsP4. The structural proteins are expressed as a separate polyprotein which is transcribed from a SG promoter, the viral proteins are separated by conserved polyprotein cleavage sites. The structural proteins are the capsid, E1 to E3, and the 6K protein (Pardigon, 2009). A diagram displaying these features along the CHIKV genome can be seen below (**Figure 2**).



**Figure 2: The CHIKV genome is translated in two polyproteins, the first expresses the nsPs and the second expresses the structural proteins.**

CHIKV has an 11.8 kb single stranded RNA genome which is 5' capped and has a 3' poly-(A) tail. The genome contains two untranslated regions (UTR) in the 5' and 3' of the genome. The first portion of the genome encodes for the nsP1 – 4 polyprotein, polyprotein cleavage sites between the viral proteins release the mature forms of the nsPs. The structural proteins are expressed in a separate polyprotein under the control of a SG promoter found in the junction of the genome. These proteins are also separated by cleavage sites along the polyprotein.

The CHIKV genome contains three untranslated regions, these are the 5' and 3' untranslated regions (UTR) and the junction region between the two ORFs (**Figure 2**). Alphavirus genomes contain strongly conserved cis acting elements and RNA structures which promote virus replication, research on the prototype SINV has identified key features which are thought to be utilised across the

different alphaviruses. A conserved RNA stem loop at the beginning of the 5' UTR has been identified as important as disrupting this stem abolished replication. The structure was proposed to initiate translation of the nsPs and to promote replication of the positive sense genome from a negative sense intermediate (Niester and Strauss, 1990).

The SG promoter is used to transcribe the structural polyprotein mRNA, the sequence is found in the nsP4 and junction region of the genome. This promoter is formed in the negative sense intermediate of the genome, the minimum amount of sequence needed to transcribe the SG region is -19 nucleotides downstream and +2 into the SG transcript coding region (Levis et al., 1990). There are transcription enhancing elements in the sequence surrounding the promoter. These regions are found at positions -40 to -20 and +6 to +14 from the start of the SG coding region, they synergise to enhance the transcription of the transcript (Wielgosz et al., 2001). In the 3' UTR the poly-(A) tail facilitates translation of the viral proteins and enhances the stability of the RNA molecule, a 19 nucleotide conserved sequence element (CSE) adjacent to the poly-(A) tail in the 3' UTR is the core promoter for negative sense intermediate synthesis (Hardy and Rice, 2005). Cis acting elements are also found in the coding regions of alphavirus genomes, a CSE in the nsP1 gene was shown to form two hairpin RNA structures which promotes replication of SINV RNA (Niester and Strauss, 1990).

### **1.3.2 Non-structural proteins**

#### **1.3.2.1 nsP1**

The alphavirus replication complex is formed by all four nsPs, each protein possesses distinct functions during the virus lifecycle to promote replication. NsP1 is the capping enzyme which adds the cap to newly synthesised CHIKV genomes. The protein caps the 5' end of both the genome and SG RNA. NsP1 transfers an guanylyl residue to the 5' end of the RNA molecule using the N-terminal guanylyltransferase domain (Ahola and Kääriäinen, 1995), another conserved methyltransferase domain in this region methylates the cap during the reaction (Rozanov et al., 1992). NsP1 recruits host membranes to large cytopathic vacuolar structures type 1 (CPV-I) invaginations to form spherules, these spherules function as the site for virus RNA synthesis (Jose et al., 2017). The protein attaches to membranes with both a palmitoylation modification (Laakkonen et al., 1996) and a amphipathic helix (Spuul et al., 2007) (**Figure 3a**).

### 1.3.2.2 nsP2

Multiple roles have been identified for nsP2, the protein functions as the viral protease, nucleoside triphosphatase (NTPase) and helicase. Three dimensional structures determined by crystallography of the C-terminal of the VEEV nsP2 demonstrated that the alphavirus conserved cysteine protease adopts a novel fold to form a catalytic dyad protease active site, the active residues Cys477 and His546 selectively digest the three target sequences (Glu-Ala-Gly-Cys) found between the nsPs (Russo et al., 2006). The N-terminal of nsP2 contains domains identified to have RNA helicase and NTP domains, SFV nsP2 was shown to have NTP dependent helicase activity on ds-RNA probes (Gomez de Cedrón et al., 1999). NsP2 functions during the capping reaction of the genome and SG RNA; the N-terminal of the protein contains NTPase activity which specific removes the  $\gamma$ -phosphate from the 5' end of RNA molecule before the capping reaction (Vasiljeva et al., 2000) (**Figure 3b**).

### 1.3.2.3 nsP3

NsP3 comprises three domains: The N-terminal macro domain, central alphavirus unique domain (AUD) and C-terminal hypervariable domain (HVD). The protein is essential for viral replication but the exact functions of the protein are currently unknown. The viral protein has roles during the early stages of infection as nsP3 is necessary for synthesis of the negative sense RNA intermediate (Wang et al., 1994) with other studies revealing importance for the transcription of SG RNA (LaStarza et al., 1994). The macro domain is a highly conserved amongst the alphaviruses and in all kingdoms of life, enzymatic assays and crystal structures of the CHIKV macro domain revealed that the region bound to short segments of RNA (adenine), displayed ADP-ribose phosphatase activity and poly ADP ribose binding (Malet et al., 2009). The nsP3 macro domain functions as a mono-ADP ribose (MAR) hydrolase which may have roles in countering the innate immune response, members of the PARP family that are IFN $\alpha$  stimulated genes are known to respond to virus infections (Ecke et al., 2017). In another study CHIKV macro domain mutants with disrupted MAR hydrolase and binding were shown to reduce virus production in both mammalian and insect cell types which lack functional IFN pathways demonstrating that the domain has additional roles in CHIKV replication which are separate to the counteracting the immune response (McPherson et al., 2017).

CHIKV infection increases cellular levels of ADP ribosylation independently from normal induction by IFN $\alpha$ , this is because nsP3 binds to ADP-ribosylated proteins to facilitate initial CHIKV infection and MAR hydrolysis is utilised during the amplification of the virus replication complex (Abraham et al., 2018).

The central portion of nsP3 is the alphavirus unique domain that has a highly conserved sequence amongst the alphaviruses. The AUD functions to bind to RNA during replication using a zinc binding domain, where four cysteine residues coordinate the zinc ion, adjacent to a patch of basic residues (Shin et al., 2012). Mutagenesis analysis of alphavirus conserved residues in AUD revealed that these were important for replication in different cell types, the synthesis of SG RNA synthesis and RNA binding activity (Gao et al., 2019).

The HVD is an unstructured flexible region of nsP3 which contains motifs which mediate protein- protein interactions, the length and sequence of the HVD varies between different alphaviruses. This intrinsically disordered region contains a proline rich region and is hyper-phosphorylated (Götte et al., 2018). The HVD contains two FGDF motifs which bind to the host G3BP proteins which function during stress granule formation. Binding to the G3BP proteins disrupts stress granule formation during infection, stabilises the replication complexes and promotes viral replication (Schulte et al., 2016). Differences between the HVDs of different alphaviruses may function to tailor the protein to efficiently aid replication of the different viruses; the New World VEEV HVD was shown to not interact with G3BP and caused the nsP3 to localise in unique spheroid structures which differ from the typical CPV-I complexes seen during Old world alphaviruses infection (Foy et al., 2013). The different alphaviruses utilise different host factors to redundantly perform the same function. The VEEV-nsP3 HVD interacts with FXR proteins with VEEV specific motifs in the C-terminal of the region, FXR proteins bind RNA and function during stress granule formation. The VEEV nsP3 to FXR interaction is analogous to the G3BP to Old World alphavirus nsP3 interaction (Kim et al., 2016).

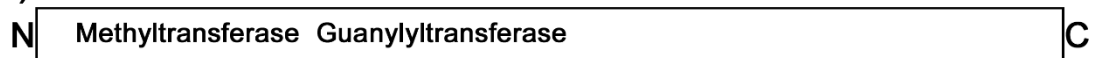
During SINV infection nsP3 is seen in both the membrane associated replication complexes and in cytoplasmic electron dense rod-like structures (Cristea et al., 2006). Similar nsP3 structures were seen during CHIKV infection as small granules which form long rod like and granular structures form over the course of infection, the different structures all sequester G3BP but there may be unknown

functions which are yet to be identified (Remenyi et al., 2017). Future research identifying novel protein and RNA interactions will be invaluable as these studies could identify currently unknown functions of the viral protein which could reveal key insights about the alphavirus lifecycle (**Figure 3c**).

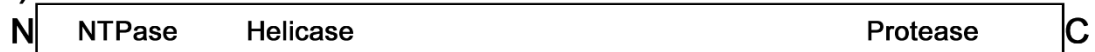
#### 1.3.2.4 nsP4

NsP4 is the RNA dependent RNA polymerase (RdRp), the core catalytic domain in the C-terminal of the protein contains a Glu-Asp-Asp motif which is characteristic of viral polymerases. The activity of polymerase is abolished when the motif is substituted for Glu-Ala-Ala (Rubach et al., 2009). The polymerase active site is present in the “palm” domain as seen in other polymerases. The catalytic domain of the protein also functions as a terminal adenylyltransferase, nsP4 adds the adenylate residues to the end of the genome and SG RNA. The tail is important for replication, translation and RNA stability (Tomar et al., 2006) (**Figure 3d**).

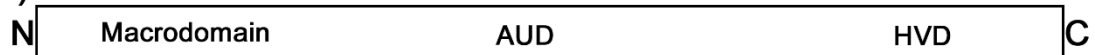
##### (a) nsP1



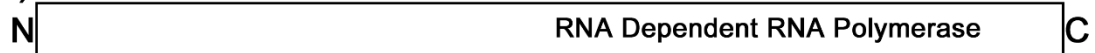
##### (b) nsP2



##### (c) nsP3



##### (d) nsP4



**Figure 3: A diagram visually depicting the known functions of the CHIKV nsP1 to nsP4.**

- (a) NsP1 contains a N-terminal methyltransferase and guanylyltransferase domain.
- (b) NsP2 has N-terminal NTPase and helicase domains and a C-terminal protease.
- (c) NsP3 has a N-terminal macrodomain, a central AUD and C-terminal HVD. (d) NsP4 has a C-terminal RNA dependent RNA polymerase domain.

### **1.3.3 Structural proteins**

#### **1.3.3.1 Capsid**

Alphavirus virions are icosahedral nucleocapsids surrounded by an envelope derived from host membranes, each virion contains one RNA genome. This nucleocapsid is formed up from 240 copies of the capsid protein and has T=4 symmetry. The capsid protein is separated into three regions, the first region of the protein is important for recognising corresponding capsid monomers during dimerization and non-specifically recognises genome RNA (Hong et al., 2006). The second region of the capsid protein contains an RNA binding region which binds to genome RNA during encapsidation (Geigenmüller-Gnirke et al., 1993). The third region of the capsid protein forms the serine protease, the active site is a catalytic triad formed by the highly conserved His, Asp and Ser residues which autocatalytically cleave the capsid away from the polyprotein precursor (Choi et al., 1991). Three dimensional structures determined by crystallography of the CHIKV capsid proteinase domain revealed that the catalytic triad are formed by the His139, Asp161 and Ser213 residues specifically and a hydrophobic pocket on the capsid protein interacts with the E2 envelope protein (Sharma et al., 2018).

#### **1.3.3.2 E1**

E1 is glycoprotein which forms heterodimers with the E2 protein on the surface of the alphavirus virions, the former is a class II fusion protein which fuses the virion envelope with host membranes in a pH dependent manner. Acidic pH in the host vesicles causes the heterodimer of E1 and E2 to undergo a conformational change which results in the formation of E1 homotrimers which form pores in the both host and virus membranes (Sanz et al., 2003). The proposed fusion pore for alphaviruses is made from five E1 homotrimers which organise into a ring structure that induces deformation in the host and viral lipid bilayer (Gibbons et al., 2004).

#### **1.3.3.3 E2**

CHIKV enters host cells by receptor mediated endocytosis mediated by binding of E2 to host glycoproteins. Alphaviruses enter host cells via clathrin-coated vesicles which form after receptor binding, the particle is transported by the endocytic pathway which is where the pH dependent rearrangement of E1-E2 heterodimers occurs (Helenius, 1995). The protein is separated into three domains labelled A, B and C. Domain A forms the tip of the E1-E2 heterodimer

spikes on the surface of the virus, domain B is found protruding sideways of the structure and domain C is near the viral membrane. Mutations in the domain A and B can alter cell type tropism and can confer evasion of host antibodies (Voss et al., 2010). Disrupting any of the three domains attenuates viral fitness, disrupted domain C displayed the most significant decrease (Weger-Lucarelli et al., 2016).

#### **1.3.3.4 E3**

The E3 structural protein is a trafficking signal which translocates the E3-E2-6k-E1 polyprotein to the endoplasmic reticulum during alphavirus replication. A precursor of E1, E2 and E3 called pE2-E1 is stabilised by E3 which prevents disassembly, E3 also assists in E2 protein folding as a chaperone, the protein stabilises the interaction between E1 and E2 and prevents premature fusion peptide formation in the ER and Golgi body (Snyder and Mukhopadhyay, 2012).

#### **1.3.3.5 6K**

6K is a small protein (6 kDa) which functions as a viroporin during alphavirus replication, viroporins function to increase membrane permeability. The channel is cation selective but does not have a role during virus entry into the cell, instead 6K is associated with the membranes of the endoplasmic reticulum, Golgi network and plasma membrane suggesting a role in the late stages of virus replication or release (Melton et al., 2002). A ribosomal frame shift (-1 position) in 6K gene results in the expression of an 8 kDa TransFrame (TF) protein which possesses an elongated C-terminal, the TF protein is incorporated into the virion during budding with possible roles in stabilising the structure (Firth et al., 2008).

#### **1.3.4 Viral lifecycle**

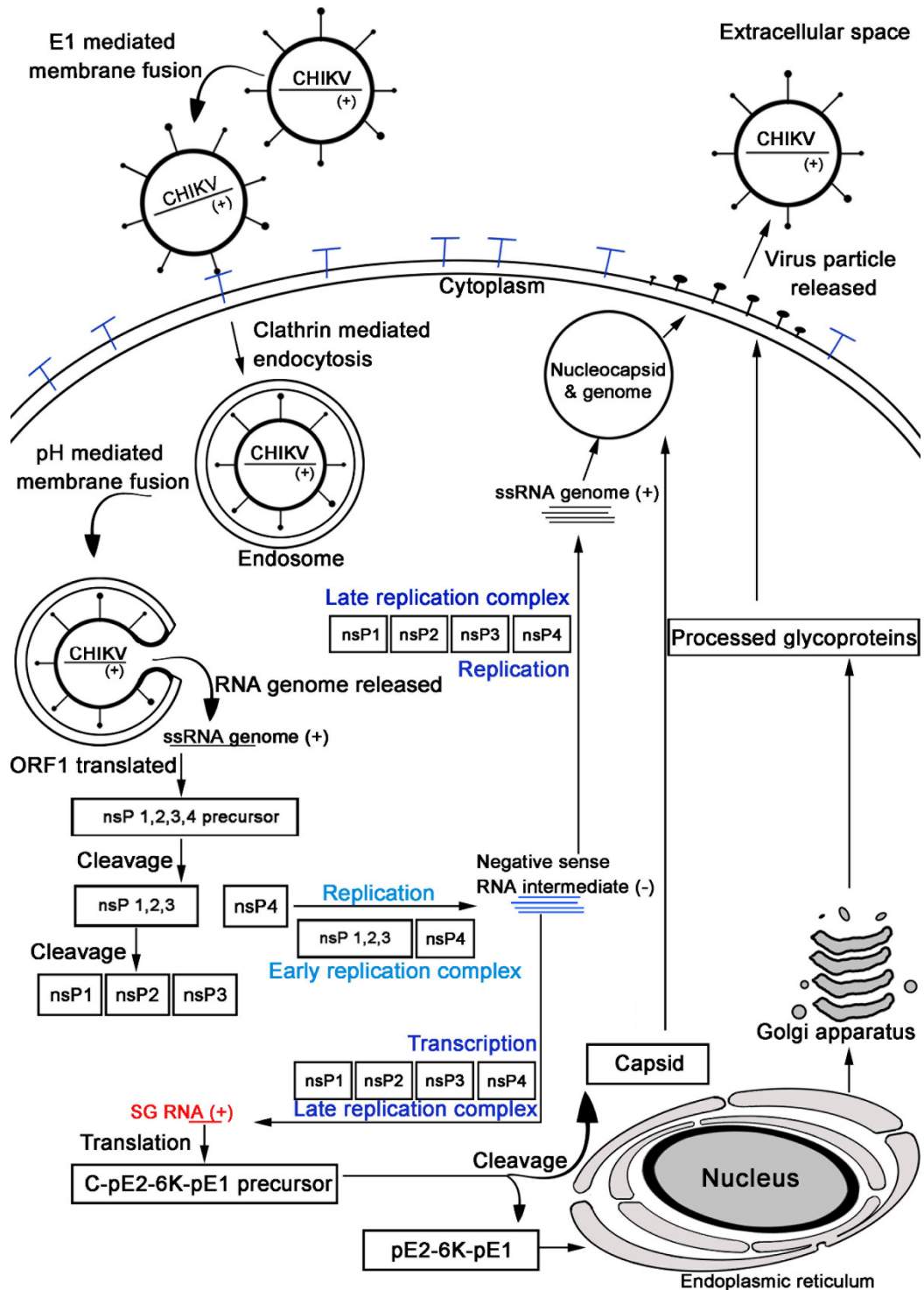
Alphaviruses enter the host cell via receptor mediated endocytosis, this occurs when the E2 envelope protein interacts with host receptor proteins on the surface of the target cell which induces clathrin coated pits to internalise the virus molecule. The CHIKV E2 protein binds to the MXR8 receptor to mediate this process (Song et al., 2019; Basore et al., 2019). The lower pH in the endosome causes the E1-E2 heterodimer reorganisation which releases the E1 envelope

protein, the E1 homotrimers function as a type II fusion peptide to fuse the host and virus membranes to release the virus into the host cytoplasm. The nucleocapsid disassembles in the cytoplasm and the RNA genome is released, the first ORF is translated to form the nsP1234 polyprotein (Leung et al., 2011).

Early in infection the nsP2 segment of the polyprotein performs an *in cis* cleavage at the nsP3/nsP4 junction to release the nsP4. The nsP123 precursor and nsP4 form the early replication complex, this recognises the RNA genome to replicate a negative strand RNA intermediate. The nsP123 precursor can only be cleaved *in trans* by nsP2 when enough of the precursor has been synthesised, this causes the release of the mature forms of nsP1, nsP2 and nsP3. The mature nsP1-4 form the late replication complex which recognises the negative sense RNA intermediate to replicate new genomic RNA and transcribe the SG RNA. The SG RNA is translated into structural polyprotein C-pE2-6K-E1 and the capsid protein is autoproteolytically cleaved in the cytoplasm. The remaining pE2-6K-E1 precursor is translocated to the ER, pE2 is formed by the E3 and E2 (Jose et al., 2009).

Signal peptidases cleave the pE2-6K-E1 precursor into pE2, 6K and E1. The structural proteins are then trafficked to the Golgi network where further processing with furin proteases cleaves the pE2 into the mature forms of E3 and E2. The processed glycoproteins are trafficked to the cell surface and the capsid proteins encapsidates the viral genome. The new virion is formed by this capsid structure and the host membrane decorated with the surface spikes (Brown et al., 2018). A diagram depicting the virus lifecycle is shown below (**Figure 4**).





**Figure 4: A diagram of the CHIKV lifecycle.**

CHIKV enters the host cell by E2 mediated membrane fusion upon binding to a host receptor, this triggers clathrin mediated endocytosis to internalise the virus. The E1 fuses the virus and endosome membrane at low pH, this releases the +ve genome into the cytoplasm and the nsP1234 polyprotein precursor is translated. The nsP4 is cleaved by nsP2 in cis, the nsP123 and nsP4 form the early replication complex to replicate the -ve RNA intermediate. When enough nsP123 is produced the mature nsP1, nsP2 and nsP3 are released in trans by nsP2. The four mature nsPs form the late replication complex which uses the -ve RNA intermediate to replicate the +ve genome and transcribe the SG RNA. The structural polyprotein is translated from the SG RNA, the capsid is autocatalytically cleaved, and the pE2-6K-pE1 precursor is processed in the endoplasmic reticulum and golgi apparatus. The processed glycoproteins, capsid and +ve genome assemble as new CHIKV particles at the host membrane before release.

#### **1.4 Differences in insect and mammalian host responses to CHIKV infection**

Mosquitoes infected with CHIKV suffer from lifelong symptomless infections while infection in mammalian hosts is acute and causes debilitating symptoms, the exact cause of this discrepancy is currently unknown. Differences in cell morphologies observed between the vertebrate and invertebrate host could explain the different pathologies seen during alphavirus infection. Alphavirus infections in mammalian cell tissue culture displayed inhibited host macromolecule synthesis and cell death as early as 24 hours post infection. This contrasts the infection in mosquito cells where no cytotoxicity or disrupted translation is seen (Sokoloski et al., 2012).

The expression of both the replicase and structural proteins in mammalian cells cause apoptosis, the accumulation of alphavirus envelope proteins in the ER activates the host unfolded protein response which causes the cell to undergo caspase mediated apoptosis. This accelerates the alphavirus induced apoptosis (Barry et al., 2010). Insect cells infected with SFV were shown to upregulate a protease inhibitor after 5 days post infection which reduced the production of mature nsPs, this mechanism could contribute to alphavirus persistence in these cells as a fall in produced virus was also seen at this time (Mudiganti et al., 2010).

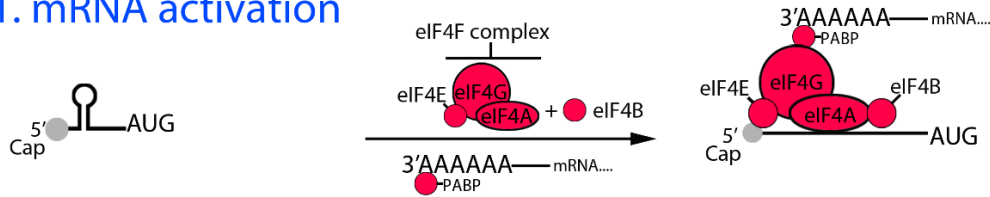
The RNA interference pathway (RNAi) is a key antiviral defense mechanism utilised by the *Aedes* mosquitoes to respond to and limit virus replication. The virus derived small interfering RNAs (siRNAs) are produced from dsRNA formed by the alphavirus genome and negative sense RNA intermediate coming together during replication and not the RNA structures in the genome (Siu et al., 2011). Suppressing the RNAi response of *Ae. aegypti* mosquitoes during SINV infection causes a higher level of virus to be produced which is fatal for the mosquito (Myles et al., 2009). This suggests that maintaining lower levels of alphavirus replication over a long period of time could be key in maintaining the lifelong infection seen in the invertebrate host.

## 1.5 Translation

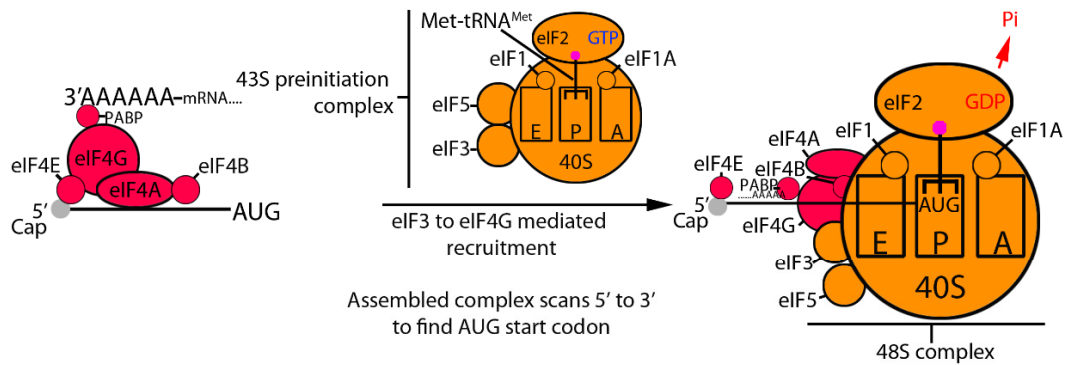
### 1.5.1 Mechanism of eukaryotic translation

Translation of mRNA into protein follows a conserved canonical pathway which is separated into several distinct processes performed by the ribosome. Translation initiation is facilitated by multiple eukaryotic initiation factors (eIFs). The ribosome is separated into the E site, P site and A site for translation. Before initiation the mRNA must be activated before it is used for translation, capped RNA with structured 5' UTRs are unwound in preparation for translation by the eIF4F complex. The eIF4F complex is formed by the cap binding eIF4E, the eIF4G scaffold, the DEAD-box RNA helicase eIF4A and eIF4B which enhances the helicase activity. The Poly-(A) tail bound Poly-(A) binding protein (PABP) interacts with the eIF4G scaffold protein, this forms a closed-loop structure which circularise the mRNA to increase the efficiency of subsequent rounds of translation (Ivanov et al., 2016). A 43S preinitiation complex binds to the activated mRNA. This complex is formed by the 40S ribosome, eIF1/1A, eIF2-GTP-Met-tRNA<sup>Met</sup> (eIF2 TC) and eIF3. An interaction between the eIF4G and eIF3 recruits the 40S ribosome to the activated mRNA (Villa et al., 2013). The eIF1/1A results in an open latch conformation in the 40S ribosome which facilitates binding and the 43S complex scans the RNA from the 5' to the 3' for the appropriate AUG start codon, upon recognition the 48S initiation complex forms where the Met-tRNA<sup>Met</sup> is positioned in place with the start codon in the P site and eIF1 is displaced to switch to a locked closed conformation. Forming the 48S complex causes the eIF5 to specifically activate the eIF2 GTPase activity which causes the protein to partially release the Met-tRNA<sup>Met</sup>. The 60S ribosome and eIF5B are recruited to the 48S complex, this causes the release of the other eIFs, this forms 80S initiation complex which causes eIF5B GTP hydrolysis and release of eIF5B and eIF1A. The assembled 80S complex undergoes elongation to translate the protein (Jackson et al., 2010). This mechanism is depicted in **(Figure 5)**.

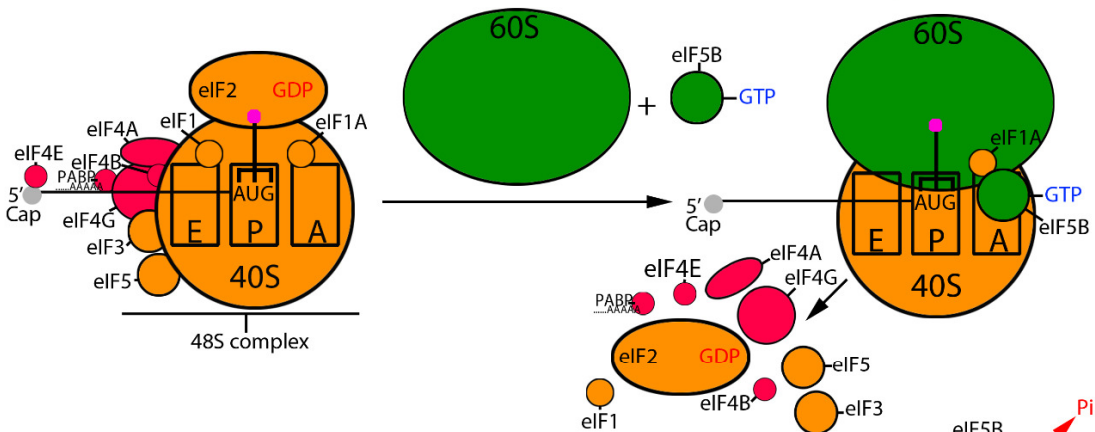
## 1. mRNA activation



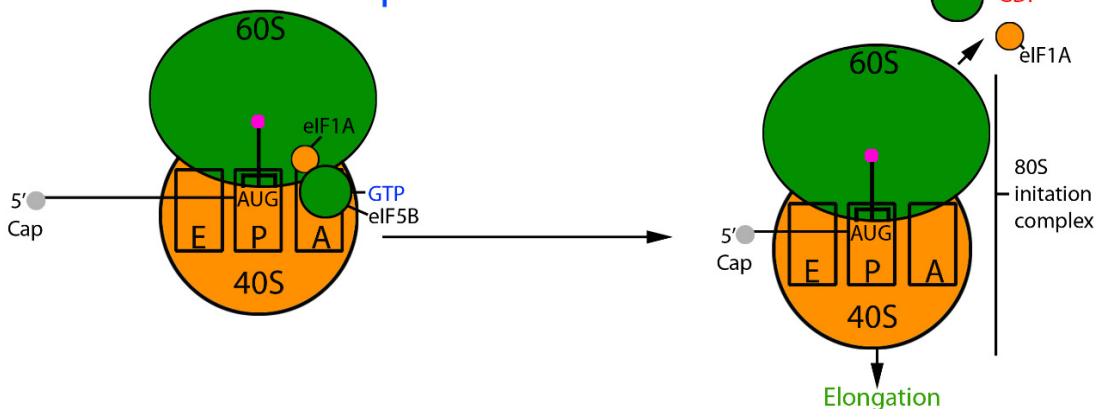
## 2. 43S complex AUG recognition/ 48S complex formation



## 3. 60S subunit and eIF5B recruitment and eIF displacement



## 4. 80S initiation complex formation

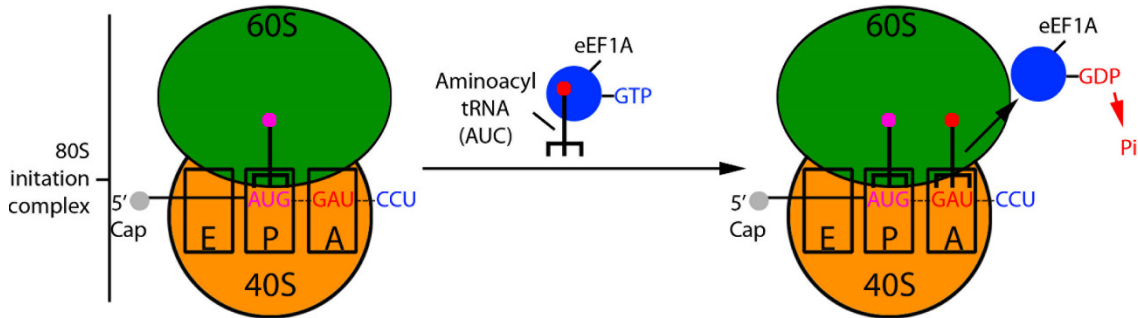


**Figure 5: Eukaryotic translation initiation.**

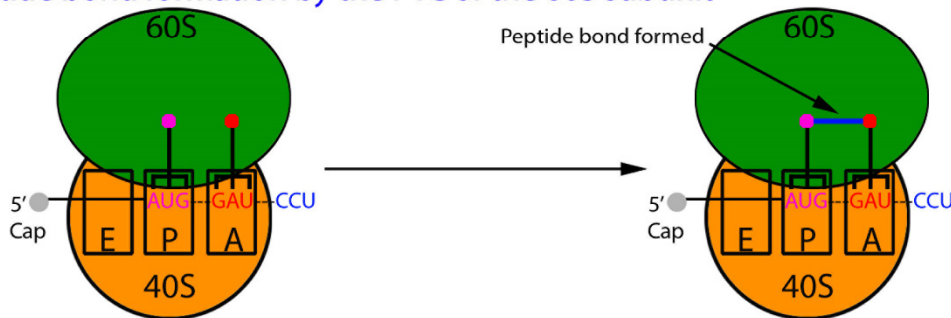
(1) Capped mRNA recruits the eIF4F complex (eIF4E+eIF4G+eIF4A) and eIF4B which unwind the RNA structures to activate the RNA for translation. The PABP interaction with eIF4G causes the mRNA to form a closed-loop structure (2) The 43S preinitiation complex is recruited to the activated mRNA by the eIF3 to eIF4G interaction, it then scans for the AUG start codon in the 5' to 3' direction, upon recognition the 48S complex is formed, the Met-tRNA<sup>Met</sup> is in the P site with the start codon, eIF1 is displaced and eIF2-GTP is hydrolysed. (3) 60S ribosome subunit and eIF5B is recruited to the 48S complex, this displaces most of the other eIFs. (4) The 80S initiation complex is formed, hydrolysis of the eIF5B releases the eIF5B and eIF1A, the 80S complex progresses to elongation.

Elongation of the synthesised peptide chain is catalysed by the eukaryotic elongation factors (eEF). The Met-tRNA anticodon is base-paired with the AUG codon in the P site of the ribosome with the subsequent codon of the sequence in the A site. The matching aminoacyl-tRNA to the codon in the A site is recruited by the eEF1A, codon-anticodon recognition results in GTP hydrolysis which releases the eEF1A and leaves the tRNA in the A site. The eEF1A is recycled by the guanine nucleotide exchange factor eEF1B which cycles the GDP for GTP. The peptidyl transferase centre (PTC) in the 60S ribosome catalyses the peptide bond formation between the tRNA in the P and A site to insert the next amino acid into the peptide chain. GTP hydrolysis and phosphate release catalysed by eEF2 translocates the tRNA molecules and mRNA to the next ribosome site such that the components in the P site move to the E site and those in the A site goes to the P site of the ribosome. This process causes the deacylated tRNA from the P site to shift into the E site for exit and the new peptidyl tRNA to move into the P site, the empty A site is free to accept the next aminoacyl-tRNA matching the next codon in the sequence (Dever and Green, 2012). This process of elongation repeats until translation is signalled to stop (**Figure 6**).

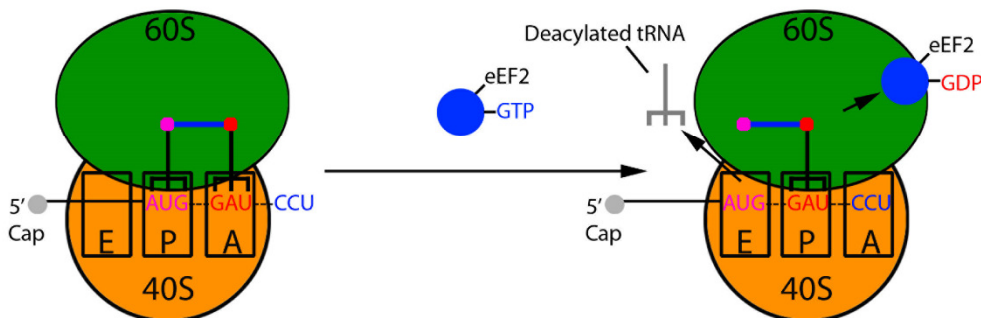
### 1. Aminoacyl-tRNA matching to codon in the A site recruitment by eEF1A



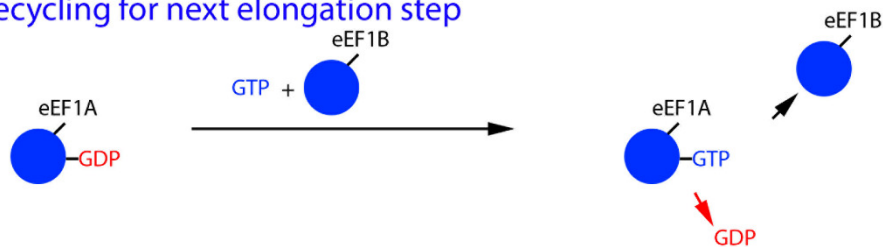
### 2. Peptide bond formation by the PTC of the 60S subunit



### 3. eEF2 mediated tRNA translocation



### 4. eEF1A-GDP recycling for next elongation step



### Figure 6: Eukaryotic translation elongation.

(1) Aminoacyl-tRNA which complements the mRNA codon in the A site of the 80S complex is recruited by eEF1A-GTP. Successful codon-anticodon pairing causes eEF1A GTP hydrolysis which releases the protein and leaves the tRNA in the A site. (2) The 60S subunit PTC forms the peptide bond between the amino acids in the P and A site. (3) eEF2-GTP causes the P and A site tRNA and codons to translocate to the E and P sites respectively, the deacylated tRNA is released from the E site and the eEF2-GDP is released. The next codon in the mRNA sequence is in the A site to receive the next aminoacyl-tRNA. (4) eEF1A-GDP is recycled for the next elongation step by eEF1B which cycles the GDP for GTP.

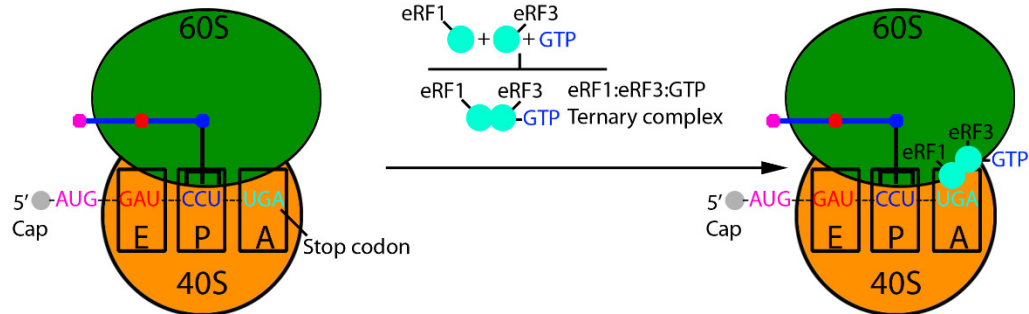
Translation termination is initiated when one of the three stop codons enters the A site, these are the opal (UGA), ochre (UAA) or amber (UAG) stop codons. Termination is mediated by two eukaryotic release factors (eRF) eRF1 and eRF3, the eRF1 can recognise all three stop codons and hydrolyses the peptide chain and eRF3 functions as a GTPase which catalyses the cleavage. The eRF1 is separated into 3 domains, the N-terminal domain (NTD) which recognises all three stop codons, the middle (M) domain which contains the Gly-Gly-Gln (GGC) loop which catalyses peptide bond hydrolysis when in the PTC and the C domain which binds GTP (Des Georges et al., 2014). The C-terminals of eRF1 and eRF3 interact with GTP to form into a ternary complex (eRF1:eRF3:GTP), this complex interacts with the ribosome forming the pre-termination complex. The eRF1 decodes the stop codon in the A site which causes a conformational change that translocates the codon into the P site. Stop codons are identified by eRF1 by a two-step sampling process. In the first step the NTD of eRF1 initially binds the first two nucleotides of the stop codon in the ribosome A site, this recognition causes the second step where eRF1 conformational changes alter the residue contacts with the second nucleotide and facilitates the ribosome recognition of the third nucleotide. Alternatively, during the second step the ribosome recognises the second nucleotide and the rearranged eRF1 binds to the third nucleotide (Kryuchkova et al., 2013). The eRF3 catalyses the hydrolysis of GTP upon stop codon recognition which rearranges the eRF1 in the P site into a favourable conformation by either releasing the eRF3:GDP from the ribosome or altering the protein interaction with eRF1. In this conformation the GGC loop of the eRF1 inserts into the PTC, this catalyses the hydrolysis of the peptidyl-tRNA to release the peptide chain from the ribosome (Alkalaeva et al., 2006).

After peptide release post termination complexes (post-TCs) need to be recycled for the next round of translation by releasing the 60S/40S ribosome subunits, the bound mRNA and deacylated tRNA. The recycling of these components is initiated by the eIF3 binding to the 40S subunit to release the 60S subunit, this separation is enhanced by the eIF1 and eIF1A which subsequently bind to the 40S subunit. The eIF1 dissociates the deacylated tRNA from the 40S subunit and the mRNA is released by the eIF3j subunit of eIF3. Another protein named the ATP-binding cassette sub family E member 1 (ABCE1) uses ATP binding and hydrolysis to recycle the 80S ribosome into the 40S and 60S ribosome subunits

(Schuller and Green, 2018). The eIF3, eIF1 and eIF1A remain attached to the 40S subunit, this prevents the rebinding of the 60S subunit and prepares the 40S subunit for the formation of the next 43S preinitiation complex (Pisarev et al., 2007) (**Figure 7**).



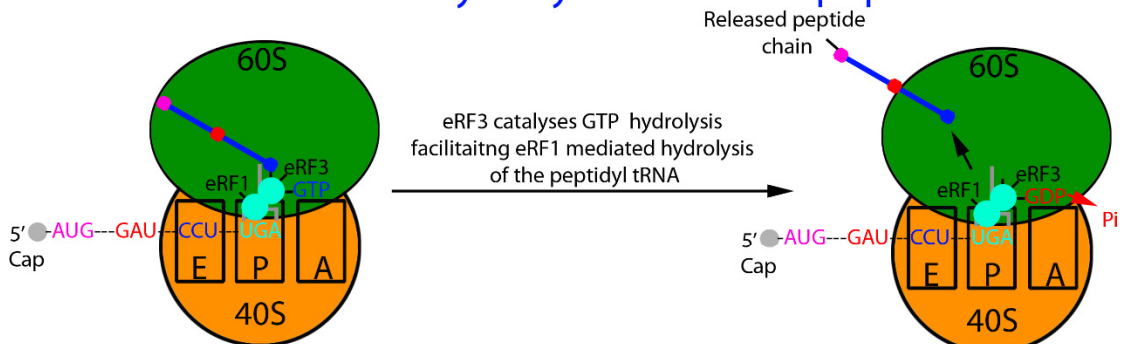
## 1. eRF1:eRF3:GTP complex recruitment to A site stop codon



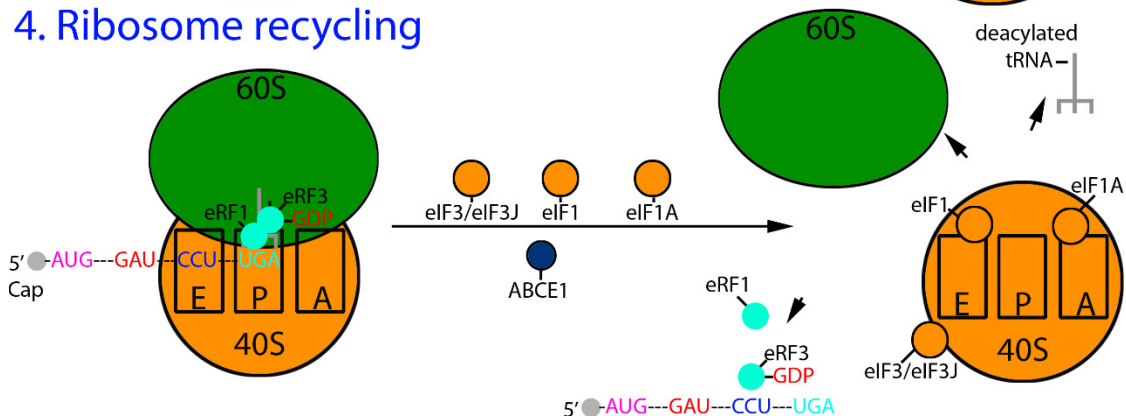
## 2. Ternary complex/ stop codon translocation into the P site



## 3. eRF3 mediated GTP hydrolysis facilitates peptide release



## 4. Ribosome recycling



**Figure 7: Eukaryotic translation termination.**

(1) The stop codon in the A site is recognised by the eRF1:eRF3:GTP ternary complex to begin the termination process. (2) Successful two-step sampling of the stop codon by eRF1 causes the stop codon and the ternary complex to translocate into the P site. (3) eRF3 catalysed hydrolysis of GTP rearranges eRF1 into a favourable position in the P site, this facilitates the eRF1 insertion into the PTC to hydrolyse the peptidyl-tRNA therefore releasing the protein from the ribosome. (4) The translation machinery is recycled for the next round of translation by the eIF3/eIF3j, eIF1, ABCE1 and eIF1A which release the 40S and 60S subunit, eRFs, deacylated tRNA and mRNA. The eIFs remain bound for the next round of initiation

### 1.5.2 Stop codon read-through

Stop codon read-through occurs when the ribosome does not terminate translation at a stop codon but instead continues elongation past the stop codon until the next stop codon in the coding sequence is reached. The aminoacyl tRNA which is inserted at the stop codon position is a near-cognate tRNA (nc-tRNA) which can bind to 2/3 of the codon in the ribosome A site. Basal stop codon read-through in mammalian cells is inefficient and only occurs up to a frequency of 0.1%. The context surrounding the stop codon is important for determining the read-through efficiency. During the standard termination sequence eRF3 interacts with poly-A binding proteins to activate the GTPase activity necessary for termination, this interaction can only occur when the stop codon is within proximity with the poly-A tail which is usually the case for stop codons at the end of mRNA sequences (Dabrowski et al., 2015).

Viruses commonly employ stop codon read-through mechanisms to maximise the coding potential of relatively small genetic material, these stop codons are found upstream of the regions coding for essential viral proteins such as the viral polymerase to control expression levels. Three different read-through signals which promote stop codon read-through have been identified. The type 1 read-through signals which are found in plant viruses involve two codons following the stop codon. An amber stop codon upstream of the Tobacco mosaic virus (TMV) RdRp encodes for the insertion of a tyrosine 5% of the time. A hexa-nucleotide CAR-YYA motif (R= purine and Y= pyrimidine) present in the two subsequent codons are essential for efficient read-through, the motif also promotes read-through of a substituted opal and ochre stop codon in this position albeit at a lower frequency (Skuzeski et al., 1991).

Type 2 read-through signals are found in the nucleotide context downstream of the stop codon, read-through of some viruses was shown to be promoted by the CUA or CGG codon which immediately followed the stop codon. This read-through signal is found in both plant and animal viruses, the most common stop codon was the opal stop codon but some cases possessed the ochre. The Tobacco rattle virus (TRV) read-through of the opal stop codon is necessary for replicase expression, replacing the last two nucleotides of the CGG codon significantly attenuates read-through of the stop codon (Beier and Grimm, 2001).

Many alphaviruses possess an opal stop codon in the nsP3 coding sequence upstream of the nsP4 coding region. Initial studies identified a type 2 read-through signal for the stop codon as a single cytidine nucleotide. For SINV the cytidine alone in the downstream CUA codon can facilitate read-through of the opal stop codon up to a frequency of 10%, removing this cytidine causes the read-through to drop to basal levels of read-through (Li and Rice, 1993). An additional read-through promoting downstream RNA structure has been identified for many type 2 read-through cases. In SINV and VEEV a large conserved downstream RNA structure up to 200 nucleotides in length was shown to promote read-through by 10 fold. The conserved bottom stem of the RNA structure is separated from the opal stop codon by an 8 – 9 nucleotide spacer segment. Compensatory mutations in this bottom stem are seen between some of the alphavirus strains therefore implicating the importance of this structure for virus replication. The exact mechanism of the RNA structure is currently unknown (Firth et al., 2011).

Type 3 read-through signals are typically seen for amber stop codon read-through, these signals are formed by the nucleotides 3' of the stop codon and a downstream RNA pseudoknot which forms long distant interactions along the genome. Read-through of the Moloney murine leukaemia virus (MuLV) amber stop codon in the gag structural protein coding sequence results in the expression of the polymerase (pol) gene, this read-through occurs at a frequency of 5 to 10%. This read-through is mediated by an RNA pseudoknot which is separated from the stop codon by a 8 nucleotide spacer (Alam et al., 1999).

### **1.5.3 Opal stop codon read-through in the CHIKV nsP3**

An opal stop codon (UGA) has been identified toward the end of the CHIKV nsP3 coding sequence at the same position described for the other alphaviruses, in some strains however this opal stop codon is substituted for an arginine codon instead (AGA or CGA). An arginine at this position would result in the expression of the full length polyprotein upon each translation, the exact phenotype behind the presence or absence of the opal stop codon on CHIKV replication is not fully characterised (K.C. Chen et al., 2013). The infrequent opal stop codon read-through would result in build-up of nsP123 precursor without expressing nsP4, the reduced nsP4 production could modulate virus replication (Hwang Kim et al., 2004).

Different isoforms of nsP3 are expressed when either the read-through or termination event at the opal stop codon occurs, the exact function of the two forms of nsP3 and the role during the virus lifecycle are currently unknown. The majority of SFV code for an arginine at the stop codon position and therefore expresses only the longer nsP3 isoform, the last 6 C-terminal residues of SFV nsP3 were shown to function as a degradation signal which reduced the half-life of the expressed nsP3 by 8 fold. The corresponding degradation signal was also found in the C-terminal of the longer isoform of SINV nsP3 but not the shorter isoform (termination product), the last six residues of the stop codon read-through product were shown to promote protein degradation (Varjak et al., 2010). It is possible that this degradation signal is conserved amongst alphaviruses, the opal stop codon read-through mechanism may function to express two populations of nsP3 which function in different aspects of the virus lifecycle.

The role of the opal stop codon during the alphavirus lifecycle is not understood, a Caribbean strain of CHIKV was shown to be a quasispecies where both the arginine and opal stop codon were detected in different isolates. Substituting the opal stop codon of a Sri Lankan strain of CHIKV did not affect virus replication *in vitro* or *in vivo* but caused a decreased pathogenesis in an infected mouse model, the sense codon mutant caused a delayed proinflammatory chemokine and cytokine recruitment phenotype in mouse models which consequently demonstrated reduced ankle swelling compared to the opal stop codon mutant (Jones et al., 2017). The authors highlight how the reduced phenotype seen for the arginine mutation contrasts the results seen previously for other alphaviruses; reintroduction of the opal stop codon into a neurovirulent strain of SINV reduced virus virulence by 25% in a mouse model (Suthar et al., 2005) and the presence of an arginine codon at the position for SFV strains is a requirement for neurovirulence with the mutation alone facilitating morbidity in mouse models (Tuittila and Hinkkanen, 2003). These results demonstrate that influence of stop codon read-through on virus replication may vary between different alphaviruses as the opal stop codon enhances the pathogenesis of CHIKV rather than attenuates it.

Different selection pressures cause the codon to switch between the opal stop codon or the arginine codon suggesting that both have a function during the alphavirus lifecycle. Passaging an opal stop codon containing strain of ONNV in mammalian Vero cells results in the sequence mutating to an arginine codon (Lanciotti et al., 1998), this result suggests that the arginine may enhance replication in mammalian cell types over the opal stop codon. Corresponding evidence of the preference for the opal in the invertebrate host has been reported as the opal stop codon doubles ONNV infectivity in the *Anopheles gambiae* vector and mosquitoes infected with both arginine and opal stop codon containing ONNV select for the opal stop codon by 10 days post infection. The opal stop codon ONNV produced lower titres in both BHK mammalian and the C6/36 *Ae. albopictus* mosquito cell lines tested in the study, the authors suggested that the stop codon could limit the production of virus in certain cells to facilitate virus persistence as seen in the vector (Myles et al., 2006). Another study demonstrated that infecting host cells with depleted levels of polyamines (carbon chains with positively charged amino groups) with an opal stop codon containing strain of CHIKV resulted in the virus substituting the stop codon for an arginine at this position. Polyamines are used by CHIKV during RdRp activity so removing the opal stop codon increased the overall expression of nsP4 to overcome deficiency, interestingly the study also found that the arginine coding CHIKV produced a decreased titre in the C6/36 cell line compared to the opal stop codon virus (Mounce et al., 2017). Interestingly the opal stop codon enhancing CHIKV replication in C6/36 cells contrasts the result seen for ONNV, further research into the role of the opal stop codon in mosquitoes is necessary.

A recent study used selective 2' hydroxyl acylation and primer extension (SHAPE) to experimentally determine the opal stop codon recoding large RNA structure in short sequences of the African/Asian or Caribbean strains of CHIKV inserted into bis-cistronic reporter plasmids. The SHAPE data confirmed the presence of a large RNA structure downstream of the CHIKV opal stop codon. The structure from both strains of CHIKV recoded opal stop codon read-through at equal efficiency when measured as a function of downstream luciferase expression (Kendra et al., 2018). Additional SHAPE experiments on the opal stop codon recoding RNA structure in the context of the full RNA genome would be valuable as the results would reveal if the structure is potentially influenced by

long range RNA interactions. It is currently unknown whether this RNA structure forms in mosquitoes, investigating how this structure folds at lower temperatures could provide an insight into the mechanism behind the different opal stop codon CHIKV replication in vertebrate and invertebrate host cells.

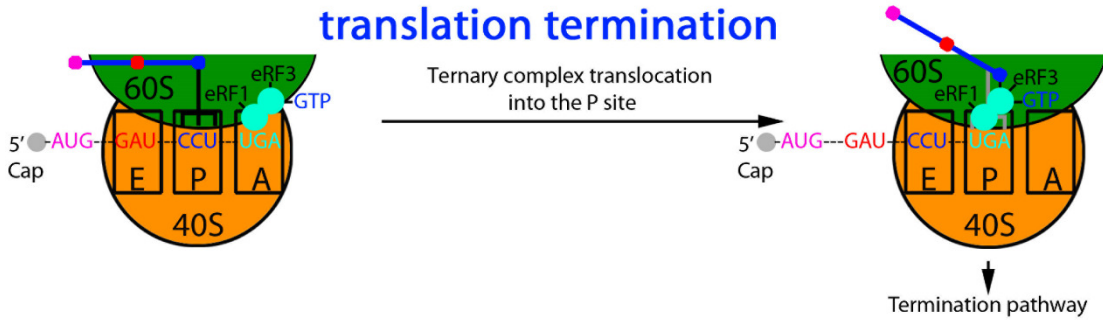
#### **1.5.4 Potential mechanism of stop codon read-through**

The exact mechanism for stop codon read-through is not understood but some insights into this process have been found. The three stop codons possess different basal read-through potential, the opal stop codon causes the highest levels, the amber stop codon is in the middle and then the ochre stop codon causes the lowest rate of “leaked” read-through. The immediate downstream nucleotide is thought to promote stop codon read-through, a cytosine at this position promotes the read-through of the opal UGA to a frequency 4% in mammalian cells but results in a milder increase for the amber (2%) and ochre (0.5%) stop codon read-through (Manuvakhova et al., 2000). It is proposed that stop codon programming is dictated by these four nucleotides (the stop codon and +1 nucleotide) rather than the stop codon alone, the hierarchy of read-through potential for the downstream nucleotides inducer goes C > U > G > A (Jungreis et al., 2011). These results highlight the importance of the combination of the stop codon and the downstream nucleotide for programmed stop codon read-through. The eRF1 which as previously described is essential for terminating peptide synthesis was shown to specifically recognise stop codon sequences on mRNA, the factor also directly interacts with the 3' nucleotide immediately downstream of a stop codon (Bulygin et al., 2002). In addition to this when the stop codon and eRF1 bind during translation termination it causes the A site to accommodate the extra 3' downstream nucleotide and the mRNA adopts a compact configuration (Brown et al., 2015). It is possible that the release factor interacting with the adjacent cytosine creates unfavourable termination conditions therefore permitting stop codon read-through.

The amino acid inserted in place of termination at the stop codon is determined by which stop codon is in the read-through position, successful read-through occurs when the nc-tRNA only mismatches at one of the three positions of the codon to anticodon interaction; in this situation two bonds form between the two complementary base pairs but a third “wobble” position which cannot form a complementary interaction is ignored and the amino acid is incorporated into the

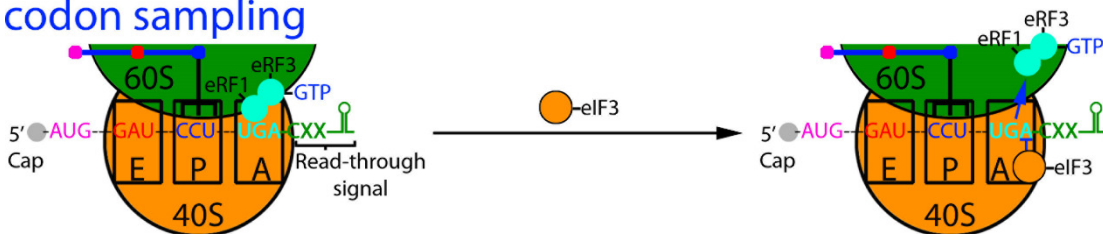
peptide chain regardless (Beier, 2001). The pool of which nc-tRNA capable of inserting at these sites varies between the three stop codons; the opal stop codon is mainly decoded by Trp or Cys tRNA and the amber and ochre are both decoded by Tyr, Gln and Lys tRNA. The preference for which nc-tRNA is incorporated in place of the stop codon during read-through is determined by the immediately 3' base following the stop codon, for example the UGA-cytosine base combination seen in alphavirus stop codon read-through preferably incorporates a cysteine residue (Beznosková et al., 2016). The nc-tRNA competes with the eRF1 for the A site of the ribosome containing the stop codon position, this competition mainly occurs when unfavourable termination conditions are created by the contextual stop codon read-through signals. The initiation factor eIF3 was shown to promote stop codon read-through in the presence of the downstream nucleotide read-through signals; when a stop codon is positioned in an unfavourable termination context the eIF3 associates with the pre-termination complex during the eRF1 two step sampling process. The eIF3 specifically alters the decoding property of 3<sup>rd</sup> nucleotide wobble position of the stop codon to both destabilise the interaction between the eRF1 and promote nc-tRNA recognition and incorporation at this position (Beznoskova et al., 2015) (**Figure 8**).

## A. Typical stop codon sampling by eRF1 to initiate translation termination

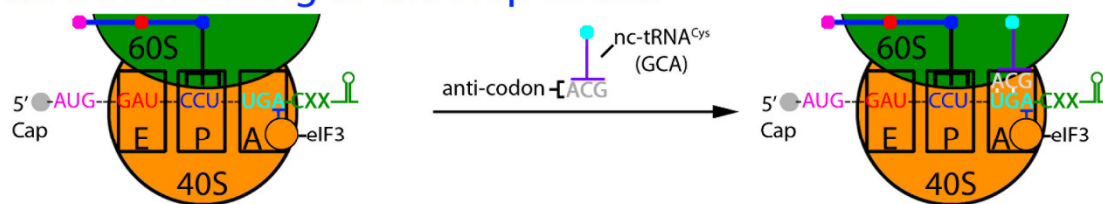


## B. Stop codon read-through

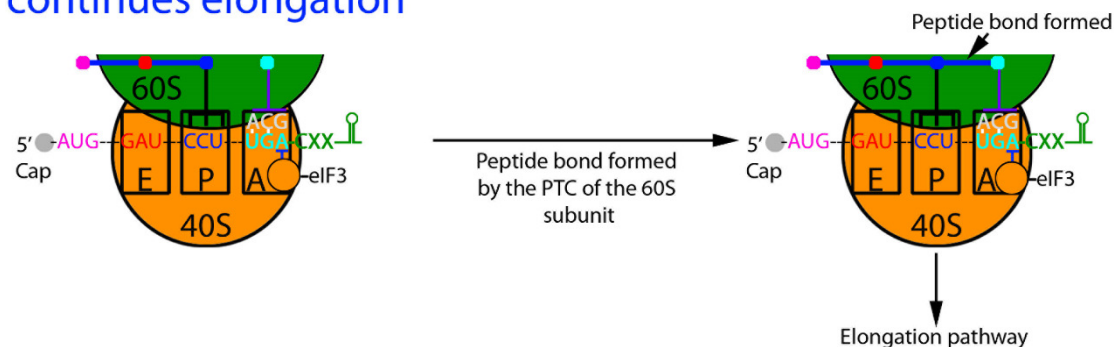
1) Read-through signals promote eIF3 mediated wobble position nucleotide (3rd) interference to disrupt eRF1 stop codon sampling



2) Wobble position nucleotide interference facilitates nc-tRNA binding to the stop codon



3) The incorporation of nc-tRNA at the stop codon continues elongation



**Figure 8: Potential stop codon read-through mechanism.**

(A) Typical stop codon decoding by eRF1:eRF3:GTP ternary complex as described previously. (B) Stop codon read-through mechanism. (1) Read-through signals near the stop codon promote eIF3 mediated disruption of the 3<sup>rd</sup> wobble position nucleotide of the stop codon. This inhibits eRF1 decoding of the stop codon. (2) The eIF3 interference also promotes nc-tRNA binding and successful recognition of the stop codon in the A site (in this case nc-tRNA<sup>cys</sup> to opal stop codon). (3) Incorporation of the nc-tRNA into the peptide chain by the PTC continues elongation until the next stop codon is reached.



The mechanism for stop codon recoding RNA structures is currently unknown but some speculated functions have been suggested; the structures could enhance read-through by directly interacting with the ribosome to enhance the incorporation of the nc-tRNA, inhibit eRF1 binding or by sterically hindering the termination factor (Firth and Brierley, 2012). As these structures are present in some but not all stop codon read-through scenarios, they may instead mainly function to synergise with additional read-through signals to enhance read-through.

### **1.5.5 Pharmacological modulation of stop codon read-through**

Compounds targeting stop codon read-through typically promote stop codon read-through rather than termination at this position, these compounds are typically used to target host genes which contain premature stop codons that results in C-terminally truncated non-functional proteins. The aminoglycoside family of antibiotics which include gentamicin and geneticin (G418) promote read-through of premature stop codons by interacting with rRNA of the ribosomal decoding centre to disrupt the proof reading during translation, for prokaryotic ribosomes this causes severe misincorporation of nc-tRNA throughout the elongation process. In eukaryotes the aminoglycosides do not cause significant miscoding during translation, they also promote the incorporation of nc-tRNA at the premature stop codons (Keeling et al., 2014).

Non-aminoglycoside antibiotics which promote stop codon read-through have been discovered, one compound of interest called ataluren (PTC124) is an anoxadiazole that was identified by high-throughput Firefly luciferase (Fluc) assay to promote stop codon read-through of the three stop codons. This compound was used to suppress the nonsense termination mutation in the dystrophin gene, which causes the condition Duchenne Muscular Dystrophy (DMD), in a human primary cell line and in a mouse model (Welch et al., 2007). The validity of ataluren was questioned when results were published suggesting that the compound increased read-through Fluc levels not by promoting read-through but instead directly preventing Fluc degradation (Auld et al., 2009). Despite this the use of ataluren to treat DMD was approved by the European Union, the compound was shown to promote read-through of non-luciferase based assays. The exact mechanism of the compound ataluren is unknown but

it is thought to interact with the ribosome to promote nc-tRNA incorporation (Siddiqui and Sonenberg, 2016).

The effect of these read-through promoting compounds on MuLV amber stop codon type 3 read-through during Gag-Pol expression was investigated, G418 and ataluren were shown to enhance the read-through of the stop codon past levels normally achieved using the read-through signal. This increased read-through altered the ratio of Gag: Pol expression which was shown to have a modest effect on virus production (Goff and Green, 2015).

### **1.5.6 Other virus mechanisms of recoding translation**

There are other mechanisms which are exploited by viruses to maximise the coding potential of small virus genomes. Some viruses utilise a cap independent translation initiation mechanism by utilising a large 5' RNA structure internal ribosome entry site (IRES), these structures are used to initiate translation on uncapped RNA genomes using only a proportion of the host initiation factors. These HCV or HCV-like IRES RNA structures directly recruit the host translation machinery such as the 40S ribosome subunit and eIF2/eIF3 (Lukavsky, 2009). A mechanism used to express alternative forms of proteins is the phenomenon of programmed frameshifting, this is where the reading frame is shifted in either the positive or negative direction by a base during translation to express an alternative coding sequence. The frameshift occurs at a slippery site of several nucleotides followed by a spacer which is followed by a downstream structural element (typically a pseudoknot) which enhances frameshifting. The RNA structure causes the ribosome to pause within the slippery site which causes the ribosome to shift from the current reading frame codon to then realign with the new codon that is in the different reading frame. Viruses use programmed frame shifting to modulate levels of viral protein expression during the virus lifecycle (Dinman, 2012).

### **1.6 Aims and objectives of the research**

CHIKV is considered a global threat due to high risk of outbreaks caused by the worldwide distribution of the *Aedes* mosquitoes. Naïve populations are at risk of epidemics and there are currently no available vaccines or antivirals which specifically target the virus. CHIKV causes severe morbidity in humans but causes lifelong asymptomatic infections in mosquitoes, the mechanism behind this difference in pathologies is unknown. Deeper understanding about the

CHIKV lifecycle in both hosts is crucial as this would reveal essential virus processes which could be targeted by antivirals. Identifying processes essential for CHIKV replication will provide the basis for rational drug and vaccine design. The focus of this study was to investigate stop codon read-through during CHIKV replication to gain insight into how this feature promotes replication in the human and mosquito hosts.

To do this the opal, ochre and amber stop codon were introduced into an ECSA strain of CHIKV which encodes an arginine at this position in the context of an SG replicon with dual luciferase reporter proteins. A panel of cell types were used to screen these constructs to investigate cell type specific phenotypes for the stop codon read-through mutants. These mutations were then investigated in the context of a full infectious clone of the ECSA CHIKV strain.

The second objective of this focus was to provide further insight into the mechanism of stop codon read-through for CHIKV in the context of mammalian host and mosquito vector. This was accomplished by experimentally determining the conserved stop codon read-through RNA structure in the context of the full infectious genome folded at temperatures corresponding with human and mosquito body temperature. Preliminary work investigating the impact of disrupting this RNA structure on CHIKV replication in a mammalian and mosquito cell line was also started in this research.

The final objective of this research was to develop an affinity tagged nsP3 in the context of virus lifecycle, this is an invaluable tool for investigating the protein as it paves the way for future studies investigating novel interactions and functions of the enigmatic viral protein. Studies identifying novel protein and RNA interactions will provide candidates for antiviral treatment.

The results presented here reveal cell type specific phenotypes for the stop codon mutants and insights into the formation of the stop codon recoding RNA structure within the mosquito vector. This research provides insights into the complex host specific adaption observed during alphavirus replication.

## Chapter 2: Materials and Methods

### 2.1 DNA and RNA manipulation

#### 2.1.1 Plasmid and virus constructs

A selection of CHIKV subgenomic replicons (SGRs) derived from an ECSA strain of CHIKV was kindly gifted by Prof Andres Merits (University of Tartu). The constructs included the following: Dual luciferase CHIKV SGR which possesses a renilla luciferase (Rluc) inserted into the nsP3 HVD coding region and firefly luciferase (Fluc) under control of the SG promoter. Another construct with a Gaussia luciferase (Gluc) reporter in the SG region with an untagged nsP3 was also gifted. A replication deficient Dluc CHIKV SGR was made in the lab previously, the active site Gly-Asn-Asn was substituted to Gly-Ala-Ala.

SG Gluc CHIKV replicons with FLAG-HA-streptavidin binding peptide or a FLAGX3His tags inserted into the hypervariable domain of nsP3 generated previously in the lab was also used in this study. A non-cytotoxic CHIKV replicon which contained a SG ZS green reporter, a puromycin acetyltransferase gene and a Rluc inserted into the nsP3 HVD was used to clone in a twin-strep tag (TST) in this study (Utt et al., 2015). An infectious clone of CHIKV derived from the Integration of Chikungunya REsearch (ICRES) strain based on the LR2006-OPY1 ECSA isolate of CHIKV (Tsetsarkin et al., 2006) was also kindly gifted by Prof Andres Merits. A replication deficient ICRES CHIKV with Gly-Ala-Ala substituted active site was used.

#### 2.1.2 Cloning strategy

The opal, ochre and amber stop codon were substituted into the Dluc CHIKV SGR using site directed mutagenesis as described in section 2.1.8. The stop codons were also incorporated into the ICRES CHIKV infectious clone, a two-step cloning strategy was used to do this. First a 1017 bp GeneART (Thermo Scientific) synthesised DNA fragment matching the ICRES CHIKV sequence between the KfII (4683) and AgeI (5684) except with two unique SpeI (5222) and SacI (5559) restriction sites was subcloned into the plasmid to generate an intermediate cloning product. Second, the SpeI (6173) and AgeI (6638) restriction

sites were used to isolate three fragments of nsP3 containing the stop codon positions from the stop codon containing Dluc CHIKV SGRs (opal, ochre and amber), these fragments were then subcloned into the SpeI SacI ICRES CHIKV using the same enzymes.

The TST was cloned into the ICRES CHIKV infectious clone using a two-step cloning method. A 129 bp GeneART (Thermo Fisher Scientific) DNA fragment of the TST sequence flanked by two SpeI restriction sites was synthesised, this was subcloned into a Fluc CHIKV SGR which contained a unique SpeI site (5222) as an intermediate plasmid for cloning. The TST-nsP3 fragment was excised from this plasmid using the KflI (4683) and the AgeI (5777) restriction sites, this fragment was sub cloned into the ICRES CHIKV plasmid using the equivalent restriction sites.

The Rluc NCT CHIKV SGR which can replicate non-cytotoxically in Huh7 cells was used to express nsP3 with no insert and TST nsP3. To do this DNA fragments were cloned into the BstBI (5567) and AgeI (6641) restriction sites on the plasmid. Two GeneART (Thermo Fisher Scientific) DNA fragments were synthesised, both fragments contained the BstBI and AgeI restriction sites and the sequences which matched the nsP3 sequences, one was designed so the Rluc reporter was removed to make the nsP3 with no insert. The other contained a TST insert in place of the Rluc reporter to make the TST nsP3 NCT CHIKV.

### **2.1.3 Replication of plasmids in bacteria**

DNA plasmids were transformed into Mix & Go Competent cells- DH5 alpha *Escherichia coli* (Z DH5 alpha) (Zymo) as per the manufacturer's protocol. Briefly, 1 ng of DNA was mixed with 50 µl of DH5 alpha bacteria. The mixture was placed on ice for 5 minutes, the whole mixture was spread on set Luria-Bertani (LB) agar plates with 100 µg/ml working concentrations of ampicillin. These plates were cultured overnight at 37°C before picking colonies. Individual colonies were picked from these plates to inoculate 5.5 ml starter cultures made up of LB broth with 100 µg/ml working concentration of ampicillin. The starter cultures were incubated for 8 hours at 37°C. A glycerol stock of the bacteria was made by mixing 500 µl of the bacteria mixture with 500 µl of a 50 % glycerol suspension

diluted in dH<sub>2</sub>O in a cryovial. The cryovial was quickly transferred to the -80°C freezer for storage.

The starter culture was used to inoculate 5 ml or 50 ml LB broth at a dilution of 1 in 1000. The 5 ml cultures were pelleted using a microcentrifuge (6800 x g, 2 minutes, room temperature), these bacteria pellets were purified using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) as per the manufacturer's guidelines. The 50 ml cultures were pelleted using a low speed centrifuge (4600 x g, 10 minutes at 4°C), the DNA was extracted from these pellets using the GeneJET Plasmid Midiprep Kit (Thermo scientific) as per the recommended protocol. The DNA samples were stored at – 20°C.

#### **2.1.4 DNA and RNA Quantification**

Nucleic acid concentration for DNA and RNA was measured using a NANODrop 1000 Spectrophotometer (Thermo Scientific). Absorbance at 260 nm is used to calculate the ng/ µl. The 260/280 and 260/230 ratios are calculated to provide an assessment of DNA or RNA purity.

#### **2.1.5 DNA and RNA agarose gel electrophoresis**

The integrity of all restriction digested DNA was verified using gel electrophoresis. DNA integrity was visualised using 1% (w/v) agarose gels in 1X TAE buffer (20 mM acetate, 1 mM EDTA and 40 mM Tris) set with SYBR Safe DNA gel stain™ (Thermo Fisher Scientific) diluted 1:10000. Set gels were placed into gel tanks with 1X TAE buffer. DNA samples were diluted using Gel loading dye, purple (6X) (NEB) before loading alongside the 1 kb or 100 bp Hyperladder™ (Bioline). The gels were resolved at 80 V for 1 hour before imaging with UV in the GENE GENIUS Bioimaging system (Syngene).

The integrity of all RNA used in this study was validated using gel electrophoresis. RNA agarose gels were used to visualise RNA sample integrity, these gels were made using 1% (w/v) agarose diluted in 1X MOPS buffer (2 mM sodium acetate, 1 mM EDTA, 20 mM MOPS) and 1.4 ml formaldehyde (Stock solution at 37% concentration). These gels were set with SYBR Safe DNA gel stain™ (Thermo Fisher Scientific) diluted 1:10000 as before. RNA samples were diluted with RNA

gel loading dye (2X) (Thermo Fisher Scientific) and RNase free H<sub>2</sub>O, these samples were boiled at 65°C for 5 minutes to denature RNA structures. The gels were placed into gel tanks containing 1X MOPS buffer, the samples were loaded on these gels with an ssRNA ladder (NEB). Gels were resolved at 80 V for 45 minutes before visualising with UV light in the GENE GENIUS Bioimaging system (Syngene).

### **2.1.6 Linearisation and purification of DNA**

The CHIKV SGR and CHIKV ICRES infectious clone DNA plasmids were linearised overnight at 37°C using the NotI-HF<sup>®</sup> restriction enzyme (NEB). The reaction was inactivated by adding 1 µl of 10% SDS, each reaction was made up to a total volume of 500 µl using RNase free H<sub>2</sub>O. An equal volume of UltraPure™ Phenol: Chloroform: Isoamyl alcohol (25: 24: 1) (Invitrogen) was added before vortexing for 1 minute. The samples were centrifuged (13,000 xg, 5 minutes, room temperature) before collecting the aqueous layer in a new tube. Next an equal volume of chloroform was added before vortexing for a further 1 minute and centrifuging again (13,000 xg, 5 minutes, room temperature). The aqueous phase was collected and incubated with 3M sodium acetate pH 5.3 (1 in 10), and an equal volume of 100% isopropanol. The samples were incubated overnight at -20°C. In the following day the DNA was precipitated by centrifugation (13,000 xg, 20 minutes, 4°C). The supernatant was removed, and pellets were washed once with 70% ethanol (13,000 xg, 5 minutes, 4°C). The ethanol was removed, and the pellet was left to air dry. The linearised DNA pellet was resuspended in RNase free H<sub>2</sub>O before storage at -20°C. All digested DNA was visualised using DNA gel electrophoresis (section 2.1.5).

### **2.1.7 *In vitro* transcription of RNA**

The purified linearised DNA was used as a template for capped *in vitro* RNA transcription using the mMESSAGE mMACHINE® SP6 kit (Thermo Fisher Scientific), 1 µg of DNA was used per 20 µl reaction performed as described by the manufacturer. RNA reactions were incubated at 37°C for 3 hours, after this reaction mixture was treated with TURBO DNase at 37°C for 15 minutes to remove template DNA.

RNA was purified using lithium chloride (LiCl) precipitation, to do this 30 µl of RNase free H<sub>2</sub>O and 30 µl of LiCl was added to each RNA reaction. This mixture was incubated at -20°C for 3 hours before centrifugation (maximum speed, 15 minutes, 4°C). The pellet was washed with 70% ethanol made using RNase free H<sub>2</sub>O before centrifugation (maximum speed, 5 minutes, 4°C). The RNA pellet was resuspended in up to 20 µl of RNase free H<sub>2</sub>O, the RNA suspensions were aliquoted 1 µg per tube before storage at -80°C.

### **2.1.8 Site directed mutagenesis**

Amino acid substitutions into the CHIKV SGR and ICRES CHIKV sequences were performed using the Q5® Site-Directed Mutagenesis Kit (NEB). Changes in the DNA sequence were made using back to back forward and reverse primers designed using the NEBasechanger™ software (see appendix table 1 for list of primers used). The reactions were set up as suggested by the manufacturer's protocol. The DNA products were transformed into zDH5 alpha *E. coli* as described in section 2.1.3. Successful mutagenesis was confirmed using sequencing reactions. Restriction enzymes with flanking (upstream and downstream) recognition sites of the mutated sequence were used to subclone the region into the parent plasmid backbone to ensure no offsite mutations were unintentionally incorporated into the mutants.



### 2.1.9 Sub-cloning and ligation

The vector backbone and insert fragments were generated using restriction enzymes which resulted in complementary overhangs for ligation. The DNA was incubated with the restriction enzymes at the individual optimal temperatures for recommended incubation times. The backbone fragments were incubated with calf intestinal phosphatase (CIP) (NEB) at 37°C for 1 hour to prevent re-ligation. The digested DNA fragments were separated using gel electrophoresis, the desired DNA fragments were excised from the gel for purification using the Monarch® DNA Gel Extraction Kit (NEB). The purified fragments were ligated overnight at 16°C with the T4 DNA Ligase (NEB), the correct amount of insert to 50 ng of vector was calculated using the NEBcloner® software (NEB). The zDH5 alpha *E. coli* was transformed using 5 µl of the ligated DNA. The bacteria were recovered using 100 µl of LB broth for 1 hour at 37°C. This mixture was plated on the LB agar plates containing ampicillin, these plates were incubated overnight at 37°C. Colonies were picked the following morning.

## 2.2 Tissue culture

### 2.2.1 Cell lines

A panel of cell lines were used in this project to investigate the replication of CHIKV SGR and CHIKV ICRES (Roberts et al., 2017). The mammalian cells used were the mouse C2C12 cells (muscle, myoblast) (Parker et al., 2016) and hamster BHK-21 cells (kidney, fibroblast) (Pohjala et al., 2011). The human cell lines used were the Huh7 cells (liver, hepatocellular carcinoma) (Ross-Thriepland and Harris, 2014); human SVG-A cells (brain, astroglia) (Schweighardt et al., 2001) and the human RD cells (muscle, rhabdomyosarcoma) (Hinson et al., 2013). The mosquito cell line C6/36 (*Ae. albopictus*, embryonic) used in this study had a mutant Dcr2 gene which results in a defective RNAi response (Siu et al., 2011).

### 2.2.2 Passaging cell lines

Mammalian cells were maintained in complete Dulbecco's modified Eagle's medium (DMEM) (Sigma) supplemented with 10% Fetal Calf Serum (FBS) (Gibco) (or 20% FBS for C2C12 cells), 1% Penicillin-Streptomycin (100 IU penicillin/ml and 100 µg/ml streptomycin) (Gibco), 10% Non-essential amino acids (NEAA) (Thermo Fisher Scientific) and 10 mM Gibco HEPES (Thermo Fisher Scientific). To passage cells the complete medium was removed before washing cells with phosphate buffered saline (PBS) (VWR Lifescience), the cells were then incubated with trypsin EDTA (Sigma) at 37°C for 5 minutes. Complete medium was added to the resuspension to inactive the trypsin, cells were pelleted with centrifugation (1000 xg, 5 minutes, room temperature) before resuspending in 10 ml of complete medium. The appropriate proportion of cells was used for passage or to seed cells for experiments.

The C6/36 mosquito cell line was maintained in Leibovitz's L-15 Medium (Thermo Fisher Scientific) supplemented with 10% FBS, 1% Penicillin-Streptomycin (100 IU penicillin/ml and 100 µg/ml streptomycin) (Gibco), and 10% Tryptose phosphate broth (Thermo Fisher Scientific). To passage the insect cells first the media was removed before washing cells with PBS (VWR, Lifescience). The cells were then scraped into new complete media using Corning® cell scrapers (Sigma-Aldrich), this suspension was then used to passage cells or seed cells for experiments. Mammalian cells were maintained in humidified incubator at 37°C with 5% CO<sub>2</sub> and insect cells were maintained in incubator at 28°C.

### 2.2.3 Transfection using lipofectamine 2000

Cells required for transfection were seeded as  $1 \times 10^5$  cells per well of a 24 well plate. The following morning the media was removed from each well and the cells were washed using 1 ml of DEPC PBS before replacing the media with 500 µl of opti-MEM™ (Thermo Fisher Scientific). Master mixtures of RNA, lipofectamine 2000 (Thermo Fisher Scientific) and opti-MEM™ were made: Each well required a mixture of 250 ng of RNA, 1 µl of lipofectamine and 100 µl of opti-MEM™. The complete mixture was left for 5 minutes at room temperature to incubate. After this incubation 100 µl of the transfection master mix was added to each well, cells

were left for 4 hours in cell incubator (37°C or 28°C). After the incubation the transfection reagent was removed, cells were washed with 1 ml of DEPC PBS and 1 ml of complete media was added per well.

The same procedure was done for 6 well plate transfections however  $4 \times 10^5$  cells per well were seeded and the master mixes were made up so that each well contained 1 µg of RNA, 5 µl of lipofectamine 2000 and 500 µl of opti-MEM™. The whole mixture was added per well for transfection.

#### **2.2.4 Electroporation**

Cells were trypsinised as described in section 2.2.2, the cell pellet after centrifugation (1000 xg, 5 minutes, room temperature) was washed in 10 ml of DEPC PBS not complete media. The cells were pelleted again (1000 xg, 5 minutes, room temperature), this pellet was resuspended in 5 ml of ice cold DEPC PBS. These cells were counted using a hemocytometer (Neubauer), cell suspensions were pelleted (1000 xg, 5 minutes, room temperature) and then resuspended in ice cold opti-MEM™ at a concentration of  $2.5 \times 10^6$  cells/ml. A precooled Gene Pulser 4 mm electroporation cuvette (BIO-RAD) with  $1 \times 10^6$  cells in 400 µl and 1 µg of RNA was assembled. The cuvette was electroporated using a Gene Pulser Xcell™ Electroporation systems (BIO-RAD) with the following square wave electroporation settings (260 V, 25 ms, 1 pulse, 0 seconds delay, in a 4 mm cuvette). The electroporated cells were resuspended in 10 ml of complete media to seed the cells in the appropriate plates or flask.

#### **2.2.5 Generating and maintaining a cell line stably expressing the non-cytotoxic CHIKV replicon**

Huh7 cells which stably expressed the NCT CHIKV replicons (nsP3 and TST nsP3) were generated as follows: Huh7 cells were electroporated with the replicon RNA as per section 2.2.4. Successful transfection was tested after 24 hours by detecting ZS green expression with the 470 nm wavelenth using the Nikon Eclipse TS100 microscope (Nikon) with CoolLED pE200 adapter (CoolLED). Cells expressing the NCT CHIKV replicons were selected using 5 mg/ml working concentrations of puromycin (Thermo Fisher Scientific) diluted in complete media. The puromycin media was replaced every 3 days until the stable

cell line was selected, this was seen when all cells were expressing ZS green and survived puromycin treatment.

### **2.2.6 Protein extraction**

Protein lysates were harvested using Glasgow lysis buffer (GLB) (10 mM PIPES pH 7.2, 120 mM KCl, 30 mM NaCl, 5 mM MgCl<sub>2</sub>, 1% Triton-x-100 and 10% glycerol) supplemented with protease inhibitors (leupeptin (1 µg/ml), pepstatin A (1 µg/ml), aprotonin (2 µg/ml) and AEBSF (0.2 mM)). To do this the cells were first washed with 10 ml of PBS, the cells were then scraped into 5 ml PBS before pelleting (1000 xg, 3 minutes, room temperature). The cell pellets were then resuspended in 500 µl of PBS to transfer into 1.5 ml Eppendorf tubes, this suspension was centrifuged (4000 xg, 5 minutes, room temperature) to pellet the cells again. This pellet was lysed by resuspending the pellet in 500 µl of GLB, the lysates were incubated on ice for 30 minutes before centrifugation (4000 xg, 5 minutes, 4°C). The supernatant was collected and stored at -80°C.

## **2.3 Molecular Biology**

### **2.3.1 Dual luciferase reporter assay**

Dual luciferase assays were performed using 24 well plates transfected with CHIKV replicon RNA as described in section 2.2.3. These cells were washed with 1 ml of PBS in triplicate before 100 µl of passive lysis buffer (PLB) (Promega) was added per well. The plate was left on rocker for 15 minutes to lyse before collecting the lysates. The samples were analysed using the Dual-luciferase<sup>®</sup> Reporter Assay System (Promega), 20 µl of sample was loaded on to a Corning<sup>®</sup> 96 well plate (white bottom) (Merck). The Rluc signal and Fluc signal were detected using 50 µl of LARII and subsequently 50 µl of Stop-N-Glo<sup>®</sup> respectively. The signal of these two reporters were detected using the FLUOstar Omega Microplate reader and Optima software (BMG LABTECH).

### **2.3.2 DNA sequencing**

DNA plasmids and cDNA products were sent to GENEWIZ® for Sanger sequencing using primers designed to specifically anneal to the CHIKV genome (Appendix table 1). The results were analysed using the DNA Dynamo Sequence analysis software (Blue Tractor Software).

### **2.3.3 RNA extraction**

Trizol Reagent (Thermo Fisher Scientific) was used to extract RNA from cells and from supernatants. Cells were first washed with 1 ml ice cold DEPC PBS before 1 ml of Trizol was added. Cells were scraped into the Trizol before transferring the suspension to 1.5 ml Eppendorf tubes. Tubes were left to sit at room temperature for 5 minutes. Chloroform (250 µl) was added to the solution, the tube was shaken vigorously for 15 seconds and incubated at room temperature for 5 minutes before centrifugation (10,000 rpm, 5 minutes, room temperature). The aqueous layer was transferred to a new Eppendorf tube, 550 µl of isopropanol was added. Tubes were left to incubate at room temperatures for 5 minutes before another centrifugation step (14,000 rpm, 30 minutes, room temperature). The tubes were then placed on ice, the isopropanol was discarded, and pellet was washed with 1 ml of 75% ethanol made with DEPC H<sub>2</sub>O. The ethanol was removed (9500 rpm, 5 minutes) before the RNA pellets were left to air dry before being resuspended in RNase free H<sub>2</sub>O. RNA was extracted from supernatants using the same process except 800 µl of Trizol reagent was added to 200 µl of supernatant.

### **2.3.4 Reverse transcription PCR**

RNA extracted from cells was prepared for sequencing reactions using reverse transcription PCR. Reverse transcription reactions were performed using the recommended protocol for SuperScript™ III Reverse Transcriptase (Thermo Fisher Scientific). First strand cDNA synthesis for sample RNA was done using 200 ng of Random Hexamer primers (Thermo Fisher Scientific) using the recommended protocol. The cDNA was used as a template for PCR using primers specific to the CHIKV genome (see appendix table 1) to amplify specific

regions of sequence. The PCR reactions were made up in 50 µl reactions containing 1X PCR buffer (20 mM Tris-HCL (pH 8.4), 50 mM KCl), 1.5 mM MgCl<sub>2</sub>, 1 mM dNTP mixture, 1 µM forward primer, 1 µM reverse primer, 2U of Vent DNA polymerase, 2 µl of cDNA product and reaction made up to 50 µl with dH<sub>2</sub>O. PCR reactions were carried out with the following cycles; 96°C for 2 mins, 20 cycles of (96°C for 15 seconds, 58°C for 30 seconds, 72°C for 30 seconds), 72°C for 10 minutes and a 4°C hold step. The PCR products were analysed using gel electrophoresis as per section 2.1.5 and products were sequenced as per section 2.3.2.

### **2.3.5 Quantification of protein concentration**

Protein concentration of lysed protein extracts were calculated using the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific). The recommended protein standards for this assay were made up with Bovine Serum Albumin (Merck) in GLB lysis buffer. The protein standards and the tested proteins were loaded clear bottom Corning® 96 well plates (Merck), the BCA reagent was added to these samples as recommended by the manufacturer. The absorbance at 570 nm was detected using an infinite F50 plate reader and the Magellan software (Tecan).

### **2.3.6 Streptactin column affinity purification**

Protein extracts containing the TST nsP3 were column purified using the Streptactin gravity flow columns (IBA lifesciences). The procedure outlined by the manufacturer was followed. Columns were used to purify 1 mg of protein lysates in a total volume made up to 1 ml made up with GLB lysis buffer. The column was washed with five 1 ml washes with the commercial wash buffer (IBA lifesciences). Proteins were eluted by six applications of 500 µl elution buffer (IBA lifesciences) which contains 2.5 mM desthiobiotin.

### 2.3.7 Streptactin batch purification

Batch purification of the TST tagged nsP3 was done using the 50% streptactin sepharose resin (IBA lifesciences). To do this 100  $\mu$ l of the resin suspension was used per 1 mg of protein lysates. The protocol was as follows, the 100  $\mu$ l 50% streptactin sepharose resin was centrifuged (3000 xg, 5 minutes, 4°C) and the supernatant was removed. The resin was washed with 500  $\mu$ l commercial wash buffer (IBA lifesciences) and then 500  $\mu$ l of GLB lysis buffer (3000 xg, 5 minutes, 4°C). The resin was next incubated with 1 mg of protein lysate in a total volume of 500  $\mu$ l (made up with GLB) overnight on a rotator at 4°C. The following day the flowthrough was collected (3000 xg, 5 minutes, 4°C), the resin was washed with 500  $\mu$ l of commercial wash buffer (IBA lifesciences) 4 times (3000 xg, 5 minutes, 4°C). Proteins were released from the resin by boiling in 1X SDS running buffer (4% SDS, 40% glycerol, 10%  $\beta$ - mercaptomethanol, 120 mM Tris-HCL ph 6.8 and bromophenol blue) for 5 minutes.

### 2.3.8 Protein separation using SDS-PAGE gels

Protein samples were separated using SDS-PAGE gels, these gels were made in the lab using the following recipe Resolving gels (10%) were made using 3.3 ml 30% acrylamide, 2.5 ml of 1.5 M Tris-HCL pH 8.8, 4 ml of dH<sub>2</sub>O, 100  $\mu$ l of 10% SDS, 100  $\mu$ l of 10% (w/v) ammonium persulphate (APS) and 10  $\mu$ l of TEMED. This mixture set in glass plates held by mini-gel apparatus on a casting stand, isopropanol was used as an overlay. The gel was left to set before removing the isopropanol. Next the 5% stacking gel was made using the following recipe, 0.83 ml 30% acrylamide, 0.63 ml 1M Tris-HCL pH 6.8, 3.4 ml dH<sub>2</sub>O, 50  $\mu$ l 10% SDS, 50  $\mu$ l (APS) and 5  $\mu$ l TEMED. The stacking gel and the desired comb insert was used to form complete SDS-PAGE gel when set.

The gels were used for protein separation using the following method. The gels were assembled in gel tanks and the 1X SDS-PAGE running buffer (50 mM Tris, 380 mM glycine, 0.1% (w/v) SDS) was added, Proteins were diluted in 5X SDS-PAGE loading buffer such that the working concentration was 1X (4% SDS, 40% glycerol, 10%  $\beta$ - mercaptomethanol, 120 mM Tris-HCL ph 6.8 and bromophenol

blue) before boiling samples at 95°C. The samples were added to the wells of the gel and the proteins were separated at 180 V for 60 minutes.

### **2.3.9 Western blot**

The SDS-PAGE gels (section 2.3.8) were soaked in towbin buffer (25 mM Tris, 192 mM glycine and 20% (w/v) methanol) after removing from the gel tank. One gel size of PVDF membrane (Millipore Immobilon FL) was cut and activated in 100% methanol for 20 seconds. Thick blotting paper was soaked in towbin buffer. The semi-dry transfer blot machine was assembled as directed by the manufacturer. The transfer was performed at 15 V for 1 hour. After transfer the membrane was blocked using 50 % Licor Blocking buffer (LI-COR) and 50% TBS (25 mM Tris, 137 mM NaCl) for 1 hour. The membrane was washed using TBS-Tween (25 mM Tris, 137 mM NaCl, 0.1% Tween-20) (TBS-T) for 5 minutes (repeated 4X). Primary antibodies (see appendix table 5) were mixed with 25% LI-COR Blocking buffer and 75% TBS to make up desired antibody concentration. The membranes were incubated with primary antibody solutions overnight at 4°C in 50 ml falcon tubes on a rotator. The membranes were then washed with TBS-T 4 times. Secondary antibodies were resuspended in solutions containing 25% LI-COR Blocking buffer and 75% TBS. Both the IRDye® 800CW Donkey anti-Rabbit IgG (LI-COR) and the IRDye® 680RD Donkey anti-Mouse IgG (LI-COR) were used where appropriate. The membranes were incubated with the secondary antibodies for 1 hour before washing as before. The western blots were left to dry between filter papers before imaging with an Odyssey LI-COR imaging system (LI-COR).

### **2.3.10 Genome alignment**

Genome sequences of CHIKV and related alphaviruses were acquired from the National Center for Biotechnology Information (NCBI) GenBank software (NCBI Resource Coordinators, 2018). These sequences were aligned using the MAFFT Multiple sequence alignment software (Katoh et al., 2017) and the resultant alignment was viewed using the Jalview software (Waterhouse et al., 2009).



## **2.4 Infectious clone of CHIKV experiments**

### **2.4.1 Generating virus stocks by electroporation**

Capped RNA of the infectious clones of CHIKV generated via *in vitro* transcription (section 2.1.7) was electroporated into BHK cells (section 2.2.4). The supernatant was collected either 24 or 48 hours post electroporation, the virus stocks were clarified (1000 xg, 5 minutes, room temperature) before aliquoting in 1.5 ml Eppendorf tubes. Virus stocks were stored at – 80°C.

### **2.4.2 Generating virus stocks using Lipofectamine 2000 transfection**

C6/36 cells were seeded  $4 \times 10^5$  cells each well of a 6 well plate, the following day these cells were transfected with 1 µg of RNA, 5 µl of lipofectamine 2000 and 500 µl of opti-MEM as described in section 2.2.3. The transfection mixture was removed after 4 hours and the wells were washed with PBS. Complete media (2 ml) was added to the wells. Virus stocks were harvested after 48 hours post transfection, the virus stocks were clarified (1000 xg, 5 minutes, room temperature) before aliquoting in 1.5 ml Eppendorf tubes. Virus stocks were stored at – 80°C.

### **2.4.3 Plaque assay**

BHK cells were seeded in either 6 or 12 well plates with  $4 \times 10^5$  or  $2 \times 10^5$  cells per well respectively. The following day serial dilutions of the virus stocks from neat up to  $10^{-6}$  dilution in complete media were performed. The media was removed from each well and the cells were washed with PBS, 100 µl of the virus dilutions were added to corresponding wells. These plates were incubated at 37°C for 5 minutes on a rocker, these plates were transferred to the 37°C incubator for 1 hour. The virus was removed, and cells were washed gently with 1 ml of PBS. An overlay of 2 ml of the 1.6% methyl cellulose (MC) mixed with complete media (1:1) for 6 well plates or 1 ml for 12 well plates. The plaque

assays were incubated in 37°C incubator for 48 hours. The MC overlay was removed, and cells were fixed with 4% paraformaldehyde (PFA) for 30 minutes. After fixation the PFA was removed and cells were stained with 0.5% (w/v) crystal violet solution. The assays were washed with water until plaques were visible.

#### **2.4.4 Infectious centre assay (ICA)**

BHK cells were seeded as  $4 \times 10^5$  cells per well of a 6 well plate on the day before experiment. Another batch of BHK cells was trypsinised and pelleted (1000 xg, 5 minutes, room temperature). The cell pellet was washed with 10 ml of DEPC PBS before pelleting the cells again (1000 xg, 5 minutes, room temperature). The cells were resuspended in ice cold DEPC PBS and were counted. The cells were pelleted again (1000 xg, 5 minutes, room temperature) and resuspended in ice cold opti-MEM at a density of  $2.5 \times 10^6$  cells per ml. Two electroporation reactions were performed per infectious clone. To do this, two pre-chilled electroporation cuvettes each containing  $1 \times 10^6$  cells in 400  $\mu$ l of optimum and 1  $\mu$ g of RNA were electroporated (square wave protocol, 260 V, 25 ms, 1 pulse, 0 seconds delay, in a 4 mm cuvette). The electroporated cells were pooled and a dilution series going from  $10^{-1}$  to  $10^{-6}$  using complete media. The remaining neat electroporated cells were seeded separately in a 6 well plate to harvest virus stocks (section 2.4.1) after 24 hours. The diluted transfected cells (1 ml) were seeded on to the pre-plated 6 well plates of BHK cells. The cells were incubated in the 37°C cell incubator for 1 hour, the media was removed and then 2 ml of the 1.6% MC mixed with complete media (1:1) was added. The cells were incubated at 37°C for 48 hours. After this the MC mixture was removed and cells were fixed with 10% PFA for 30 minutes. The cells were stained with 0.5% (w/v) crystal violet, the plates were washed with water until plaques were seen.

#### **2.4.5 Virus infection for protein extraction**

T75 flasks were seeded with  $2 \times 10^6$  cells with either Huh7 or RD cells. The media was removed from these cells, the cells were washed with 5 ml of PBS and the cells were then infected with the viruses at a MOI of 10 in a total volume of 2 ml. The cells were incubated on rocker at room temperature for 10 minutes, cells were incubated in the 37°C cell incubator for 1 hour. The virus was removed, and

cells were washed with 5 ml of PBS, before replacing with 10 ml of complete media. After 24 hours the media was removed, cells were washed with 5 ml of PBS. The cells were lysed in GLB as before: Cells were scraped into 5 ml of PBS and then pelleted (1000 xg, 3 minutes, room temperature). The cell pellets were then resuspended in 500 µl of PBS and transferred to 1.5 ml Eppendorf tubes for centrifugation (4000 xg, 5 minutes, room temperature). The cell pellet was lysed with 500 µl of GLB supplemented with protease inhibitors, the lysates were incubated on ice for 30 minutes before centrifugation (4000 xg, 5 minutes, 4°C). The supernatant was collected and stored at -80°C.

#### **2.4.6 One step growth curve**

RD or BHK cells were seeded in 12 well plates as  $1 \times 10^5$  cells per well a day before the experiment. The media was removed, and cells were washed with 1 ml of PBS. Each well was infected with chosen virus at a MOI of 0.01 in a total volume of 1 ml diluted with complete media. The cells were incubated on rocker at room temperature for 5 minutes, a 0 hour timepoint was taken by transferring the supernatant from a well into a 1.5 ml Eppendorf. The samples were centrifuged (4000 xg, 5 minutes, room temperature) before transferring into a new 1.5 ml Eppendorf before storage at -80°C. Additional timepoints were taken at 4, 12, 24 and 48 hours post infection. The virus stocks were plaque assayed on BHK cells (section 2.4.3).

#### **2.4.7 RNA extraction from transfected cells**

BHK cells were electroporated as per section 2.4.1 and C6/36 cells were transfected using lipofectamine 2000 as per section 2.4.2. The remaining transfected cells left after the virus stocks were harvested were used to extract intracellular RNA when needed. To do this the cells were scraped into 5 ml of PBS before pelleting (1000xg, 5 minutes, room temperature), the cell pellets were resuspended in 500 µl of PBS before transferring to a 1.5 ml Eppendorf. The cells were pelleted (4000xg, 5 minutes, room temperature) and then incubated with 1

ml of Trizol Reagent (Thermo Fisher Scientific) for 30 minutes. The mixture was then extracted as per manufacturer's protocol (section 2.3.3).

## **2.5 Nucleic acid structure determination**

### **2.5.1 Mfold prediction of RNA structure**

RNA structures for the 677 base long RNA sequence (positions 5218 to 5894) from the ICRES CHIKV genome were predicted using the Mfold software (Zuker, 2003). The RNA structures were simulated for 37°C and 28°C, the structures were annotated by p-num values (the table 4 of the P-Num values/colours scheme from the Mfold software can be found in the appendix). The diagrams of predicted structure and vienna files dot bracket for the predicted RNA structure were obtained.

### **2.5.2 *In vitro* transcription of uncapped ICRES CHIKV RNA**

Uncapped RNA for the RNA structure experiments were generated using the MEGAscript™ SP6 kit (Thermo Fisher Scientific) using the manufacturers protocol. Each reaction was performed using 1 µg of the linear WT ICRES CHIKV DNA as a template. The resultant RNA was purified using LiCl precipitation as described in section 2.1.7. The RNA pellet was resuspended in 0.5X TE buffer (Thermo Fisher Scientific) before being stored in 10 µg aliquots at -80°C.

### **2.5.2 QuSHAPE**

Reverse primers with either a 5' FAM or 5' HEX fluorescein modification which target upstream of the predicted structure of the CHIKV sequence were synthesised (IDT DNA technologies). Three primers which bind downstream of the RNA structure were chosen to target an unstructured predicted region, each primer was used in independent experiments. Three primers were designed as this would theoretically maximise the chance of successfully binding to the RNA and can provide redundancy. These primers were resuspended in RNase free H<sub>2</sub>O at a concentration of 10 µM and stored at -20°C, the primers can be seen in appendix table 1. The QuSHAPE reactions were performed using the uncapped

WT ICRES CHIKV RNA as the use of capped RNA for the reaction would considerably increase the cost of the experiment. The protocol was performed as follows: The 10 µg of uncapped RNA was made up to 20 µl using 0.5X TE buffer, these samples were folded by first heating the samples to 95°C for 2 minutes and then incubating on ice for 2 minutes. Next 103 µl of RNase free H<sub>2</sub>O and 45 µl of RNA folding buffer (300 mM HEPES pH 8.0, 300 mM NaCl and 20 mM MgCl<sub>2</sub>) and 2 µl of RNasin Ribonuclease inhibitor (Promega) was added to the RNA. The samples were incubated at 37°C or 28°C in a heat block for 30 minutes. The reactions were split into NMIA positive (+VE) and negative (DMSO) (-VE) reaction using half the sample each (72 µl), 8 µl of NMIA or 8 µl of DMSO was added to the reaction. These samples were heated at 37°C or 28°C for 30 minutes. The samples were then ethanol precipitated by adding 4 µl of NaOAc (3M, pH 5.2), 2 µl 100 mM EDTA and 350 µl 100% EtOH. These samples were incubated at -80°C overnight.

The following morning the RNA was pelleted (14,000 xg, 30 minutes, 4°C) and washed with 70% EtOH (14,000 xg, 5 minutes, 4°C). The pellets were resuspended in 10 µl of 0.5X TE buffer. The NMIA positive and negative (DMSO) reactions were set up as follows: 5 µl of the +VE or -VE reaction, 6 µl of RNase free H<sub>2</sub>O and 1 µl of the 10 µM 5' FAM labelled reverse primer were added to a PCR tube. A PCR thermocycler was used to run the following cycle: 85°C for 1 minute, 60°C for 10 minutes and 35°C for 10 minutes. A reverse transcription master mix which contained the following per reaction: 4 µl first strand buffer, 1 µl DTT, 0.5 µl RNasin ribonuclease inhibitor, 0.5 µl RNase free H<sub>2</sub>O, 1 µl dNTPs (10 µM) and 1 µl SuperScript™ III reverse transcriptase was added to the samples. These samples were incubated at 52°C for 30 minutes in the PCR thermocycler. After the incubation 1 µl of 4 M NaOH was added and reactions were incubated at 95°C for 3 minutes, 2 µl of 2 M HCl was added and the samples were cooled on ice. The RNA was ethanol precipitated overnight at -80°C.

The sequencing ladder reaction was prepared as follows: 10 µg of uncapped RNA, 1 µl of 10 µM 5' HEX reverse primer was added to a PCR tube. The tubes were incubated at 85°C for 1 minute, 60°C for 10 minutes and 35°C for 10 minutes in a PCR thermocycler. The reverse transcription master mix (per reaction: 4 µl first strand buffer, 1 µl DTT, 0.5 µl RNasin ribonuclease inhibitor, 0.5 µl RNase free H<sub>2</sub>O, 1 µl dNTPs (10 mM), 1 µl ddATP (10 mM) (Roche) and 1 µl

SuperScript™ III reverse transcriptase was added to the sequencing ladder. The reaction was incubated at 52°C for 30 minutes, 1 µl of 4 M NaOH was added and reactions were incubated at 95°C for 3 minutes and 2 µl of 2M HCl was added and tubes were cooled on ice. The ladder was ethanol precipitated as before.

The following morning the +VE, -VE and RNA sequencing ladder samples were pelleted (14,000 xg, 30 minutes, 4°C) and washed in 70 % EtOH (14,000 xg, 5 minutes, 4°C). The RNA pellets were resuspended in 40 µl deionised formamide. All samples were heated at 65°C for 10 minutes and then vortexed. Each 40 µl +VE and -VE SHAPE reaction sample were mixed with 10 µl RNA sequencing ladder before sending to the DNA Sequencing and Services (University of Dundee) on dry ice for analysis with capillary electrophoresis (two capillary analysis where one tube contained the +VE reaction and the other the -VE reaction). The SHAPE data was analysed using the QuSHAPE software, the CHIKV RNA fasta used to generate Mfold models was used the template (Karabiber et al., 2013), the NMIA reactivities were overlaid on the Mfold predicted RNA structure using the VARNA software (Darty et al., 2009). The colour styles for NMIA reactivity was set so the 0 to 0.3 thresholds were coloured white, the 0.3 to 0.7 were coloured orange and 0.7 to  $\geq 1.0$  was red.

### **2.5.3 RNAstructure program prediction**

Improved RNA structure models generated using the SHAPE values were generated using the RNAstructure software (Reuter and Mathews, 2010). The SHAPE reactivities were used as pseudo-free energy change terms, this forces the resultant RNA model to be constrained by the inputted NMIA values. The resultant model was used to generate a Vienna dot bracket file, this file was viewed using VARNA and the NMIA reactivities were overlaid as before.

## Chapter 3: Effect of opal stop codon read-through on CHIKV replication

### 3.1 Introduction

Stop-codon read-through is a common strategy utilised by many different viruses to modulate the expression of certain viral proteins during infection. Stop codons are commonly found upstream of regions which encode the viral polymerase, specific read-through modulates the levels of expressed polymerases and consequently replication (Csibra et al., 2014). In alphavirus genomes an opal stop codon is located prior to the nsP3-nsP4 cleavage site, the presence of the stop codon results in the expression of nsP1,2,3 precursor as the majority polyprotein product but specific read-through of the stop codon occurs 10% of the time to yield the full nsP1,2,3,4 polyprotein (Li and Rice, 1993). Expression of the full-length polyprotein produces the RdRp.

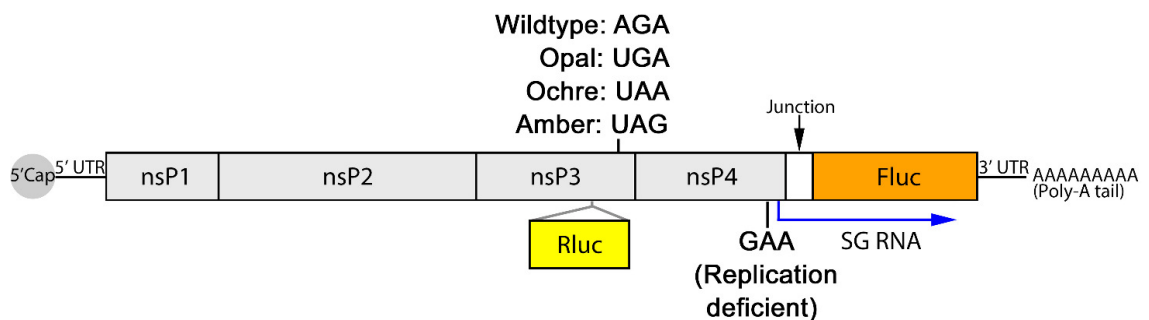
Most strains of CHIKV have an opal stop codon at this position but some possess a sense codon which commonly encodes for arginine, the presence of the arginine results in the expression of only the complete nsP1234 precursor (Götte et al., 2018). The role of the opal stop codon in CHIKV replication is currently unknown. To investigate the distribution of stop versus sense codon at this position the nucleotide sequence of multiple strains of CHIKV and the related alphaviruses; SINV, VEEV, ONNV and SFV were aligned against the strain of CHIKV studied in our lab. This panel of related alphaviruses were chosen due to their application in previous alignments of this region, SINV and VEEV predominately possess an opal stop codon at this position while the SFV and ONNV demonstrate the same fluctuation between the sense and stop codon at this position as seen for CHIKV (Firth et al., 2011). The lab strain of CHIKV was derived from the ICRES ECSA strain (Tsetsarkin et al., 2006) (**Figure 9**).





that the C-terminal amino acids present in longer nsP3 isoforms function as a degradation signal; removing 10 C-terminal amino acids from SFV nsP3 increases the half-life of the protein by 8 fold. In SFV these amino acids are Asp-Val-Leu-**Arg**-Leu-Gly-Arg-Ala-Gly-Ala and in the ICRES ECSA CHIKV strain they are Asp-Glu-Leu-**Arg**-Leu-Asp-Arg-Ala-Gly-Gly. The bold Arg is equivalent to the opal stop codon position seen in other alphaviruses (Varjak et al., 2010).

In this study stop codons were substituted into a Dluc ECSA CHIKV SGR which contained a Rluc fused to the nsP3 reading frame and a Fluc controlled by a SG promoter. This construct was used to investigate how the presence of stop codons could influence CHIKV replication. Site directed mutagenesis was used to change the wildtype arginine codon (AGA) at this position into sequences corresponding to the opal (UGA), ochre (UAA) and amber (UAG) stop codons. A replication deficient nsP4 (RdRp) mutant with a disrupted active site of Gly-Asp-Asp (GDD) to a Gly-Ala-Ala (GAA) substitution was used as a negative control for CHIKV replication (**Figure 10**).



**Figure 10: Stop codons were substituted into the ECSA derived Dluc CHIKV SGR construct.**

The SGR contained Rluc and Fluc reporters which were used to measure replication of the CHIKV SGR in different cell types.

The Dluc reporter system was chosen as both luciferases do not require post translational modifications to function and can be used in a combined system as they require different substrates to function. The Fluc reaction for photon emission occurs when the beetle luciferin is oxidised when incubated with ATP,  $Mg^{2+}$ ,  $O_2$  and coenzyme A whilst Rluc luminescence occurs with  $O_2$  and coelenterate luciferin. The signal for both luciferases are measured in tandem,

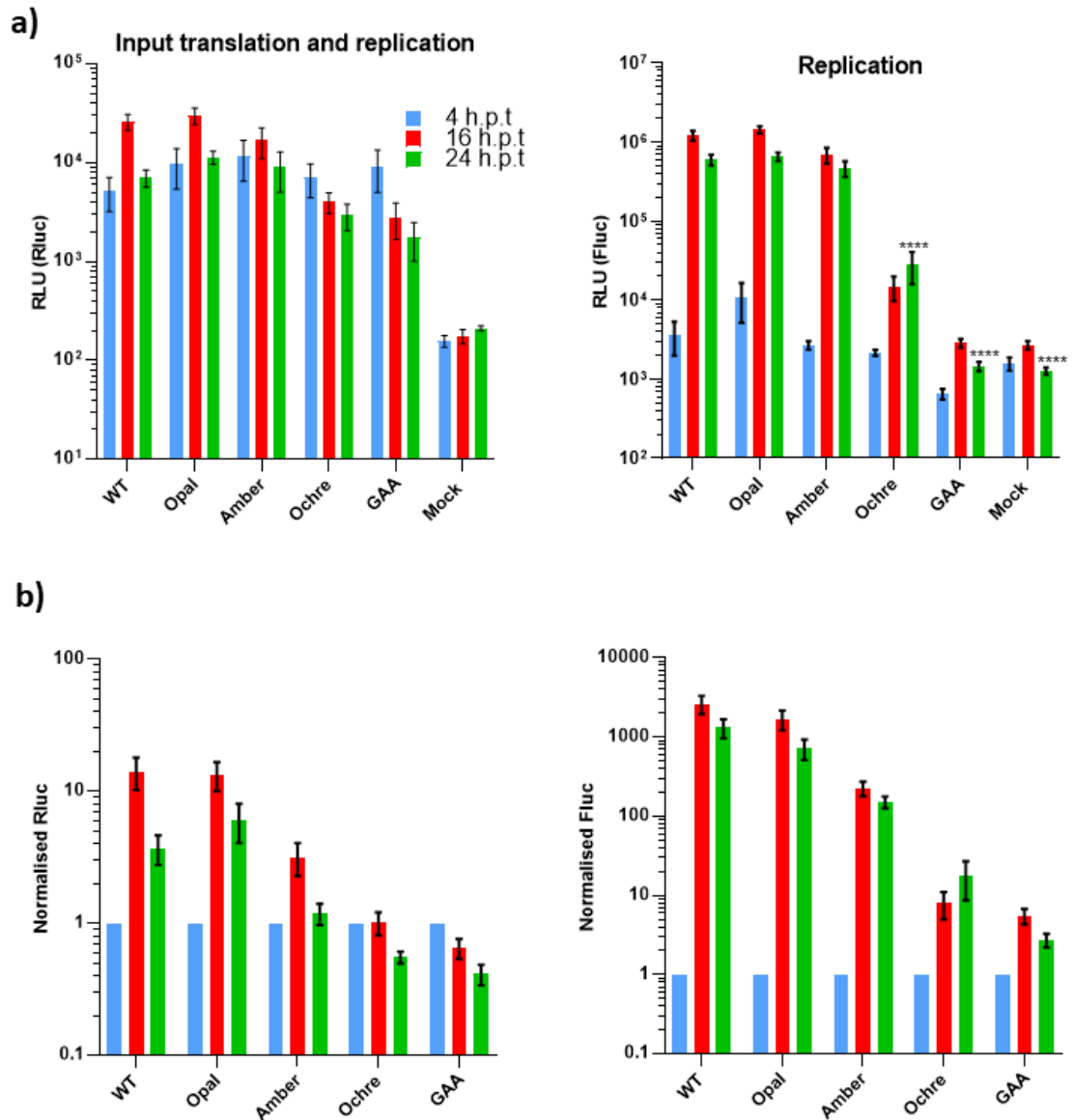
first reagent is added to induce Fluc activity and then a reagent which quenches Fluc signals and activates Rluc is added (Sherf et al., 1996). In the CHIKV Dluc SGR the amount of relative light units (RLU) for the Rluc or Fluc correspond to expression of the ORF1 or replication respectively. The expression of ORF2 is dependent on successful stop codon read-through; the stop codon is prior to the nsP4 encoding sequence and thus early termination would prevent the expression of the RdRp required for active genome replication complex.

CHIKV can replicate both in humans and *Aedes* mosquitoes, different phenotypes have been observed in the two hosts. Infected mosquitoes present a lifelong asymptomatic infection whilst infection in humans results in cytopathic infection which causes the fever associated with the virus (Rupp et al., 2015). A variety of mammalian cell types and a mosquito cell line were transfected with the Dluc SGR panel determine the effect which the stop codons would have on CHIKV replication and to investigate cell type specific phenotypes.

## 3.2 Results

### 3.2.1 The effect of stop codon read-through on CHIKV SGR replication in non-human mammalian cells

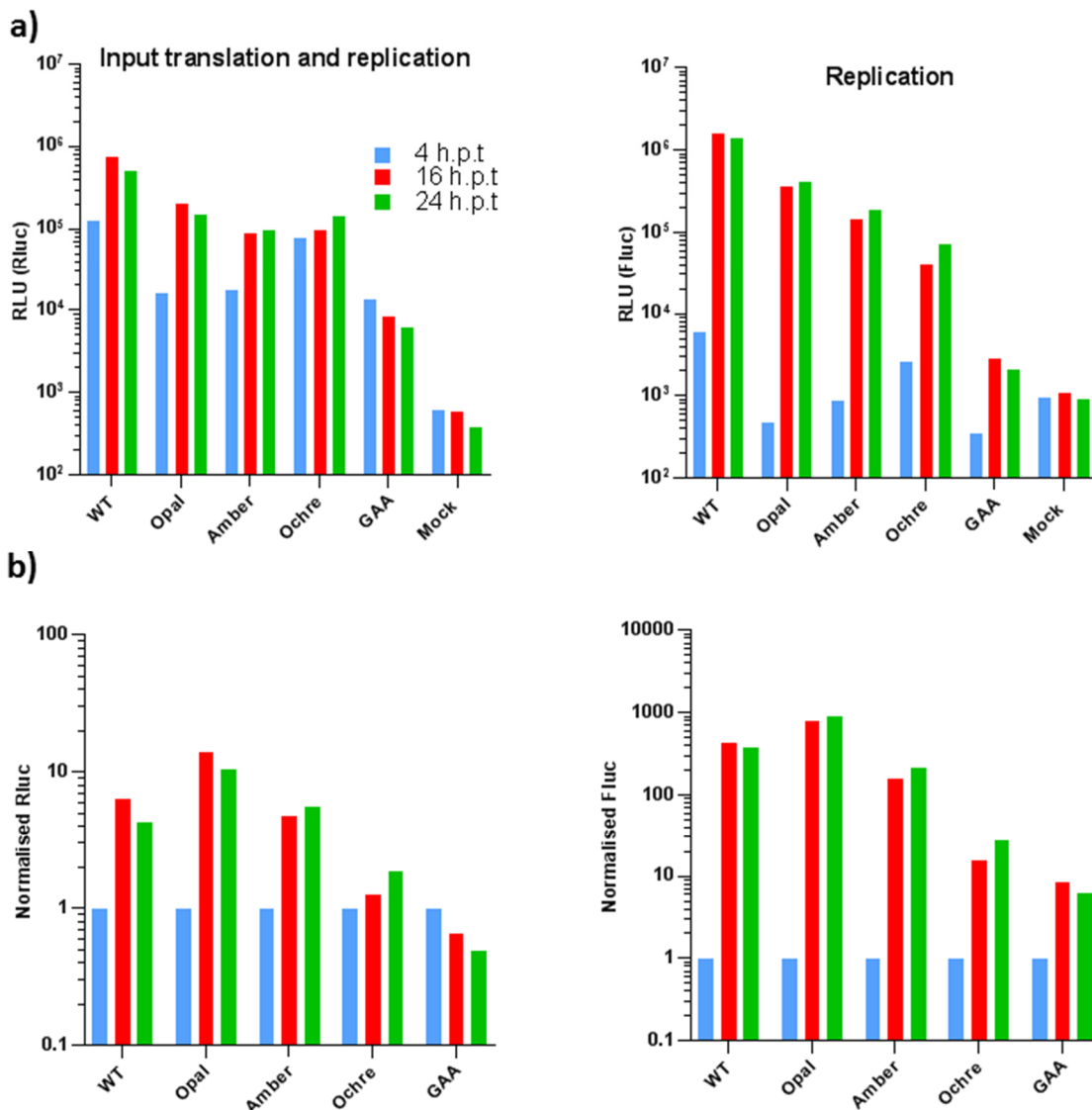
Mouse models have been used to investigate the pathology of CHIKV infection on a mammalian host due to similarities in comparison to human infection with the virus. Previous studies that have successfully demonstrated CHIKV infection in mice mimicked the expected joint damage and inflammation seen in human infection (Gardner et al., 2010). A mutant of CHIKV lacking the opal stop codon was shown to decrease CHIKV inflammation and ankle damage in mice, although replication of the virus was comparable to wildtype *in vivo* and *in vitro* (Jones et al., 2017). C2C12 (*Mus musculus*) mouse muscle cells were chosen due to the phenotypic relevance of muscle localised CHIKV infection (Yaffe and Saxel, 1977). These mouse cells were transfected with the Dluc CHIKV SGR panel, the replication deficient GAA mutant and cell only control to investigate how the mutations affected CHIKV replication (**Figure 11**).



**Figure 11: Effect of stop codon in nsP3 on CHIKV genome replication in murine mouse muscle cells (C2C12).**

(a) Lysates of the SGR transfected cells were harvested at 4, 16 and 24 hours post transfection (h.p.t). The Rluc and Fluc values can be seen on the left and right graphs. (b) All readings were normalised to the values at 4 hours for each individual construct to display the fold change overtime. Error bars represent S.E.M (One Way ANOVA versus WT at 24 h.p.t, \*\*\*\* =  $P \leq 0.0001$ ). N=3.

Baby hamster kidney (BHK) cells were also used to screen the replicon constructs as these cells are highly permissive for alphavirus replicon replication (Schlesinger and Dubensky, 1999). Infectious CHIKV has been shown to productively replicate in BHK cells, high titres of the virus were produced (Davis et al., 1971). These cells were transfected with the CHIKV SGR panel (**Figure 12**).



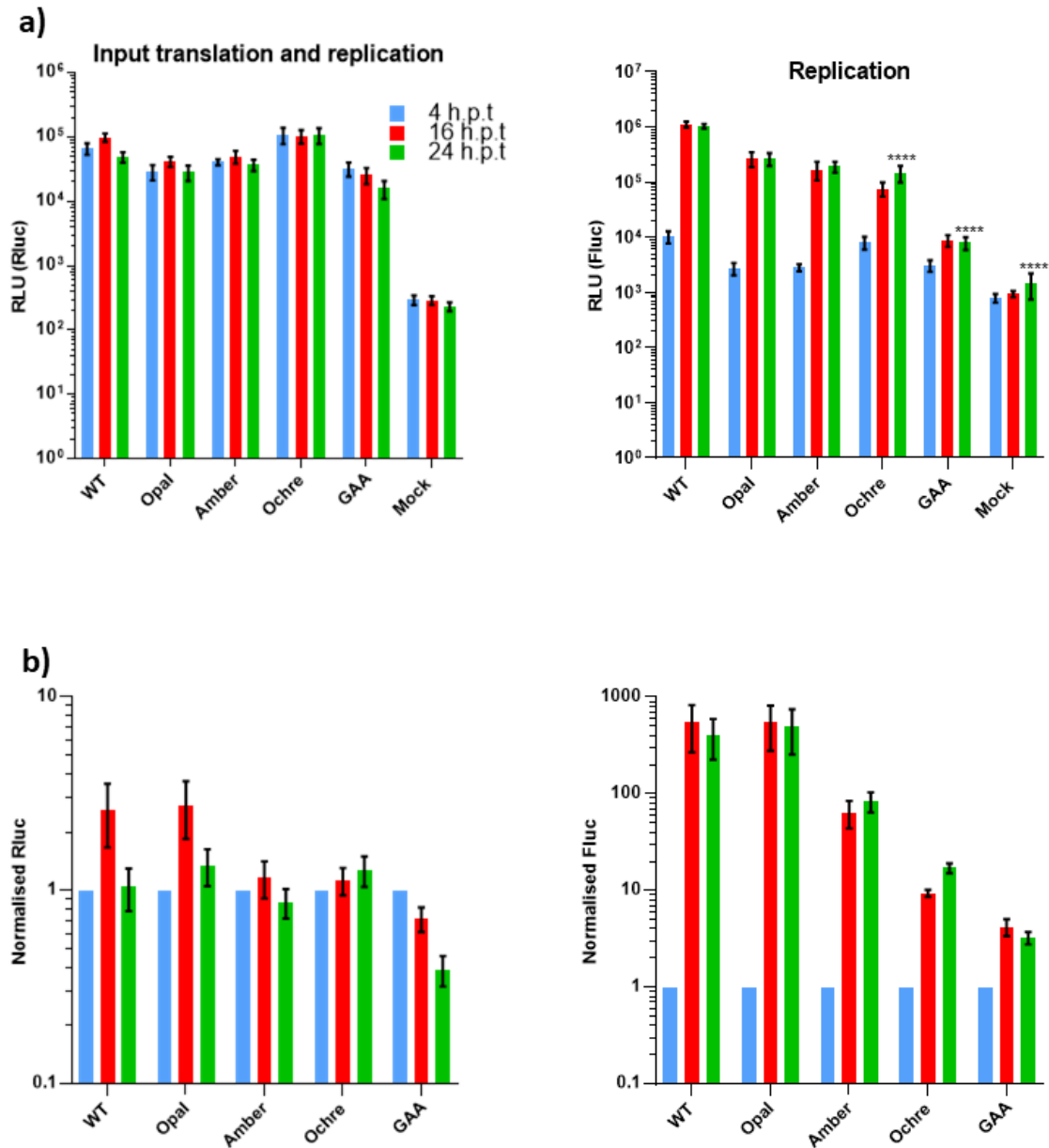
**Figure 12: Effect of stop codon in nsP3 on CHIKV genome replication in BHK cells.** (a) Lysates of the SGR transfected cells were harvested at 4, 16 and 24 h.p.t. The Rluc and Fluc values can be seen on the left and right graphs. (b) All readings were normalised to the values at 4 hours for each individual construct to display the fold change overtime. N=1.

The results of this analysis are presented in Figure 11 (C2C12 cells) and Figure 12 (BHK cells). Raw luciferase values are presented in (a), in addition, to control for experimental variability (e.g. a result of transfecting different RNAs) the data were normalised to the 4 h.p.t values (b). For the wildtype SGR, Fluc values increased by nearly 1000-fold from 4 – 24 h.p.t., indicative of the robust CHIKV genome replication and consistent with previous data (Roberts et al., 2017; Gao et al., 2019). Substitution of the wildtype Arg codon for the opal stop codon had no significant effect on replication in either cell type. Substitution for the amber stop codon resulted in a modest reduction in replication, but in contrast the presence of the ochre stop codon significantly reduced replication to levels close to that of the negative control (nsP4 GAA). These results indicate that efficient read-through of either an opal or amber stop codon near the C-terminus of nsP3 occurs during translation of the CHIKV ORF1, allowing expression of the nsP4 polymerase and subsequent genome replication. However, read-through of the ochre stop codon is inefficient, resulting in low replication. The stop codon mutant replicons were next screened against a panel of human cell types to investigate whether these phenotypes were cell type specific.

### **3.2.2 The effects of stop codon read-through in CHIKV replication in human cell types**

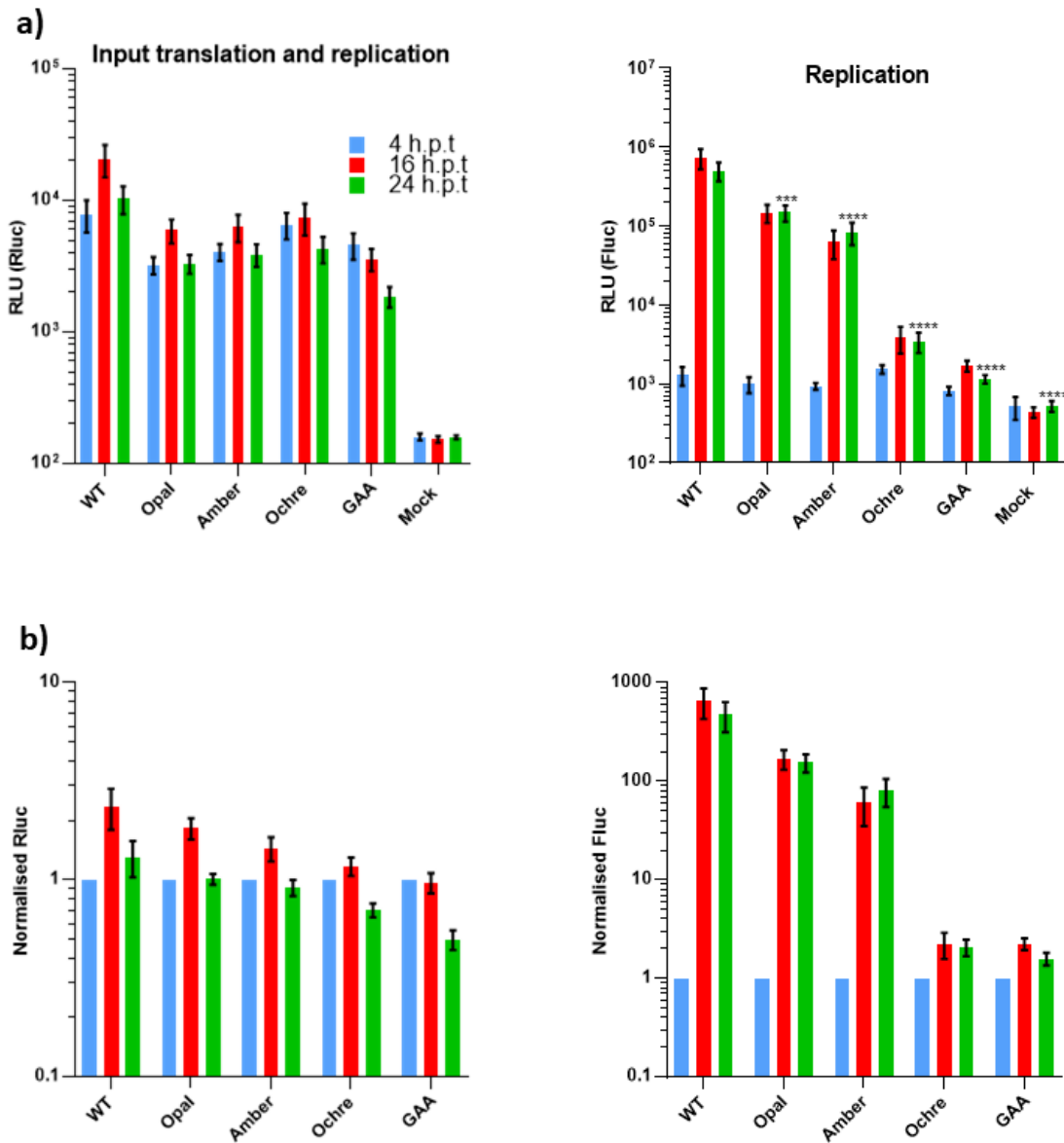
In humans CHIKV infection causes an acute infection which is characterised by severe muscle and joint pain, rash and fever. A subsequent long-term joint pain which can last months is described as the chronic stage of CHIKV infection (Goupil and Mores, 2016). The virus can infect a variety of cell types within the human host. The virus is spread to the humans when bitten by an infected mosquito, the virus replicates in host fibroblasts, epithelial and endothelial cells (Singh and Unni, 2011). CHIKV has been shown to disseminate in the host to infect organs such as the liver and brain (Jaffar-Bandjee et al., 2009). Human cell lines which originate from different organs were transfected using DLuc CHIKV SGRs, the chosen cell lines were the Huh7 (liver), SVG-A (glial) and RD cells (muscle). Previous work in the lab had identified these cell lines to be suitable models of CHIKV infection as they resulted in medium to high levels of CHIKV replication (Roberts et al., 2017).

Huh7 cells are human liver cells which have been used during Hepatitis C virus studies in the lab previously (Ross-Thriepland and Harris, 2014). CHIKV is detected in the liver at the acute and chronic phase, infected liver cells display higher levels of apoptosis mediated cell death (Broeckel et al., 2015). SVG-A cells are human fetal astrocyte cells (Schweighardt et al., 2001), these glial cells were chosen to assess how the stop codon mutants replicated in brain cells. Severe symptoms of CHIKV infection has been detected in children and elderly individuals, these infections are characterised by neurological complications including meningoencephalitis and seizures (Mehta et al., 2018). RD cells are a human muscle cell line (Hinson et al., 2013), these cells were chosen as muscle cells are a primary target for CHIKV replication (Singh and Unni, 2011). The Huh7 (**Figure 13**), SVG-A (**Figure 14**) and RD cells (**Figure 15**) were transfected with the D<sub>luc</sub> CHIKV SGRs to investigate whether cell type specific phenotypes were observed.



**Figure 13: Effect of stop codon in nsP3 on CHIKV genome replication in human liver cells (Huh7).**

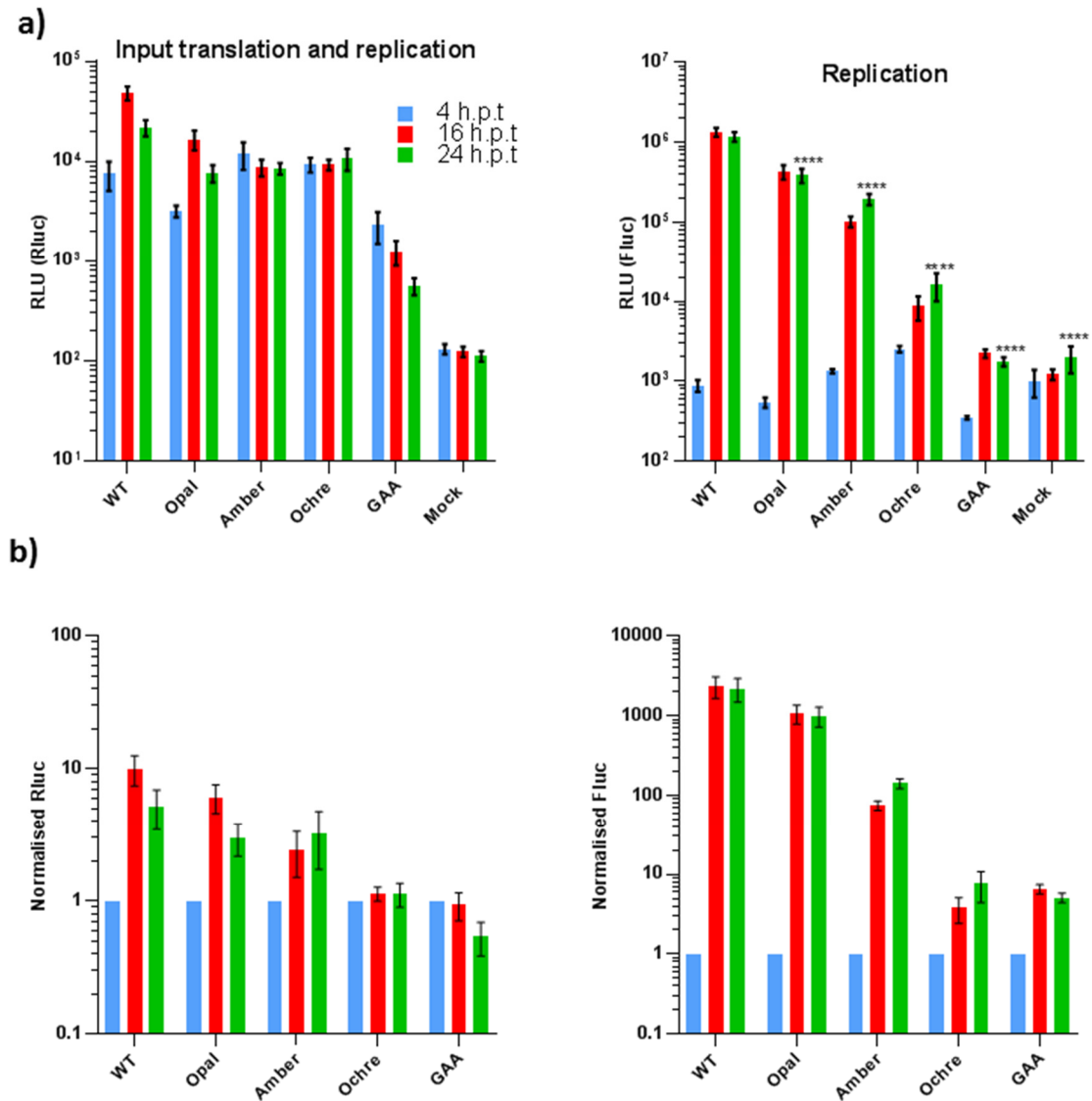
(a) Lysates of the SGR transfected cells were harvested at 4, 16 and 24 h.p.t. The Rluc and Fluc values can be seen on the left and right graphs. (b) All readings were normalised to the values at 4 hours for each individual construct to display the fold change overtime. Error bars represent S.E.M. (One Way ANOVA versus WT at 24 h.p.t, \*\*\*\* =  $P \leq 0.0001$ ). N=3.



**Figure 14: Effect of stop codon in nsP3 on CHIKV genome replication human glial cells (SVG-A).**

(a) Lysates of the SGR transfected cells were harvested at 4, 16 and 24 h.p.t. The Rluc and Fluc values can be seen on the left and right graphs. (b) All readings were normalised to the values at 4 hours for each individual construct to display the fold change overtime. Error bars represent S.E.M. (One Way ANOVA versus WT at 24 h.p.t, \*\*\* =  $P \leq 0.001$ , \*\*\*\* =  $P \leq 0.0001$ ). N=3.





**Figure 15: Effect of stop codon in nsP3 on CHIKV genome replication in human muscle cells (RD).**

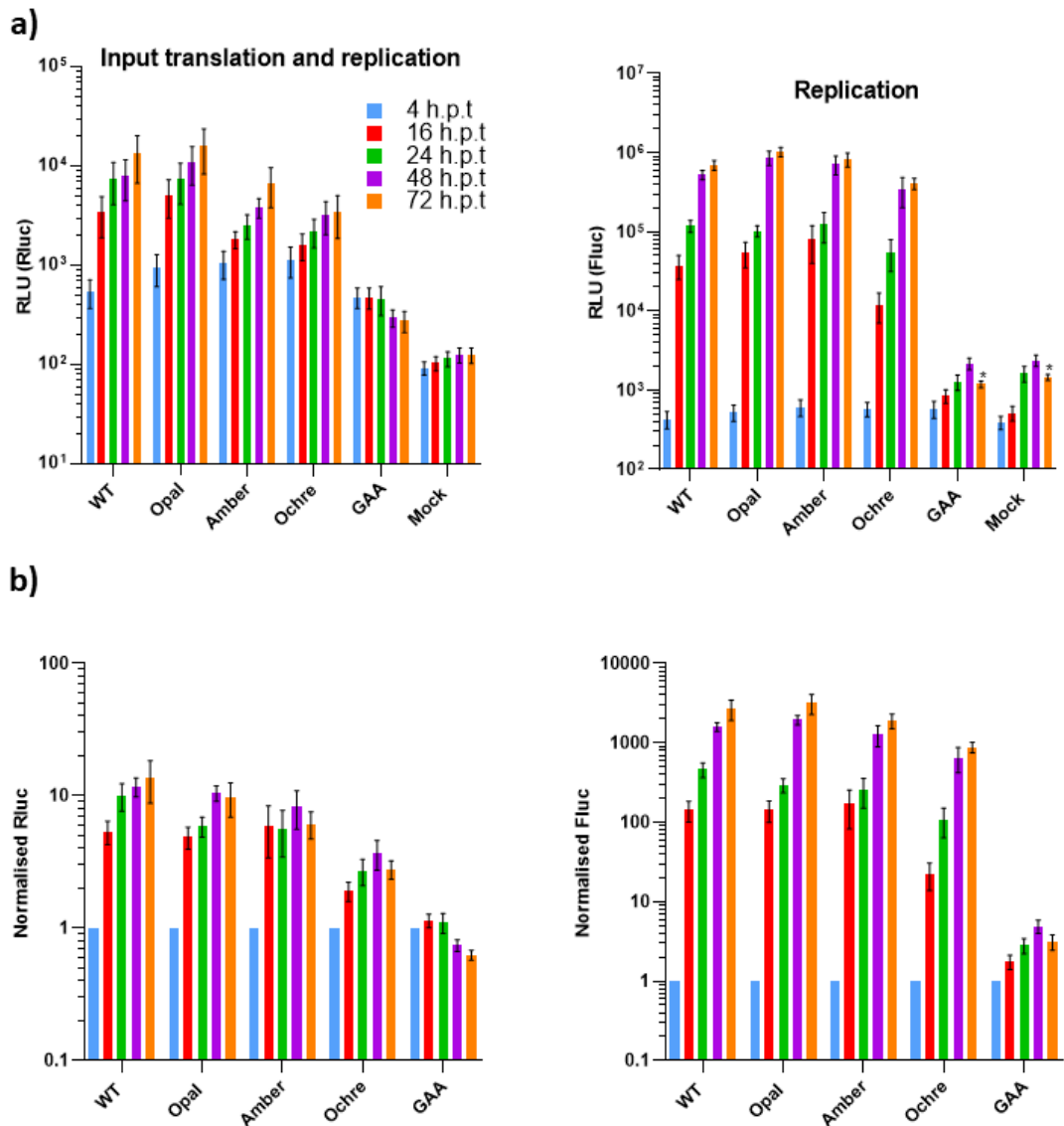
(a) Lysates of the SGR transfected cells were harvested at 4, 16 and 24 h.p.t. The Rluc and Fluc values can be seen on the left and right graphs. (b) All readings were normalised to the values at 4 hours for each individual construct to display the fold change overtime. Error bars represent S.E.M. (One Way ANOVA versus WT at 24 h.p.t, \*\*\*\* =  $P < 0.0001$ ). N=3.

Replication of the CHIKV SGR in Huh7 (**Figure 13**), SVG-A (**Figure 14**) and RD cells (**Figure 15**) are presented above. The raw luciferase values are presented in (a) and data normalised to 4 h.p.t values in (b). Fluc values of the wildtype SGR in all three human cell types displayed around a 1000-fold increase between 4-24 h.p.t, the opal stop codon had no significant effect on replication in Huh7 cells but displayed a small reduction in replication in SVG-A and RD cells. The Fluc values in all three human cell types for the amber stop codon SGR were 10-fold lower than the wildtype at 16 and 24 h.p.t, the ochre stop codon resulted in a 100-fold decrease. Inefficient read-through of the ochre stop codon but efficient read-through of the opal and amber stop codon was seen in all human cell types. The maintained phenotypes across the different cell types suggested that the stop codons did not have cell type specific phenotypes, instead the ochre stop codon had a general reduced replication in all mammalian cell types. It was of interest to investigate whether this phenotype was shared in mosquito cells. These stop codon mutants were screened in the C6/36 mosquito cell type to investigate if a phenotype could be seen in cells isolated from the *Ae. albopictus* vector.

### **3.2.3 The effects of stop codon read-through on CHIKV replication in C6/36 (mosquito) cells**

Alphaviruses produce a lifelong non-cytopathic infection in their mosquito vectors, CHIKV infects the *Ae. albopictus* and *Ae. aegypti* mosquitoes. Differences between alphavirus infection in vertebrate cells compared to invertebrate cells have been observed; for example macromolecular synthesis is shut off in the acute vertebrate infection but not in the persistent mosquito infection (Rupp et al., 2015). Stop codon read-through for CHIKV and related alphaviruses increases infectivity of the viruses in the invertebrate hosts. Cells derived from an *Ae. albopictus* mosquito (C6/36) were screened with the SGR panel. These cells lack a functional RNAi pathway (Brackney et al., 2010) and CHIKV infection of C6/36 cells results in mild cytopathic effect (CPE) and persistent infection (Li et al., 2013). These properties suggest that the C6/36 cell line can be used as a model for CHIKV infection in the mosquito vector. The stop

codon Dlu<sub>c</sub> CHIKV SGRs were transfected into the C6/36 cells to investigate the effect of stop codon read-through on CHIKV replication in mosquito cells. Timepoints post transfection were increased to 72 hours as lower CPE was expected in these cells (**Figure 16**).



**Figure 16: Effect of stop codon in nsP3 on CHIKV genome replication in *Ae. albopictus* cells (C6/36).**

(a) Lysates of the SGR transfected cells were harvested at 4, 16, 24, 48 and 72 h.p.t. The Rluc and Fluc values can be seen on the left and right graphs. (b) All readings were normalised to the values at 4 hours for each individual construct to display the fold change overtime. Error bars represent S.E.M. (One Way Anova versus WT at 72 h.p.t, \* =  $P \leq 0.05$ ). N=3.

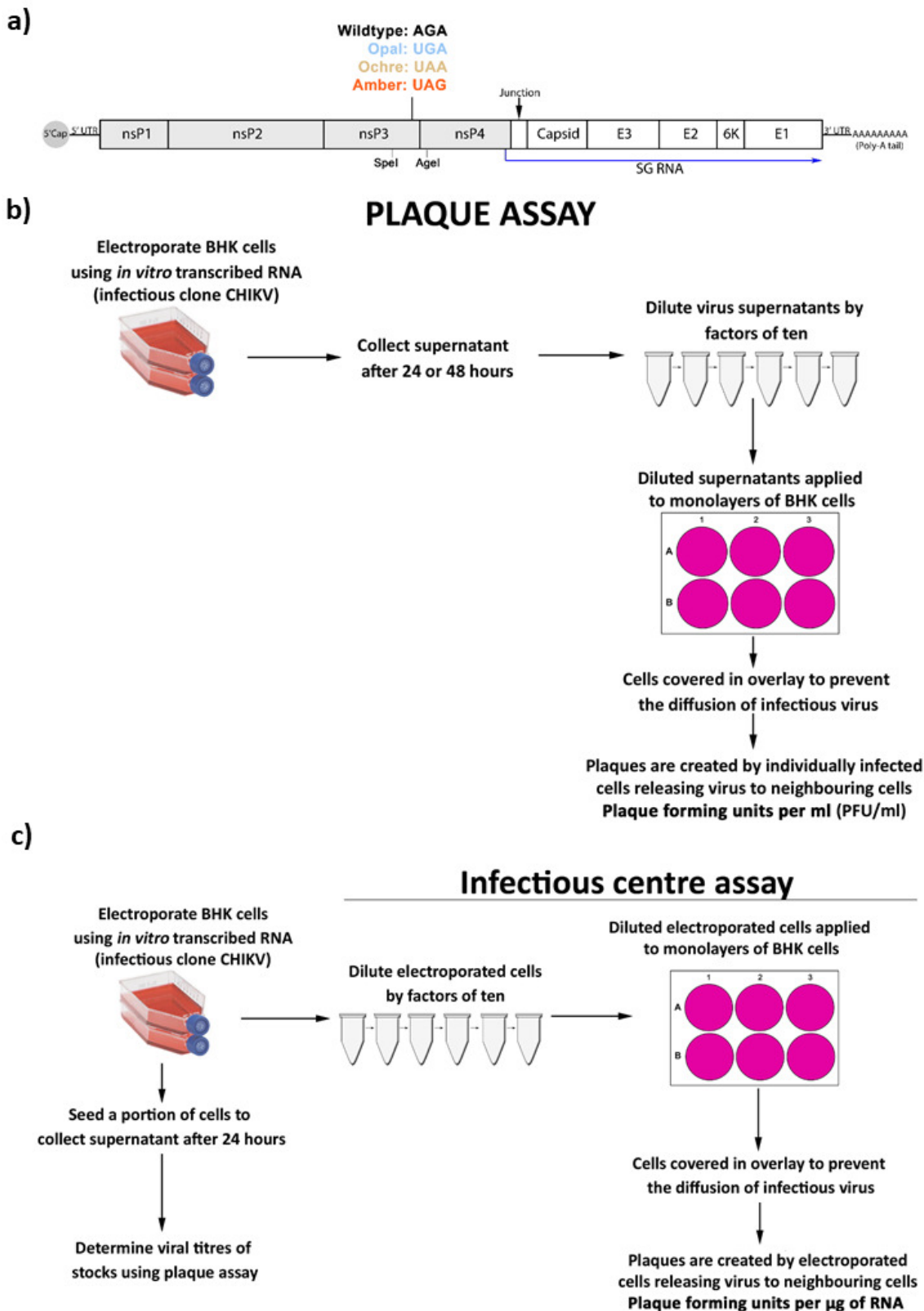
Replication of the CHIKV SGR in C6/36 cells can be seen in Figure 16. Raw luciferase values up to 72 h.p.t are shown in (a) and the graphs showing normalised to 4 h.p.t values in (b). The Rluc and Fluc values for the wildtype and stop codon SGR were shown to increase up to 48 h.p.t before plateauing at the final 72 h.p.t timepoint suggesting that the replicons were able replicate in C6/36 cells for extended lengths of time. A 1000-fold increase in Fluc values between 4-48 h.p.t and 4-72 h.pt was seen for the wildtype, opal, amber and ochre SGR; these results suggest that stop codon read-through of all three stop codons was efficient in mosquito cells. The prolonged replication of the SGR observed in the mosquito cells can be attributed to the non-cytotoxic nature of CHIKV in insect cells which would permit continuous replication of the replicons, in contrast replication in the mammalian cell types result in cell death.

#### **3.2.4 The effect of stop codon read-through on CHIKV infectivity in BHK cells**

The effect on stop codon read-through on the context of complete viral lifecycle was investigated, these stop codon mutations were introduced into the ICRES infectious clone of CHIKV which was derived from the ECSA CHIKV strain used to create the SGR (Tsetsarkin et al., 2006). To create these constructs a GeneArt fragment (Thermo Fischer) of a segment of nsP3 containing an introduced novel SpeI site which matched the corresponding restriction site in the SGR was produced, this fragment was subcloned into the ICRES CHIKV backbone. Fragments of nsP3 containing each stop codon were generated by digesting the stop codon CHIKV SGR with the SpeI and AgeI restriction enzymes, these inserts were subcloned into the ICRES CHIKV (**Figure 17a**). This system was used to investigate whether the phenotypes observed in the SGR would be consistent in the full CHIKV lifecycle.

Plaque assays were used to determine the titres of virus stocks produced, virus produced from transfected cells was serially diluted before application to seeded monolayers of cells. After adhesion, these plates were incubated for 37°C for 1 hour and covered using a methyl-cellulose (MC) overlay. Cells were then incubated for 48 hours. Plaques observed after fixing and staining corresponded

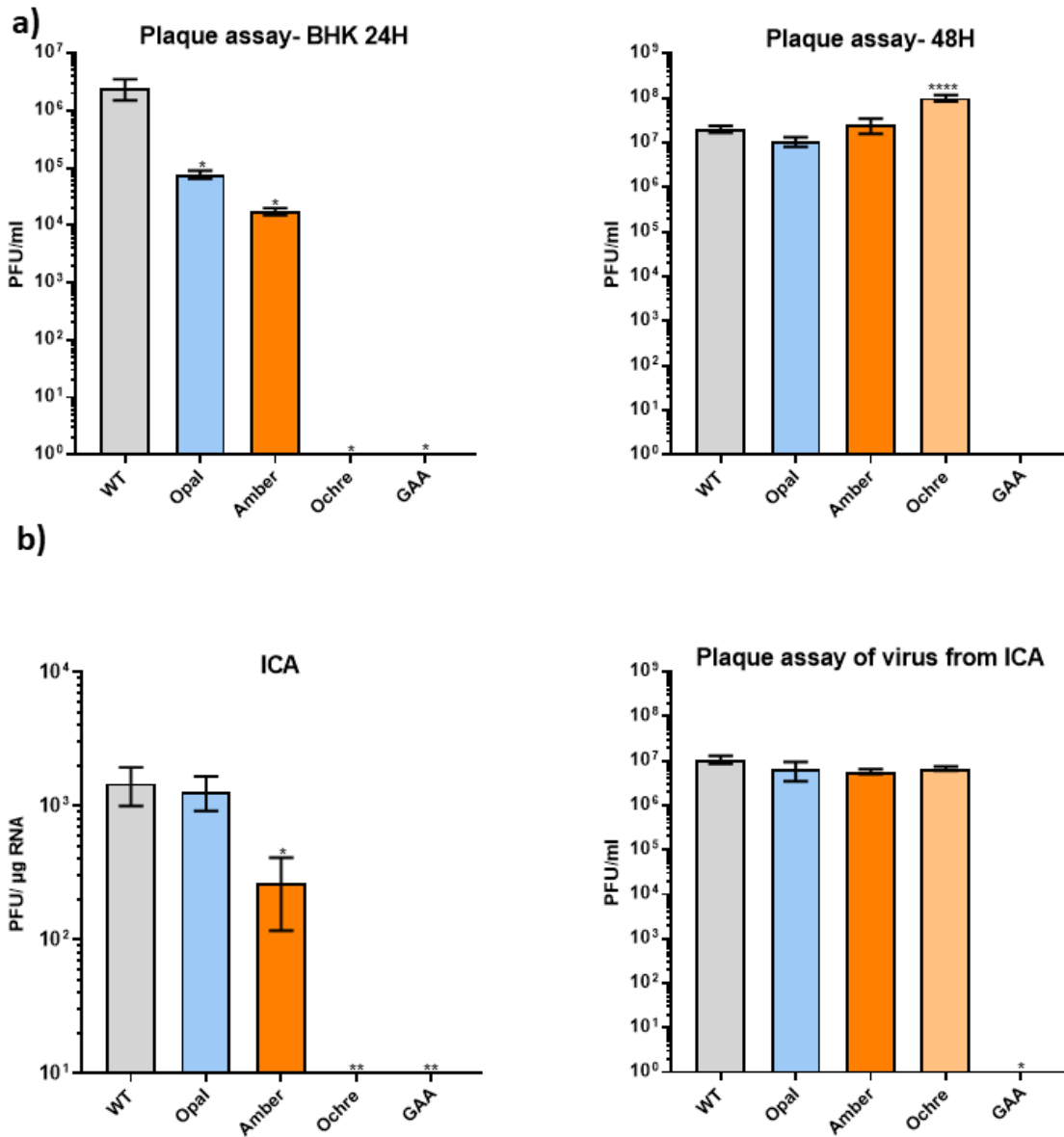
with a single virus particle spreading in a localised cluster of cells. A diagram detailing the process can be seen below (**Figure 17b**). Infectious centre assays (ICA) were performed using the stop codon mutant infectious clones to see if the mutations had an effect of the production of infectious CHIKV particles. An ICA determines the infectivity of the transfected infectious clone RNA, low RNA infectivity combined with high viral titres of stocks produced after transfection suggests a possible reversion. Briefly, the ICA involves electroporating cells with the infectious clone RNA, these transfected cells are then serially diluted and subsequently applied onto pre-seeded cells, incubated for 1 hour at 37°C before a MC overlay is applied to the cells. Plaques observed after fixing and staining correspond to transfected cells spreading infectious virus to neighbouring cells (Mutso et al., 2018) (**Figure 17c**).



**Figure 17: Stop codon mutants of infectious clone CHIKV (ICRES) and the flow diagrams depicting the plaque assay and ICA protocols.**

(a) All three stop codon mutations were introduced into the ICRES infectious clone of CHIKV. A GAA replication deficient mutant was used as a negative replication control. (b) A diagram depicting the plaque assay methodology (c) A diagram depicting the process of the ICA experiment used to investigate RNA infectivity of the ICRES CHIKV RNA.

BHK cells were transfected with the panel of ICRES CHIKV infectious clones, virus stocks were isolated 24 and 48 hours post electroporation and the virus titres were determined using plaque assays. A GAA replication deficient mutant of infectious CHIKV was used as a negative control (**Figure 18a**). The ICA for these infectious clones and the titres of the viral stock produced during this assay can be seen below (**Figure 18b**).



**Figure 18: Effect of stop codon read-through on ICRES CHIKV virus production and RNA infectivity in BHK cells.**

(a) Viral titres for virus stocks harvested 24H (N=2) or 48H (N=3) post transfection of BHK cells were determined using plaque assays. (b) The RNA infectivity of the ICRES CHIKV panel was determined using ICA on the left (N=3), the viral titre of virus stocks harvested from ICA experiments (N=2) is also shown. Error bars represent S.E.M. (One Way ANOVA versus WT, \* =  $P \leq 0.05$ , \*\* =  $P \leq 0.01$ , \*\*\*\* =  $P \leq 0.0001$ ).



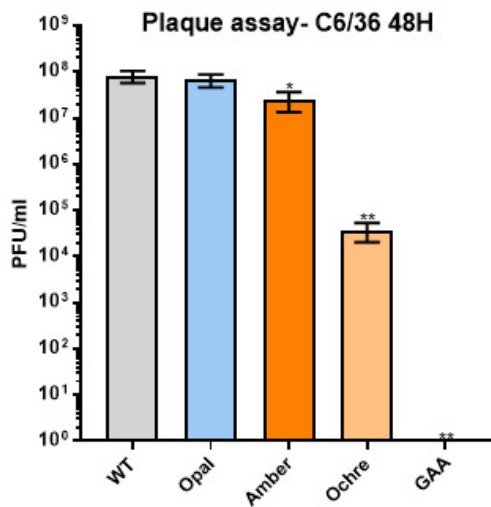
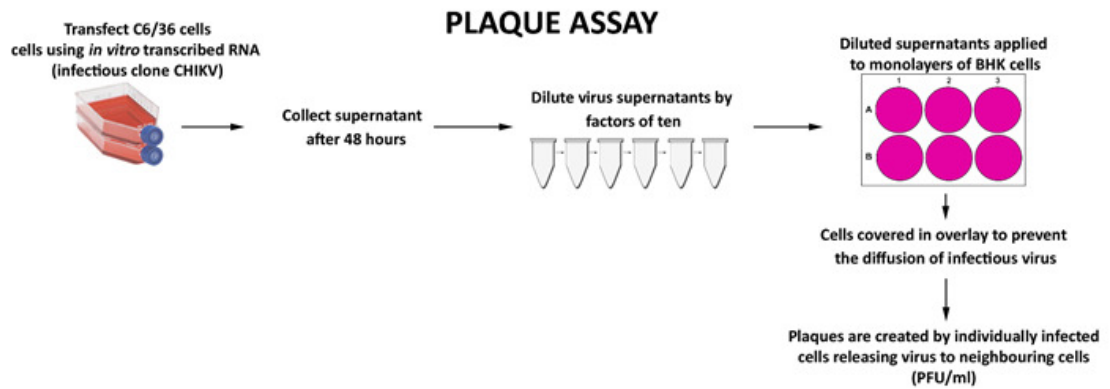
Titres of the virus stocks harvested 24 and 48 hours post transfection were determined using plaque assays (**Figure 18a**). The RNA infectivity of these infectious clones was determined using ICA, the titre of the virus produced by these transfected cells was determined using plaque assay (**Figure 18b**). The opal stop codon resulted in a 10-fold reduction in produced virus after 24 h.p.t but generated wildtype levels of virus after 48 hours, a similar phenotype was seen for the amber stop codon. The ochre stop codon resulted in no observable plaques when the virus was harvested 24 hours post transfection but resulted in 10-fold higher titres than wildtype when harvested at 48 hours. The GAA ICRES CHIKV resulted in no detected titre in either harvest, as expected the replication deficient mutant could not produce virus. The RNA infectivity of opal stop codon CHIKV was similar to wildtype but a slight reduction was seen for the amber stop codon CHIKV but the presence of the ochre stop codon resulted in no observable RNA infectivity. Virus stocks harvested during the ICA revealed that all three stop codon mutants could replicate at wildtype levels, for the ochre stop codon the low RNA infectivity but high viral titre suggests a possible reversion.

The presence of the opal and amber stop codons resulted in a slightly reduced viral titre at 24 h.p.t but wildtype levels of replication can be seen after 48 h.p.t and from the ICA isolated virus samples. These results suggest that read-through of ochre stop codon is inefficient in mammalian cells, however the mutant can produce wildtype levels of virus over time which may be a result of reversion. C6/36 cells were transfected with the same infectious clone RNA to test whether the viruses produced would also have the same phenotypes in the mosquito cells.

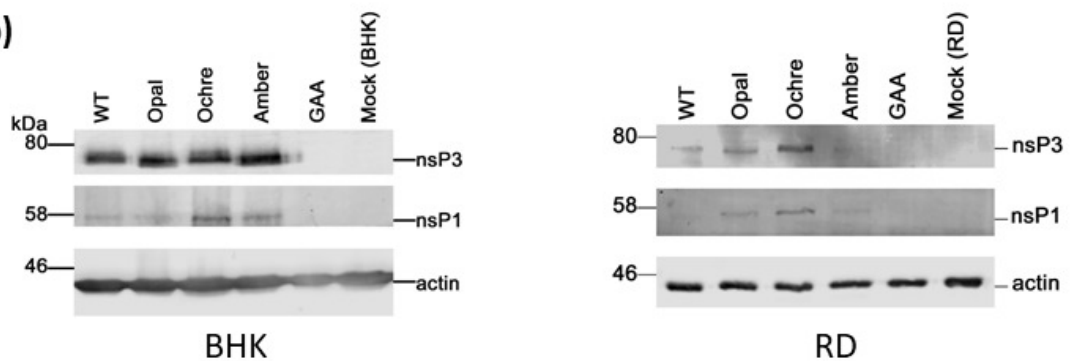
### 3.2.5 The effect of stop codon read-through on CHIKV infectivity in C6/36 cells

C6/36 cells were transfected with the ICRES CHIKV RNA to investigate whether the stop codon mutants would influence virus replication in mosquito cells. The mosquito cells were transfected with the infectious clone RNA, viral stocks were harvested from the cells after 48 hours and these virus samples were then plaque assayed using BHK cells (**Figure 19a**). Virus produced by transfected C6/36 cells were used to infect BHK and RD cells at a multiplicity of infection (MOI) of 0.01 to test if the mutant virus stocks displayed the same ability to infect cells as wildtype. Protein samples were lysed for western blot using anti-nsP3, anti-nsP1 and anti-actin (loading control) antibodies to compare the levels of nsP expression (**Figure 19b**).

a)



b)



**Figure 19: Effect of stop codon read-through on CHIKV production in C6/36 cells and nsP1 and nsP3 expression in BHK and RD cells.**

(a) Virus stocks produced using C6/36 cells (48 h.p.t with ICRES CHIKV RNA) were plaque assayed using BHK cells to determine viral titre (N=5). (b) Western blots of mammalian BHK and RD cells infected samples using antibodies specific to nsP3, nsP1 and actin. Error bars represent S.E.M. N=1. (One Way ANOVA versus WT, \* =  $P \leq 0.05$ , \*\* =  $P \leq 0.01$ ).

The titres of the infectious clones of CHIKV isolated from transfected C6/36 cells at 48 h.p.t was determined by plaque assay on BHK cells (**Figure 19a**). The expression of nsP1 and nsP3 in infected BHK and RD cells was investigated using western blots with antibodies specific to the viral proteins (**Figure 19b**). The presence of the opal stop codon had no effect on viral titre but a slight decrease was seen for the amber stop codon containing infectious clone. Presence of the ochre stop codon resulted in a 1000-fold decrease in viral titre compared to wildtype. The amount of nsP3 expressed in infected BHK and RD cells appeared to be equivalent for all stop codon mutants but a slight increase in amount for the ochre infected cells was seen in the RD cells. NsP1 was present in all stop codon mutants but lower levels of the viral protein were seen in the WT, opal and amber samples. Read-through of all three stop codons in C6/36 was observed, the ochre stop codon appeared to be tolerated in C6/36 cells albeit a reduction in comparison to wildtype was seen.

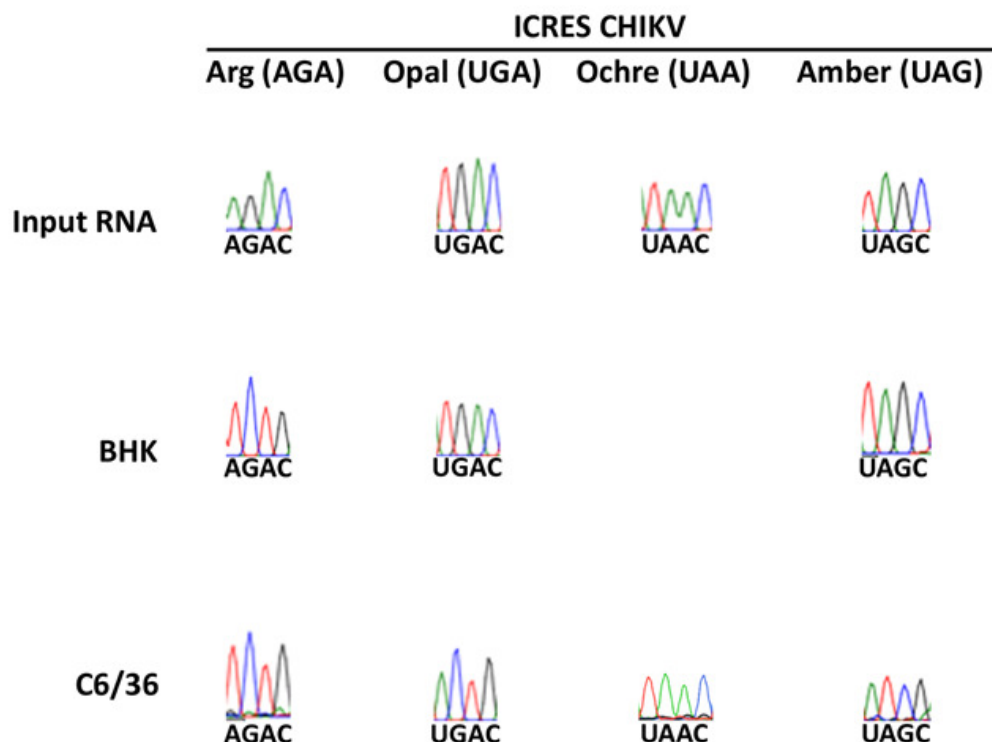
### **3.2.6 Sequencing of stop codon ICRES CHIKV produced in BHK and C6/36 cells**

Both BHK and C6/36 cells were transfected with RNA of the infectious clone CHIKV panel, intracellular RNA was extracted using Trizol at 24 and 48 h.p.t respectively. The RNA was used for reverse transcription PCR (RT-PCR). Random hexamer primers were used to generate the pool of cDNA and subsequent PCR using primers specific for CHIKV genome was done to amplify the region containing the stop codon position. The PCR products were sequenced using the same specific CHIKV primers (**Figure 20a**). The sequences found for the extracted BHK, C6/36 and input RNA can be seen below (**Figure 20b**).

a)



b)



**Figure 20: Stop codon mutations were maintained in the BHK and C6/36 extracted RNA samples.**

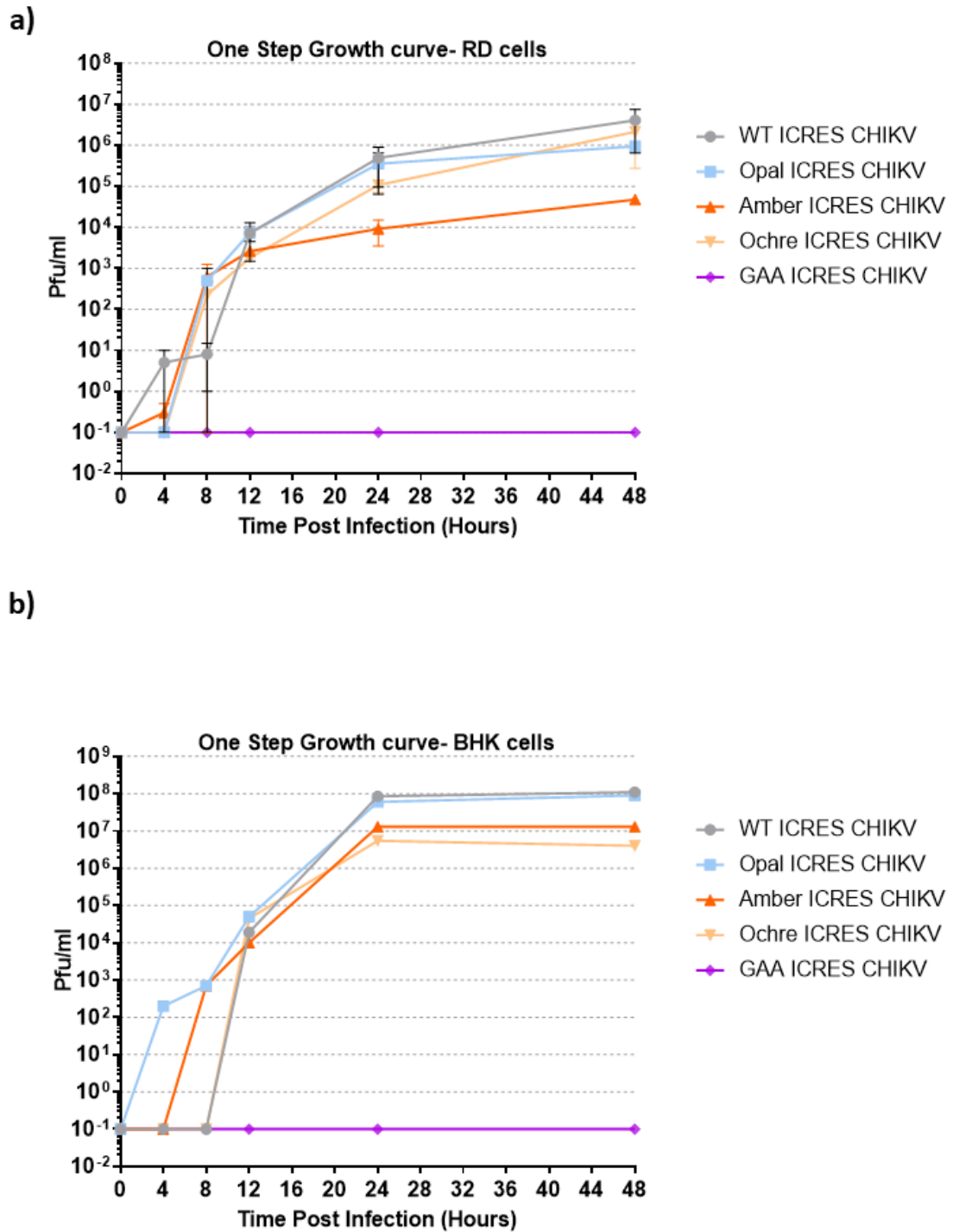
(a) A diagram depicting the procedure utilised to harvest intracellular RNA and the subsequent RT-PCR and sequencing protocol. (b) Sequence data for the input RNA, BHK and C6/36 extracted RNA post transfection with ICRES CHIKV RNA using the same primers which specifically targeted the CHIKV genome for the PCR step. RNA was extracted at 24 h.p.t (N=1).

The protocol described in Figure 20a was used to investigate if the stop codon mutations incorporated into the ICRES CHIKV were maintained post transfection into BHK and C6/36 cells. The sequences of the input RNA, BHK and C6/36 extracted RNA can be seen in Figure 20b. Sequencing the input RNA confirms

that all three stop codon mutants do contain the desired stop codon mutations, in BHK cells the opal and amber stop codons were seen to persist, but no sequence data could be obtained for the ochre stop codon sample due to low levels of replication (**Figure 18**). All three stop codons were shown to persist in the C6/36 cell samples. These results confirm that the opal and amber stop codons did not mutate when in both BHK and C6/36 cells. The ochre was also shown to not revert when in C6/36 cells.

### **3.2.7 One step growth curve of RD and BHK cells infected with stop codon infectious clone CHIKV**

Virus stocks produced in the C6/36 cells were used to infect BHK and RD cell types to investigate the growth of the stop codon mutant CHIKV over time. Virus produced using the mosquito cell line was chosen to maintain the phenotypic relevance of mosquito to mammalian host infection. The RD and BHK cells were infected at a MOI of 0.01. Resultant virus stocks were harvested at different timepoints to investigate whether the stop codon mutations would affect the production of virus versus the wildtype virus. These virus samples were plaque assayed on BHK cells (**Figure 21**).



**Figure 21: The effect of stop codons on CHIKV growth in RD and BHK cells overtime.**

RD (N=2) and BHK (N=1) cells were infected using the stop codon infectious clone CHIKV panel at a MOI of 0.01, samples were taken at timepoints up to 48 hours post infection. The titres of these samples were measured using plaque assays on BHK cells.

The growth of the stop codon ICRES CHIKV in RD cells overtime is shown in Figure 21a, the growth in BHK cells is in Figure 21b. The presence of the opal stop codon did not affect the growth of ICRES CHIKV as wildtype levels of PFU/ml were seen at all timepoints for both cell types. The amber stop codon resulted in a 10-fold decrease in viral titre for the samples harvested at 24 and 48 hours post infection. Interestingly the ochre mutant had wildtype levels of growth by 48 hours in RD cells but resulted in a 10-fold lower titre in BHK cells at the equivalent timepoint. All three stop codon ICRES CHIKV mutants replicated in both cell types but the rate of growth for each was different; presence of the opal stop codon was best tolerated of the three. The mechanism of opal stop codon read-through in alphaviruses is not fully understood but it is possible that the signals which promote opal stop codon read-through may not be fully efficient when used to read-through ochre and amber stop codons. Inefficient read-through of the ochre and amber stop codons would produce less nsP4, this would reduce the amount of RdRp produced. This reduction may contribute to the decrease in viral titres seen in the amber and more dramatically in the ochre stop codon mutant.

### 3.3 Discussion

The presence of the opal stop codon was shown to cause either no effect or a slight decrease to replication of CHIKV in the mammalian cells in both the SGR (**Figure 11 - Figure 15**) and in the infectious clone of CHIKV (**Figure 18, Figure 21**). The amber stop codon mutants had a similar phenotype to the opal stop codon but the ochre stop codon resulted in a severe decrease in replication. The viral titre for the ochre stop codon infectious clone CHIKV was shown to recover by 48 h.p.t in BHK cells (**Figure 18**) but unfortunately the sequencing for this mutant at 24 h.p.t was unsuccessful (**Figure 20**). Further sequencing experiments where RNA is extracted at 48 h.p.t for sequencing would be of interest as this timepoint corresponds to the increase in ochre ICRES CHIKV replication in mammalian cells, in addition more repeats of the BHK CHIKV SGR transfection would be desirable. The results for the infectious clone ICRES CHIKV transfection into mammalian cells and the C6/36 cells would be more comparable if both were harvested at the same timepoint (24 h.p.t), additionally it would be ideal to transfect the mammalian BHK cells using the same method as the C6/36 cell transfection to determine whether lipofectamine 2000 transfection influences the phenotypes seen for the stop codon mutants. The 4,



16 and 24 h.p.t timepoints were chosen for the SGR assay in mammalian cells as previous experiments I performed this assay demonstrated that the 12 and 16 h.p.t timepoints were equivalent and a drop off in luciferase signal was seen at 48 h.p.t. This decrease in signal most likely corresponds to the replicon causing cell death by this later timepoint. A small increase in RLuc and FLuc signal for the GAA and mock samples from 4 to 24/ 72 h.pt was seen in all CHIKV SGR assays, this is most likely due to the number of cells increasing over time and therefore represents a background reaction to the luciferase reagent. The SGR results correspond with the observations seen when the opal stop codon of the SINV was replaced with the other two stop codons; the production of the nsP34 polyprotein and the viral growth rate for the both the amber and ochre two stop codons were reduced, the ochre mutation caused the largest decrease in viral growth rate of the two (Li and Rice, 1989). The exact mechanism behind the reduced replication for the amber and ochre stop codon in comparison to the opal is currently unknown however the different stop codons may influence the read-through efficiencies for the alphavirus stop codon recoding signals.

All three stop codon SGRs replicated to wildtype levels in the C6/36 mosquito cell lines (**Figure 16**), the infectious clone mutants were also shown to replicate in the C6/36 cells (**Figure 19a**). All three stop codon mutations were present in the infectious clone RNA extracted from transfected C6/36 cells (**Figure 20**). Previous studies have implicated that the presence of the opal stop codon in alphaviruses enhances infection in the mosquito vector but not than the mammalian host. Increased ONNV infection of *A. gambiae* mosquito was attributed to the effect of the opal stop codon viral replication (Myles et al., 2006). In another study passaging an ONNV strain containing the stop codon in the mammalian Vero African Green Monkey cell line caused the stop codon to switch to an arginine at this position (Lanciotti et al., 1998), this could suggest the opal stop codon is selectively maintained in the mosquito vector but not the mammalian host. How the presence of either the arginine or opal stop codon confers host specific advantages is currently unknown and will require further research to characterise.

All three of the stop codon ICRES CHIKV mutants were able to infect both RD and BHK cells as shown by the western blots and one step growth curves (**Figure 19b and Figure 21**) however the differences in virus growth and levels of non-

structural protein expression suggest that the different stop codons do influence the viral fitness in these cells. More repeats of these experiments would be preferable as this validate the phenotype stated above. The inverse mutation of a La Réunion CHIKV strain possessing an opal stop codon into an arginine codon was shown to decrease viral titres produced in C6/36 cells ( $10^6$  vs  $10^4$ ) (Mounce et al., 2017), interestingly this phenotype was not seen in the infectious clone used in this study as the arginine ICRES CHIKV replicated to the same level as the opal stop codon ICRES CHIKV (**Figure 19a**). SINV infection of U4.4 (*Ae. albopictus*) cells resulted in prolonged detection of nsP123 precursor 5 days post infection which contributed to reduced positive sense RNA synthesis, this was hypothesised to promote persistent infection in the mosquito cells (Mudiganti et al., 2010). The stop codon mutant ECSA CHIKV replicons showed enhanced and prolonged replication up to 72 h.p.t into C6/36 cells (**Figure 16**). The presence of the three stop codons in the SGR likely increased the production of nsP123 and reduced nsP4 expression, this may have promoted CHIKV replication in mosquito cells by delaying the virus lifecycle. Western blots of these infected samples using an antibody targeting CHIKV nsP4 would be useful, this would provide a measure of the different levels of read-through for the different stop codons. Different efficiencies of stop codon read-through would influence the levels of nsP4 in the different cell types. Unfortunately, I did not have access to an nsP4 antibody to test this hypothesis. Nsp4 contains an N-end degradation signal and is therefore known to display a high rate of degradation, it was hypothesised that the unstable nature of nsP4 was a mechanism to regulate the levels of the RdRp during infection (De Groot et al., 1991). An antibody specific to CHIKV nsP4 has successfully been generated (Kumar et al., 2015), obtaining a similar antibody would be useful for investigating the effect of stop codon read-through on nsP4 expression in the different cell types.

Specific stop codon read-through can be promoted by signals upstream and downstream of the read-through site. It has been shown that the immediate cytidine residues downstream of the SINV opal stop codon promotes read-through by 10 fold (Li and Rice, 1993), this residue appears to be strongly conserved amongst the aligned CHIKV strains and other alphaviruses shown in this study (**Figure 9**). Furthermore a highly conserved downstream RNA structure was shown to also be important for both the SINV and VEEV stop-codon read-

through again enhancing read-through by up to 10 fold when present (Firth et al., 2011). The exact mechanism by which these signals promote read-through is currently unknown.

## Chapter 4: The role of RNA structure on opal stop codon read-through in the context of CHIKV replication

### 4.1 Introduction

As described in the introduction three types of stop codon read-through signals have been identified. Type 1 signals are plant virus exclusive hexa-nucleotide motifs immediately downstream of the stop codon. An example of the type 1 read-through of an amber stop codon is in the TMV which controls the expression for the RNA replicase, read-through is promoted by the CAR-YYA motif immediately downstream of the stop codon (where R signifies a purine base and Y a pyrimidine base) (Skuzeski et al., 1991). Type 2 signals are formed by the local nucleotide context following the recoded stop codon, in some cases only one of these nucleotides is required for effective read-through. Type 2 signals have been identified in plant viruses and RNA viruses, the tobacco rattle virus possess a CGG codon immediately following the opal stop codon, in this case the 2<sup>nd</sup> and 3<sup>rd</sup> nucleotides of the codon synergise to promote read-through (Beier and Grimm, 2001). Type 3 signals involve a downstream 3' RNA structure which promotes stop codon read-through, for example read-through of the stop codon in MuLV is promoted by an RNA pseudoknot structure which is located 8 nucleotides downstream of the amber stop codon (Wills et al., 1991). A diagram depicting these three types of read-through signals are shown below (**Figure 22**).

# Stop codon read-through signal

Type 1: NNN-STOP-CAR-YYA

Type 2: NNN-STOP-CGG

Type 3: NNN-STOP 

**Figure 22: Three types of read-through signals have been identified.**

Type 1 signals involve a hexa-nucleotide CAR-YYA motif (where R denotes purine bases and Y denotes pyrimidine) immediately downstream of the stop codon. Type 2 signals involve the local nucleotide context around a stop codon and type 3 involves a downstream RNA structure which may also bind to downstream sequences as a pseudoknot.

Alphavirus stop codon read-through was initially shown to rely on type 2 read-through signalling as mutagenesis of the cytidine alone of the adjacent CUA codon disrupted read-through of the SINV opal stop codon (Li and Rice, 1993), future work however identified that many cases of type 2 read-through are also promoted by a downstream RNA stem loop structure. A RNA structure identified in alphaviruses was shown to promote opal stop codon read-through in SINV and VEEV by up to 10-fold, the structure was shown to be strongly conserved amongst the alphaviruses (Firth et al., 2011). A diagram of this alphavirus read-through signal is shown below (**Figure 23**).

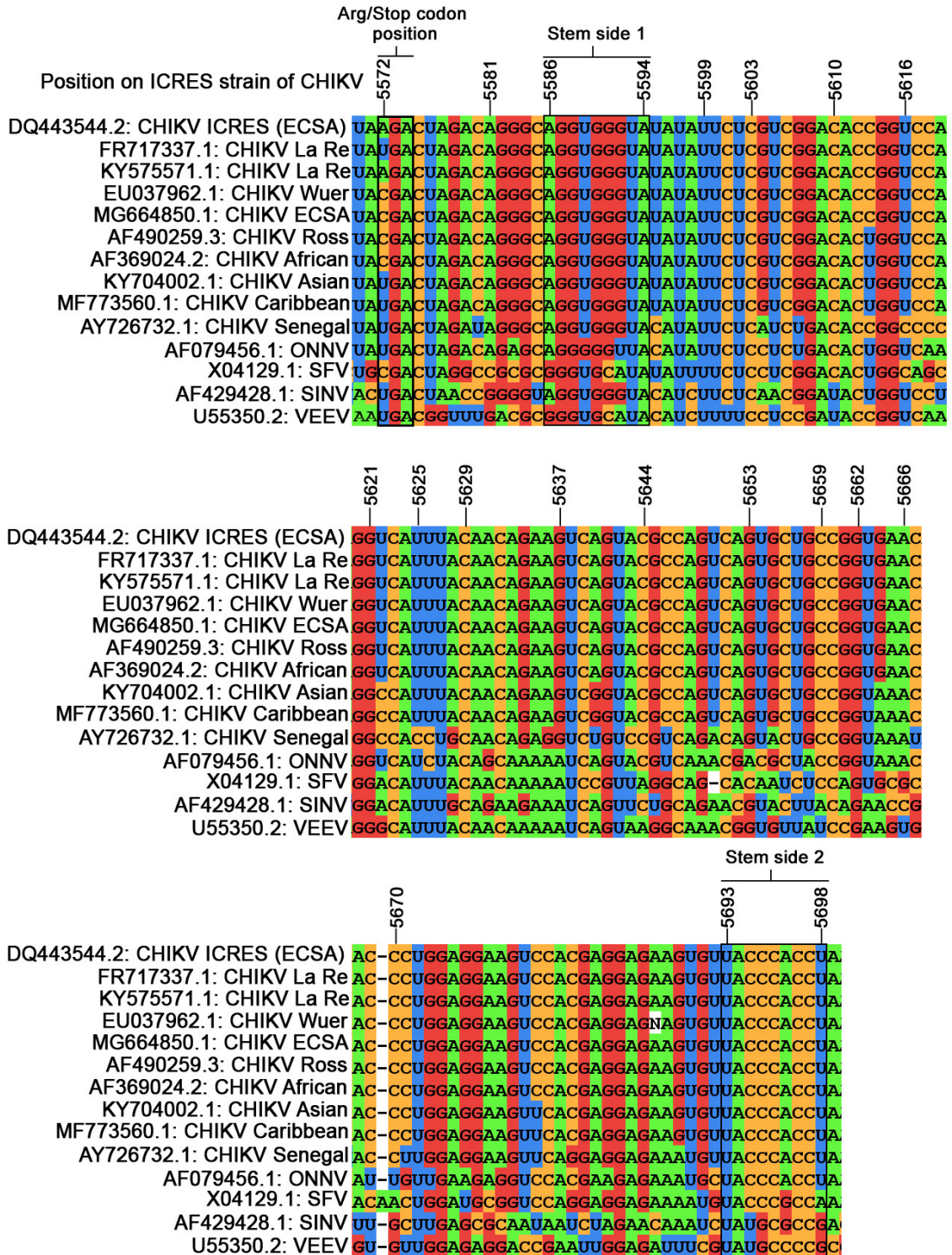
# Alphavirus read-through signal



**Figure 23: Alphavirus stop codon read-through signals.**

The cytidine nucleotide adjacent to the stop codon and a downstream RNA stem loop structure were both shown to promote stop codon read-through in alphavirus replication.

Sequences which correspond to the bottom segment of the RNA structure was identified by (Firth et al., 2011), mutagenesis of the components of the RNA structure in this region revealed the structure was important for opal stop codon read-through in the context of a Dluc reporter. The sequences for the ICRES CHIKV strain used in the lab, other strains of CHIKV and related alphaviruses were aligned to identify the sequences which match this identified bottom stem of this conserved RNA structure (**Figure 24**).



**Figure 24: The bottom stem of the RNA structure read-through signal identified in VEEV and SINV was identified in CHIKV sequences.**

Aligning the different alphavirus genomes revealed that this region was strongly conserved between them. Boxes were used to highlight the position of the Arg/opal codon and the sequences which correspond to the 2 sides of the bottom stem of the RNA structure was highlighted.

Aligning the CHIKV and alphaviruses sequences confirmed that the regions which form the bottom stem of the read-through signal RNA structure were present in the analysed CHIKV sequences (**Figure 24**). Structural regions of RNA are known to affect translational efficiency and ribosome stalling. Structures within RNA sequences can be predicted computationally and validated using experimental procedures (Mortimer et al., 2014). Programs such as Mfold can be used to analyse potential RNA structures, an array of minimum free energy models of structured RNA with favourable base pairing or single stranded regions of RNA are predicted. These predictions can be assessed using the P-num value which is a summary of the number of predicted RNA structure models which contain said base as single or double stranded (Zuker, 2003).

Selective 2'-hydroxyl acylation analysed by primer extension (SHAPE) is a technique that provides data representing RNA structure of a sampled RNA fragment at single nucleotide resolution. This technique involves chemically probing RNA transcripts using hydroxyl-selective electrophiles such as NMIA which attacks the 2'-hydroxyl group of ribose to form a 2'-O-adduct. NMIA reacts with all four RNA nucleotides, this reaction occurs preferentially on unstructured flexible nucleotides. Reverse transcription primer extensions using 5' labelled primers which anneal to sequences downstream of the probed region terminate at modified 2'-O-adduct sites, this is because the reverse transcriptase cannot read through the modified bases. The frequency of the resultant different sized cDNA products are compared to a ladder created using dideoxy nucleotides (ddNTPs) to reveal nucleotide positioning of unstructured regions of RNA (Low and Weeks, 2010).

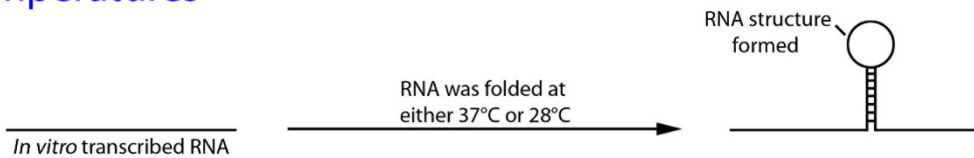
SHAPE using radioactively labelled primers was used in a recent study to investigate the presence of the read-through RNA structure using sequences based on Asian/African and Caribbean strains of CHIKV inserted into a Dluc vector. Their results validated the presence of an RNA structure in this region and suggested that the read-through signal promoted opal stop codon read-through of the Dluc replicon more effectively than the ochre and amber stop codons (Kendra et al., 2018). Their results provided a useful insight into the stop codon read-through structure however there were some limitations for this SHAPE experiment. The reactions were performed using RNA derived from dual luciferase plasmids containing approximately 200 bases of CHIKV sequence, this



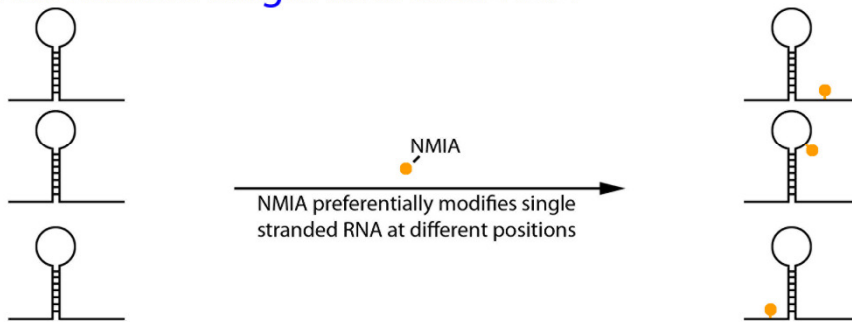
means the structures formed during folding would lack potential interactions which may occur other sections along the full length CHIKV genome. Additionally, the folding reaction in this study was solely performed at 37°C, although this temperature is phenotypically relevant to RNA structures formed during human host infection it does not provide us with much information about how these structures form at temperatures corresponding with the mosquito vector (28°C).

During this project the technique QuSHAPE, which utilises fluorescently labelled primers (FAM or HEX) analysed by capillary electrophoresis analysis (Karabiber et al., 2013), was used to investigate the presence of the read-through signal RNA structure in full-length ICRES CHIKV genome RNA folded at 37°C and 28°C. This protocol has the advantage of providing quantitative values of NMIA reactivity at a single nucleotide resolution. A diagram of the SHAPE protocol is seen in **(Figure 25)**.

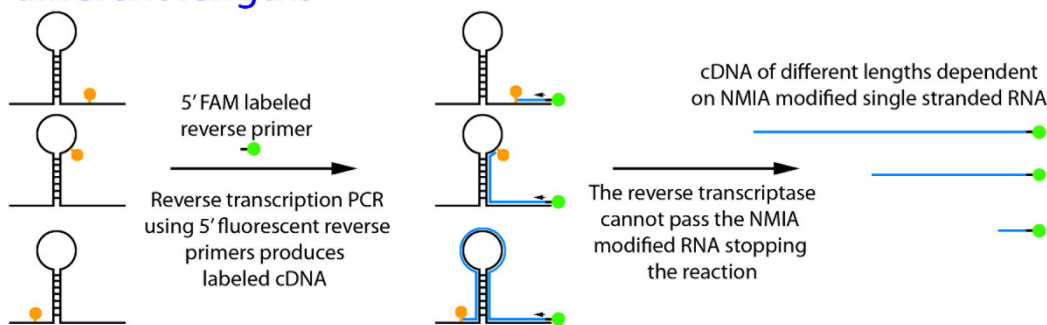
## 1. The *in vitro* transcribed RNA was folded at the desired temperatures



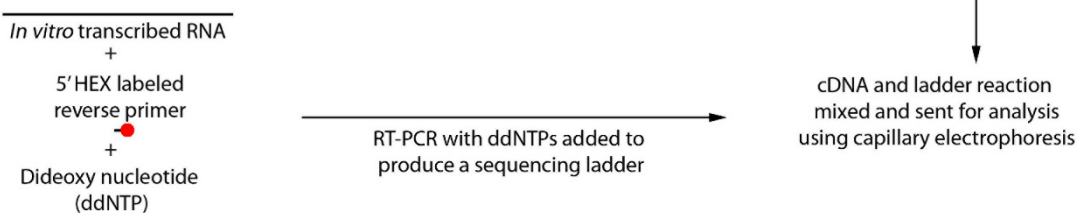
## 2. Folded RNA incubated with NMIA which preferentially modifies flexible single stranded RNA



## 3. Reverse transcription PCR (RT-PCR) using downstream 5' fluorescently labeled reverse primers produces cDNA of different lengths



## 4. Sequencing ladder reaction using unfolded RNA and 5' HEX reverse primer RT-PCR



### Figure 25: The QuSHAPE experimental procedure.

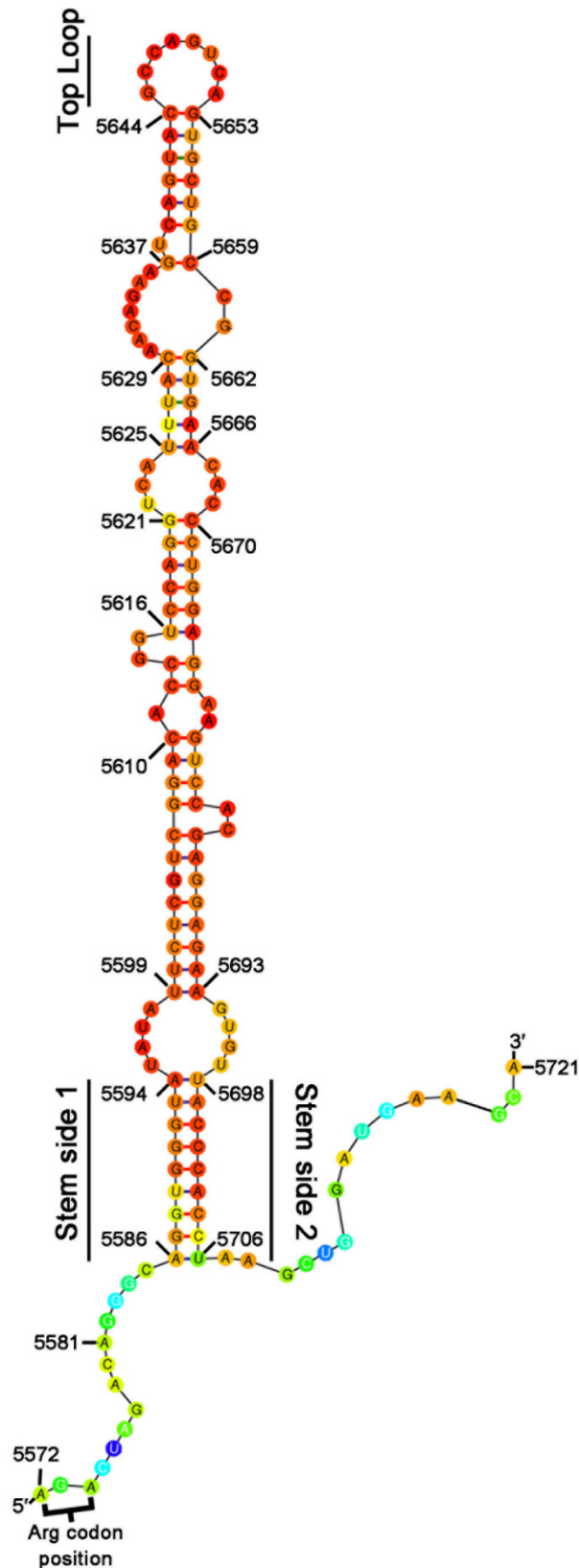
(1) *In vitro* transcribed ICRES CHIKV RNA (full length genome) was folded at 37°C or 28°C to form the RNA structures. (2) Folded RNA was incubated with the hydroxyl-selective electrophile NMIA which preferentially reacts with unstructured RNA to form 2'-O-adducts modified bases at those positions. (3) Reverse transcription PCR (RT-PCR) primed using downstream 5' FAM labelled reverse primers produces a pool of different sized labelled cDNA as the reverse transcriptase cannot pass the 2'-O-adducts modified bases. (4) A sequencing ladder made by RT-PCR with unstructured RNA, 5' HEX labelled reverse primers and an added ddNTP was generated. The cDNA from the SHAPE experiments was pooled with the sequencing ladder before analysis using capillary electrophoresis.

It was hypothesised that the different temperatures which correspond with the human and mosquito host body temperatures may result in changes to RNA folding and consequently the overall RNA structure. In addition, SHAPE experiments on the full-length CHIKV genome would reveal whether other parts of the genome could influence the formation of the read-through RNA structure. Three reverse primers which were designed to bind to an unstructured region of RNA downstream approximately 50 bases from the RNA structure were used for these SHAPE experiments, each primer was used for independent reactions with uncapped full length ICRES CHIKV genome RNA. Three independent primers were chosen to increase the chance of successful binding and to provide a robust analysis of the RNA structures formed in this region.

## **4.2 Results**

### **4.2.1 The read-through signal RNA structure was detected in the full-length CHIKV genome RNA when folded at 37°C**

A segment of the CHIKV genome was computationally folded at 37°C using the Mfold program to predict the RNA structure downstream of the opal stop codon position. The colour scheme depicts the P-num plot for this structure, on this scale red colouring (a score of 0) represents well determined participation in RNA base-pairing or if unstructured RNA up to black (a score of 1) represents no participation in RNA base pairing (Zuker and Jacobson, 1998) (**Figure 26**).

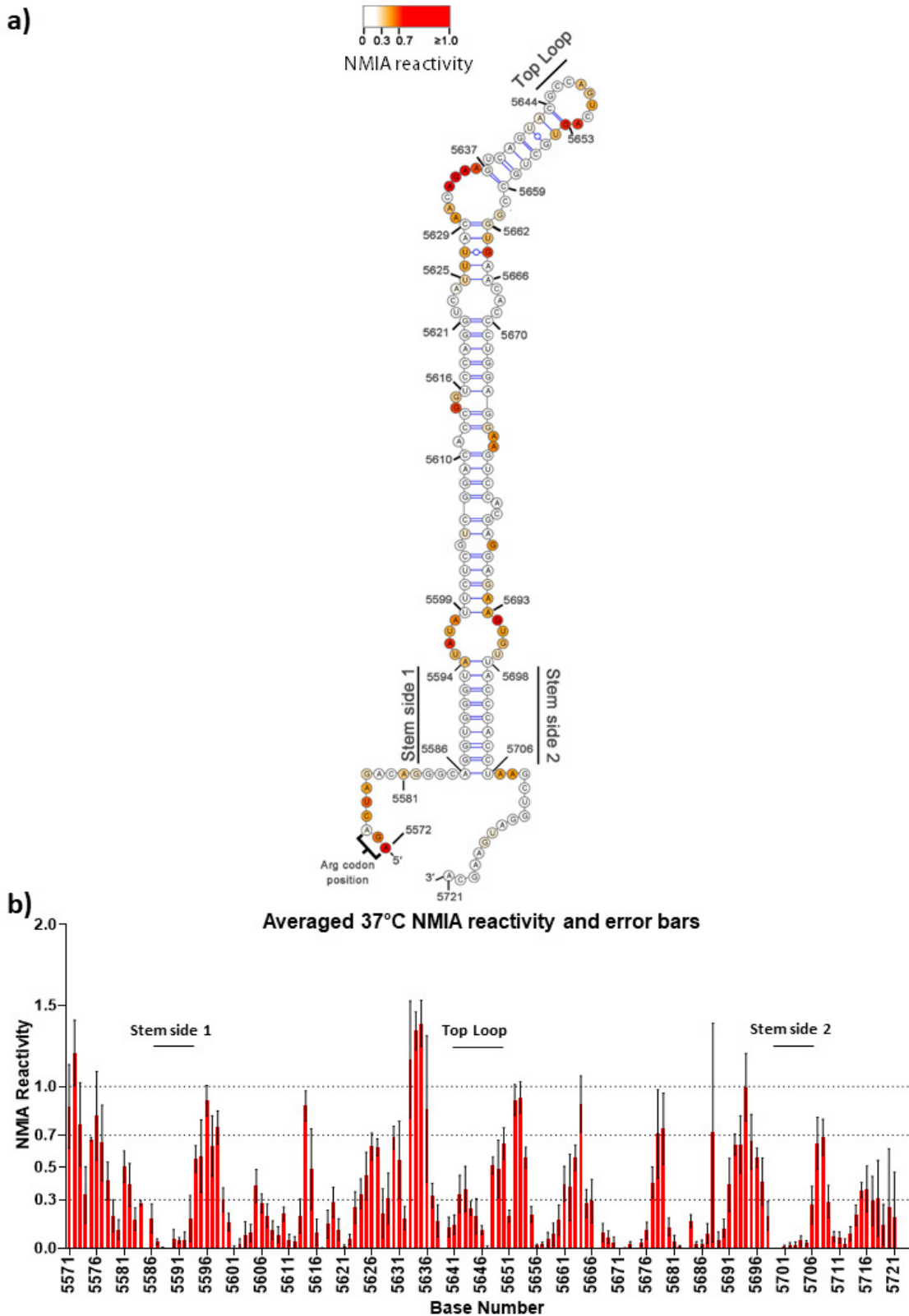


**Figure 26: The read-through signal RNA structure was predicted to fold when simulated using Mfold (37°C).**

A segment of the CHIKV genome was analysed with Mfold to predict potential RNA structures downstream of the Arg codon. The simulation was done for folding at 37°C. The P-num values were plotted for each base was plotted. The scale goes from Red (well determined prediction of RNA base pairing or unstructured), green as the midpoint and black as the highest value (poorly determined).

The read-through signal RNA structure was shown to form downstream of the Arg codon position when RNA folding was modelled using Mfold at 37°C (**Figure 26**). This structure was experimentally verified with a QuSHAPE experiment using full-length infectious ICRES CHIKV clone RNA generated via *in vitro* transcription. This RNA was then folded at 37°C to form RNA structures which would represent those present at physiological temperature. The folded RNA was treated with NMIA to form the 2'-O-adducts as previously described or DMSO as a negative control. This RNA was then subject to reverse transcription PCR using 5' FAM labelled reverse primers which specifically bind downstream of the predicted RNA structure region. A ddNTP ladder was also generated for each primer reaction using 5' HEX labelled reverse primers. The samples were mixed with the corresponding sequencing ladder for analysis using capillary electrophoresis.

The fluorescent signals which correspond to the FAM and HEX signal reveal details related to the modification of RNA with NMIA and the RNA nucleotide sequence respectively. The reactivity at a single nucleotide resolution was superimposed onto the RNA structure predicted by Mfold using the VARNA software (Darty et al., 2009) (**Figure 27a**), the averaged NMIA reactivity for each nucleotide position was plotted (**Figure 27b**).

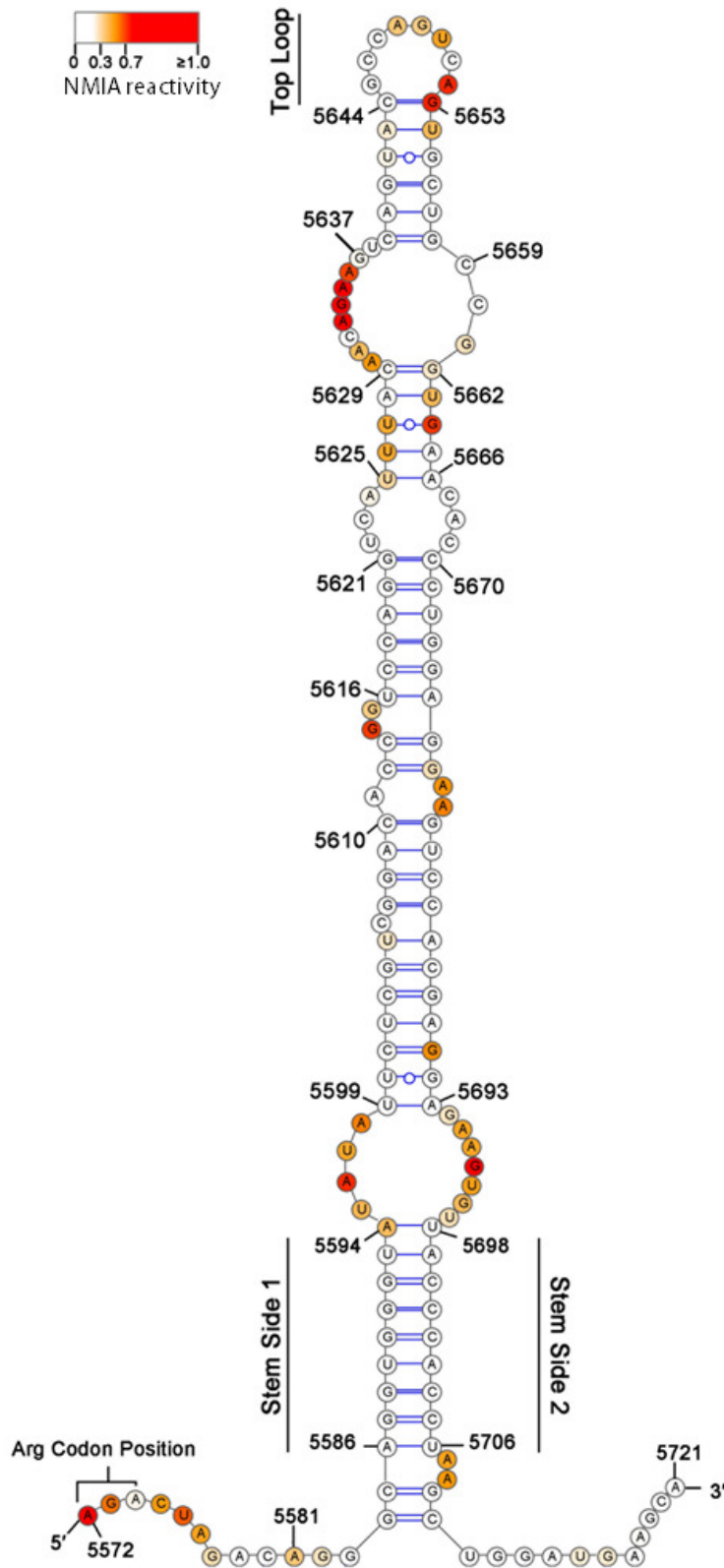


**Figure 27: The predicted RNA structure downstream of the opal stop codon was validated using QuSHAPE.**

(a) The averaged NMIA reactivity generated by QuSHAPE on RNA folded at 37°C was superimposed on the Mfold predicted RNA structure using the VARNA software. The key represents NMIA reactivity: 0 to 0.3 (white, base paired), 0.3 to 0.7 (Orange, moderate NMIA reactivity) and 0.7 to  $\geq 1.0$  (Red, high NMIA reactivity) (b) The averaged NMIA reactivities for each base were plotted. Error bars represent S.E.M. N=4.

The predicted read-through signal RNA structure was validated by the NMIA reactivity data obtained using QuSHAPE as shown in Figure 27a and Figure 27b, moderate to high NMIA reactivity strongly coincides with the regions of predicted single stranded regions by Mfold. The two sides of the bottom stem formed by nucleotides at position 5586- 5594 and the complimentary 5698- 5706 positions were shown to be unreactive to NMIA, this was to be expected as this double stranded region was identified as an important for forming the RNA structure (**Figure 24**). The moderate to high levels of NMIA reactivity seen at the top of the predicted structure (5644- 5653) and along the structure (5594-5599, 5693-5698 and 5629-5673) are indicative of single stranded regions of RNA. These single stranded regions strongly correspond with the predicted bulges and top loop of the RNA structure predicted by Mfold. Taken together the NMIA reactivity confirms the presence of the large predicted RNA structure downstream of the opal stop codon.

A predicted model generated using the RNAstructure software guided by the experimental SHAPE data can propose a more accurate model of predicted RNA structures; SHAPE reactivities are used as pseudo-free energy change terms to constrain the resultant RNA model (Deigan et al., 2009). The SHAPE data was used to improve the RNA structure present in the full-length RNA genome of CHIKV folded at 37°C (**Figure 28**).



**Figure 28: SHAPE reactivity adjusted structural prediction of the read-through RNA structure in the CHIKV genome RNA folded at 37°C.**

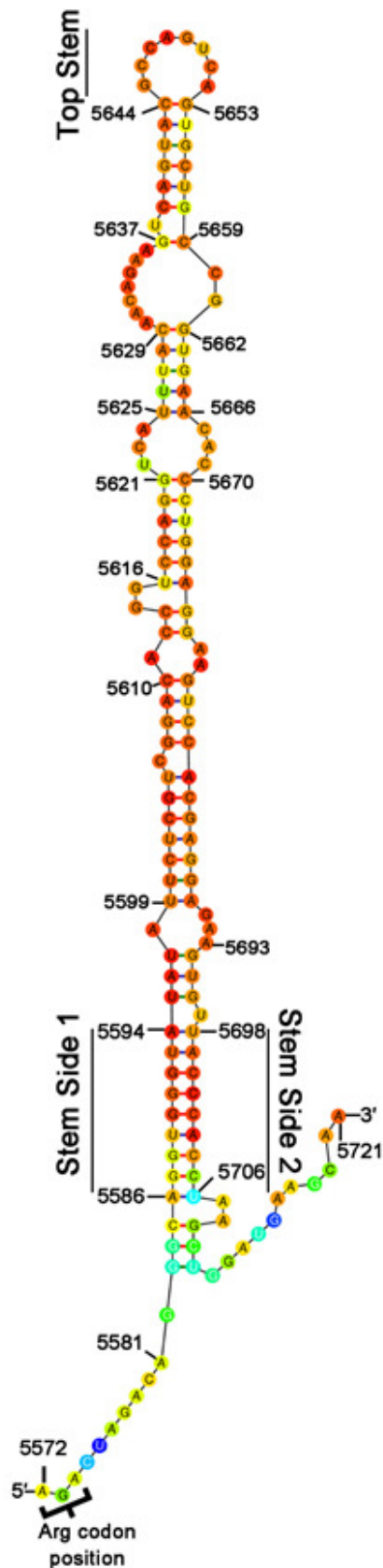
SHAPE data was used to constrain the predicted model produced using RNAstructure to increase the accuracy of the generated model. The NMIA reactivities are shown overlaid on the modelled structure. The key represents NMIA reactivity: 0 to 0.3 (white, base paired), 0.3 to 0.7 (orange, moderate NMIA reactivity) and 0.7 to  $\geq 1.0$  (Red, high NMIA reactivity).



The redrawn RNA structure using the SHAPE data as constraints can be seen in Figure 28. The experimentally determined structure closely matches the one predicted by Mfold. The two sides which make up the bottom stem and the top loop remain unchanged between the two models; some small differences can be seen such as the small bulge seen between bases 5706 to 5709. Both of these structures share features with the RNA structure proposed by (Kendra et al., 2018), this suggests that the QuSHAPE pipeline used in this study successfully predicted a highly accurate RNA structure for the recoding signal RNA structure in the context of the full length genome. Whether CHIKV forms these RNA structures in the mosquito vector is currently not known, QuSHAPE on full-length ICRES CHIKV RNA folded at 28°C was performed to investigate whether this RNA structure could form at the body temperature of a mosquito.

#### **4.2.2 The stop codon read-through signal RNA structure forms when the full-length CHIKV genome is folded at 28°C**

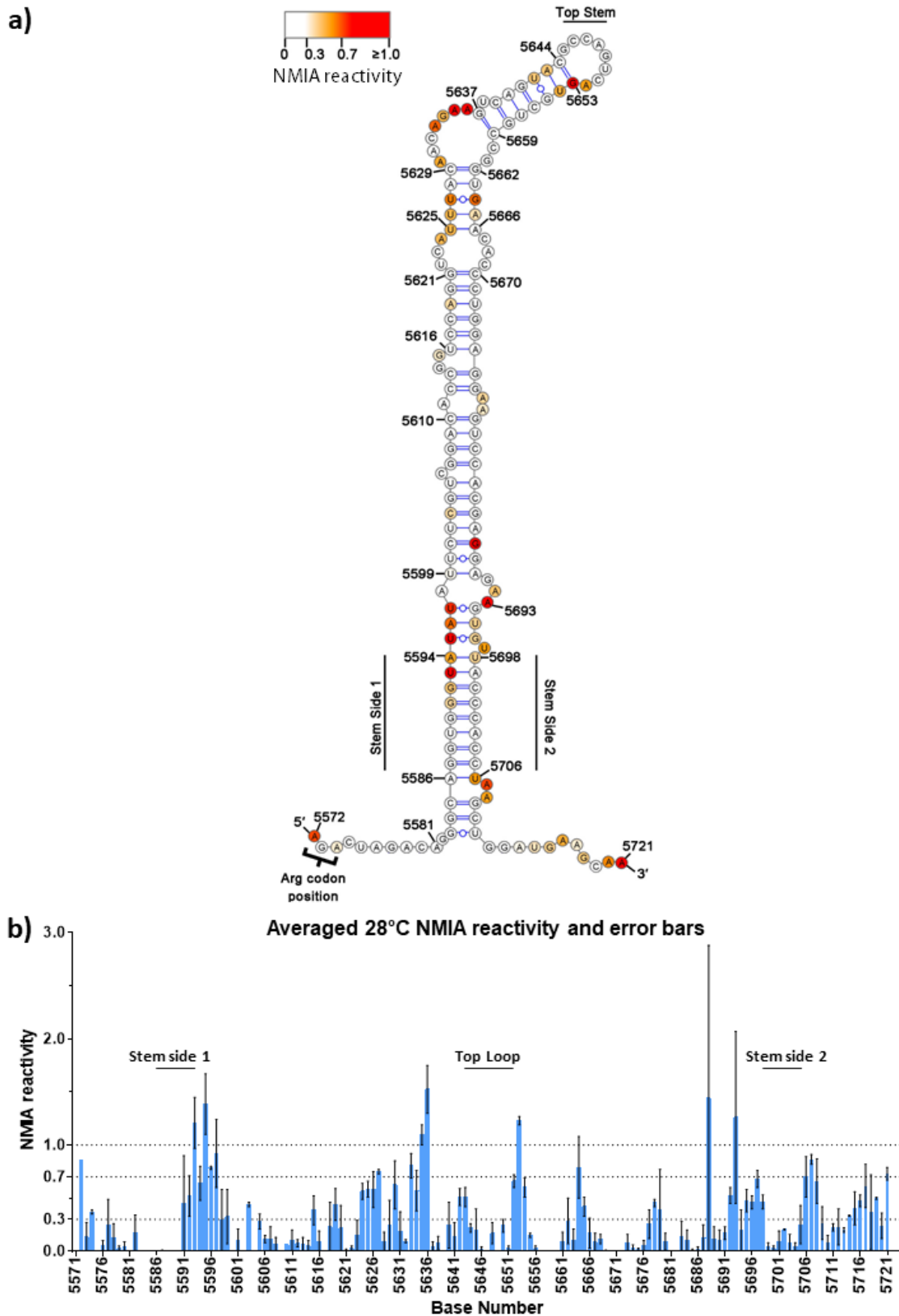
Replication of the CHIKV genome was enhanced for the three different stop codons in the context of the CHIKV SGR and infectious clones in the C6/36 mosquito cell line (**Figure 16 and Figure 19**). The mechanism behind promoting stop codon read-through is not fully understood, whether the mechanism is the same between the vertebrate and invertebrate host is also unknown. Alphavirus RNA structures have been shown to promote virus growth in the context of mosquito cells; mutations of the conserved sequence element (CSE) in the nsP1 coding region in SINV were shown to have no effect on mammalian cell replication but abolished replication in a mosquito cell line (Fayzulin and Frolov, 2004). A direct repeat RNA structure in the 3' untranslated region (UTR) of an Asian strain of CHIKV was shown to be important for adaptation for replication in mosquito cells but not in mammalian cells (R. Chen et al., 2013). Mfold was used to predict the RNA structure of the segment of CHIKV sequence when folded at 28°C (**Figure 29**).



**Figure 29: The read-through signal RNA structure was shown to fold in the CHIKV sequence when folded at 28°C.**

A segment of the CHIKV genome was analysed using Mfold simulating the folding at 28°C. The P-num values for each base were plotted. The scale goes from red (well determined RNA base pairing or unstructured RNA), to green as a midpoint and up to black (poorly determined).

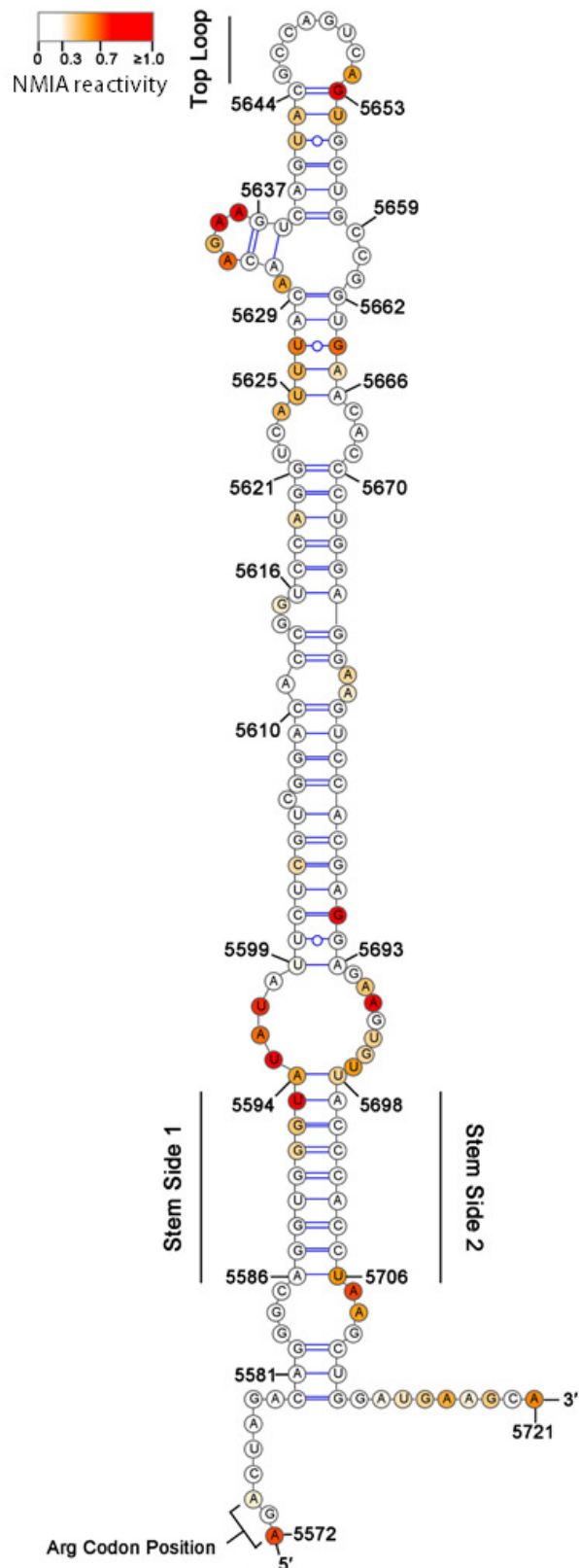
The read-through signal was shown to form in RNA structure predicted using Mfold for the lower temperature of 28°C (**Figure 29**). The previously described QuSHAPE experiment was performed on full-length genome CHIKV RNA folded at 28°C, this was done to investigate potential differences in RNA structure when folded at mosquito host body temperature which may provide insights into the increased read through efficiency of the stop codons in the mosquito cells seen in the previous chapter. The SHAPE data was overlaid on the Mfold generated structure to validate the prediction (**Figure 30a**). A graph of the NMIA reactivity of the 28°C folded ICRES CHIKV RNA was generated (**Figure 30b**).



**Figure 30: The RNA structure predicted to fold at 28°C was validated using QuSHAPE.**

(a) The NMIA reactivity from QuSHAPE experiment for RNA folded at 28°C was overlaid on top of the predicted Mfold structure. The key represents NMIA reactivity: 0 to 0.3 are shown in (white, base paired), 0.3 to 0.7 (orange, medium reactivity) and 0.7 to  $\geq 1.0$  (red, high reactivity). (b) Averaged NMIA reactivity for the QuSHAPE experiment were plotted. Error bars represent S.E.M. N=2.

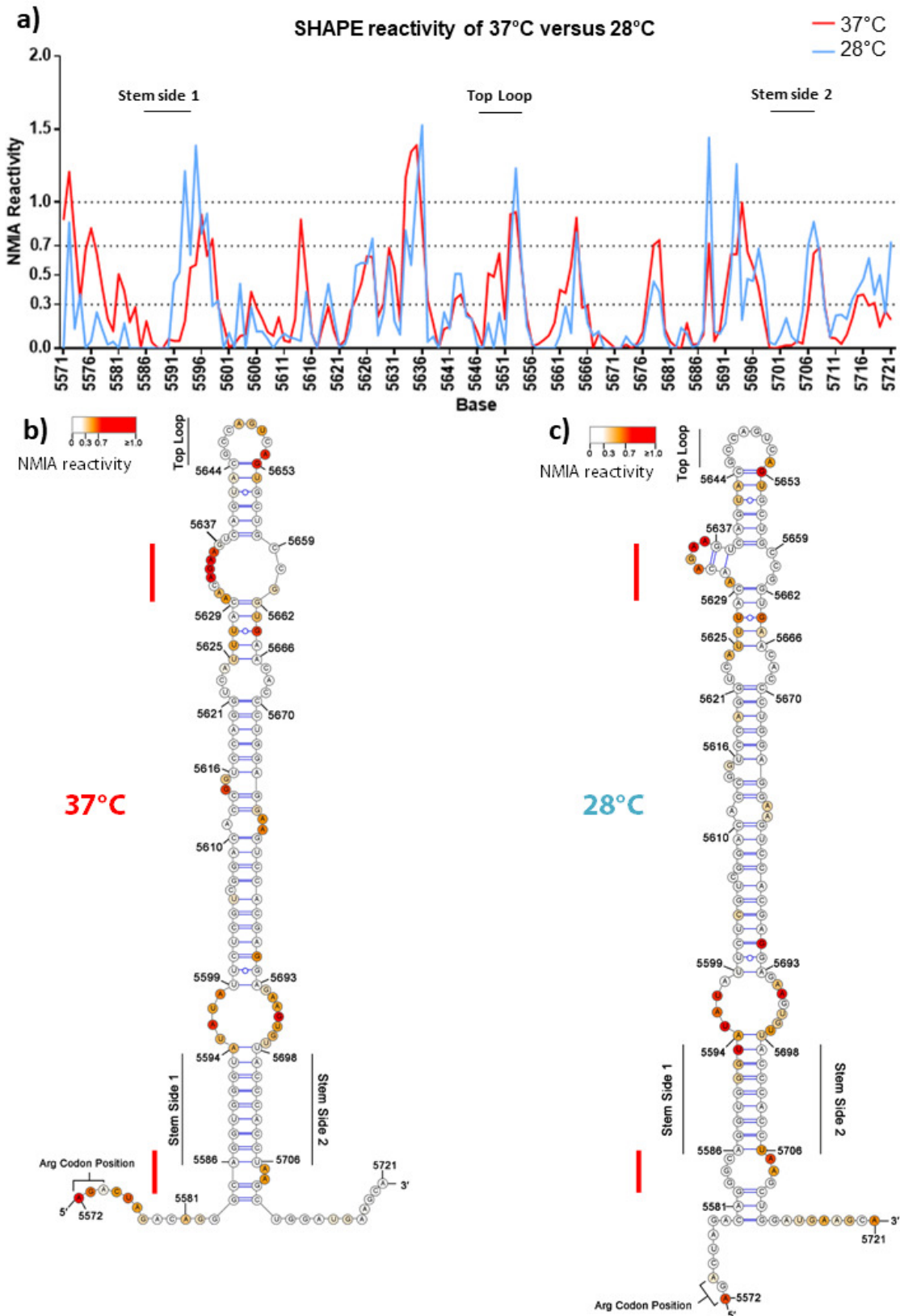
The NMIA reactivity of the 28°C folded ICRES CHIKV RNA was used to label the Mfold predicted RNA structure and a graph of the averaged values for each base can be seen in Figure 30. The large RNA stem loop was shown to still form at this lower temperature. The sides of the conserved bottom stem and the top loop were shown to be present in the above predicted RNA structure, the bulge seen at bases 5706 to 5709 resembles the one seen in the RNAstructure program predicted 37°C stem loop in Figure 28. An improved model of the RNA structure was predicted using the RNAstructure program and the SHAPE data for the 28°C fold (**Figure 31**).



**Figure 31: An improved model using the SHAPE data for the 28°C folded CHIKV RNA structure was generated.**

RNAstructure program was used to construct the RNA structure using the SHAPE data to constrain the prediction of CHIKV RNA folding at 28°C. The NMIA reactivity was overlay onto the predicted structure. A key of the NMIA reactivity is shown: 0 to 0.3 (white, base paired), 0.3 to 0.7 (orange, medium reactivity) and 0.7 to  $\geq 1.0$  (red, high reactivity).

The adjusted RNA structure for the 28°C folded CHIKV RNA can be seen in Figure 31. The RNA structure shares many common features with the SHAPE data directed predicted 37°C stem loop in Figure 28. Some differences are present however, a more prominent bulge was identified between the bases 5629 to 5637 and the reactivity of the top loop appears to be less susceptible to NMIA than seen for the 37°C RNA structure. The averaged NMIA reactivity for the 37 °C and 28°C folded RNA reactions were plotted side by side to identify regions of similarity and differences in the single stranded regions of RNA, the two experimentally determined structures are presented side by side with differences highlighted by a red line for clarity (**Figure 32**).





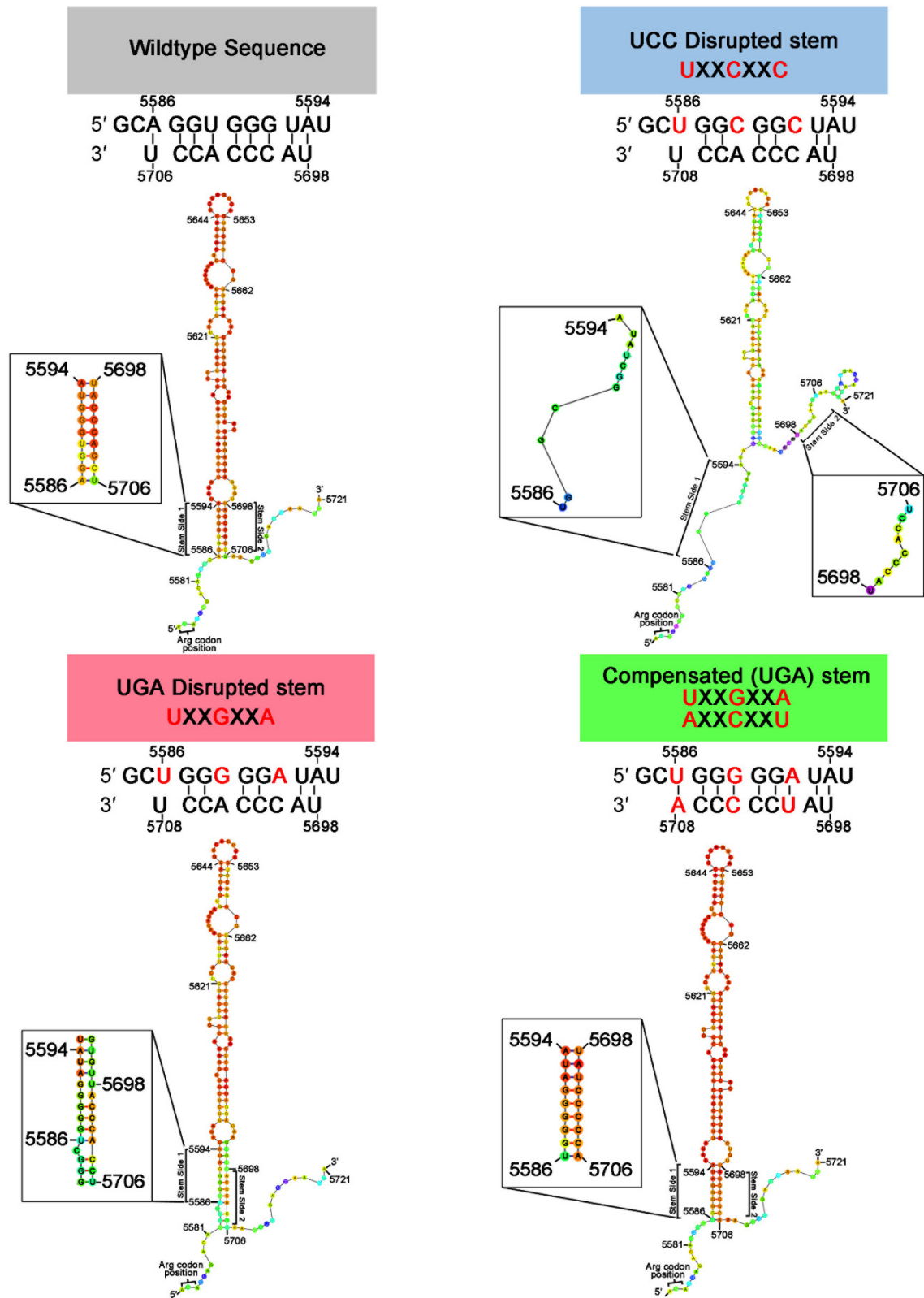
A clear depiction of shared regions of single stranded RNA between the stem loops of CHIKV RNA folded at 37°C and 28°C was seen when the averaged NMIA reactivities were plotted together and when the RNA models for folding at both temperatures are presented side by side (**Figure 32**). The reactivity between bases 5621 up to 5696 form peaks which are highly similar in size and shape. Some changes are seen for the 28°C reactivity for the stem side 1 which appears to be more reactive at base 5593. This comparison suggests that the read-through signal RNA structure can form at both 28°C and 37°C, whether this feature could promote read-through in both mammalian and insect cells is currently unknown.

#### **4.2.3 Disrupting the stop codon read-through recoding RNA structure disrupts the production of CHIKV in BHK cells**

Programmed read-through of the opal stop codon by the CHIKV read-through signal RNA structure has been shown using D<sub>luc</sub> plasmids; one luciferase is separated by the region containing the RNA structure, successful expression of the second luciferase reporter is dependent on successful stop codon read-through (Kendra et al., 2018). In another study mutagenesis of the VEEV and SINV RNA structure inserted between two luciferase reporters was shown to drastically reduce the expression of the read-through dependent luciferase; mutations which disrupted the Watson-Crick pairing in the bottom stem was shown to drastically reduce read-through efficiency but could be compensated back to wildtype levels by recovering the interaction (Firth et al., 2011).

The effect of disrupting the RNA structure in the context of the full CHIKV life cycle is currently unknown, to investigate this a panel of ICRES CHIKV mutants with disrupted and compensated RNA structures were generated. Site directed mutagenesis was used to change the primary coding sequence whilst maintaining the amino acid sequence. For all three stop codons the RNA stem loop was disrupted by changing the first, fourth and seventh nucleotide on the stem side one to either UXXCXXA (UCC disrupted stem) or UXXGXXA (UGA disrupted stem) to disrupt the Watson-Crick base pairing to the second side of

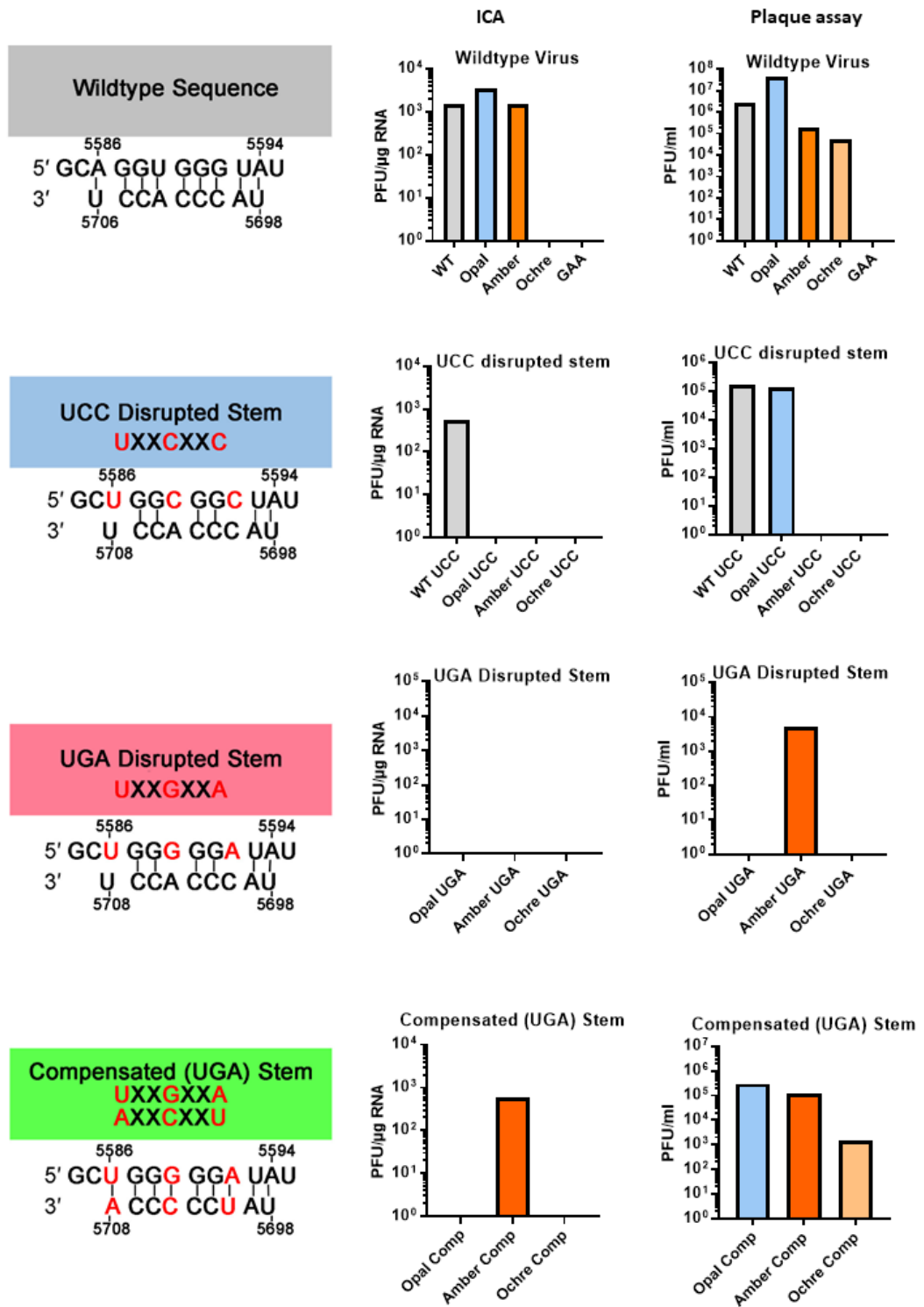
the stem. The UGA disrupted mutant was also compensated by substituting the nucleotides on the stem side 2 to UXXCXXA (Compensated (UGA) stem) to recover the base pairing between the two sides. The effect of these mutations on the RNA structure of the wildtype ICRES CHIKV when folded at 37°C was predicted using Mfold (**Figure 33**).



**Figure 33: Mfold predicted structures of disrupted and compensated RNA structures in the context of the wildtype ICRES CHIKV mutants folded at 37°C.** The base of the RNA stem loop was mutated to remove Watson-Crick interactions without disrupting the amino acid sequence. Both the UCC and UGA disrupted stem mutations caused the stem loop to reorganise whilst the compensated (UGA) stem mutant recovered the structure. The structures were simulated at 37°C. The boxes show a zoomed view of the bottom stem.

Disrupting the Watson-Crick base pairing of the bottom stem of the RNA structure changes the structure of the predicted RNA stem loop when simulated (37°C), the RNA structure can be recovered by compensatory mutations on the opposing side of the stem (**Figure 33**). Disrupting the stem with the UCC substitution causes the base of the predicted stem to no longer base pair and decreases the P-num values across the whole RNA structure. The UGA disruption mutant reduces the P-num value of the bottom of the RNA structure which means binding in this region is poorly determined, the two sides of the bottom stem also formed interactions with different bases on the opposing side of the stem. The stem was recovered back to the wildtype structure when the compensatory mutations were introduced.

The effect of the disrupted and compensated RNA structure on CHIKV replication in mammalian cells was investigated using BHK cells. An ICA experiment and plaque assay of the resultant virus stocks 24 h.p.t using the wildtype stop codon mutants, UCC disrupted stem mutants (WT, opal, ochre and amber), UGA disrupted stem mutants (opal, ochre and amber) and compensated (UGA) stem (opal, ochre and amber) was performed (**Figure 34**).



**Figure 34: Disrupting the RNA structure decreases the RNA infectivity and production of virus for the stop codon ICRES CHIKV mutants.**

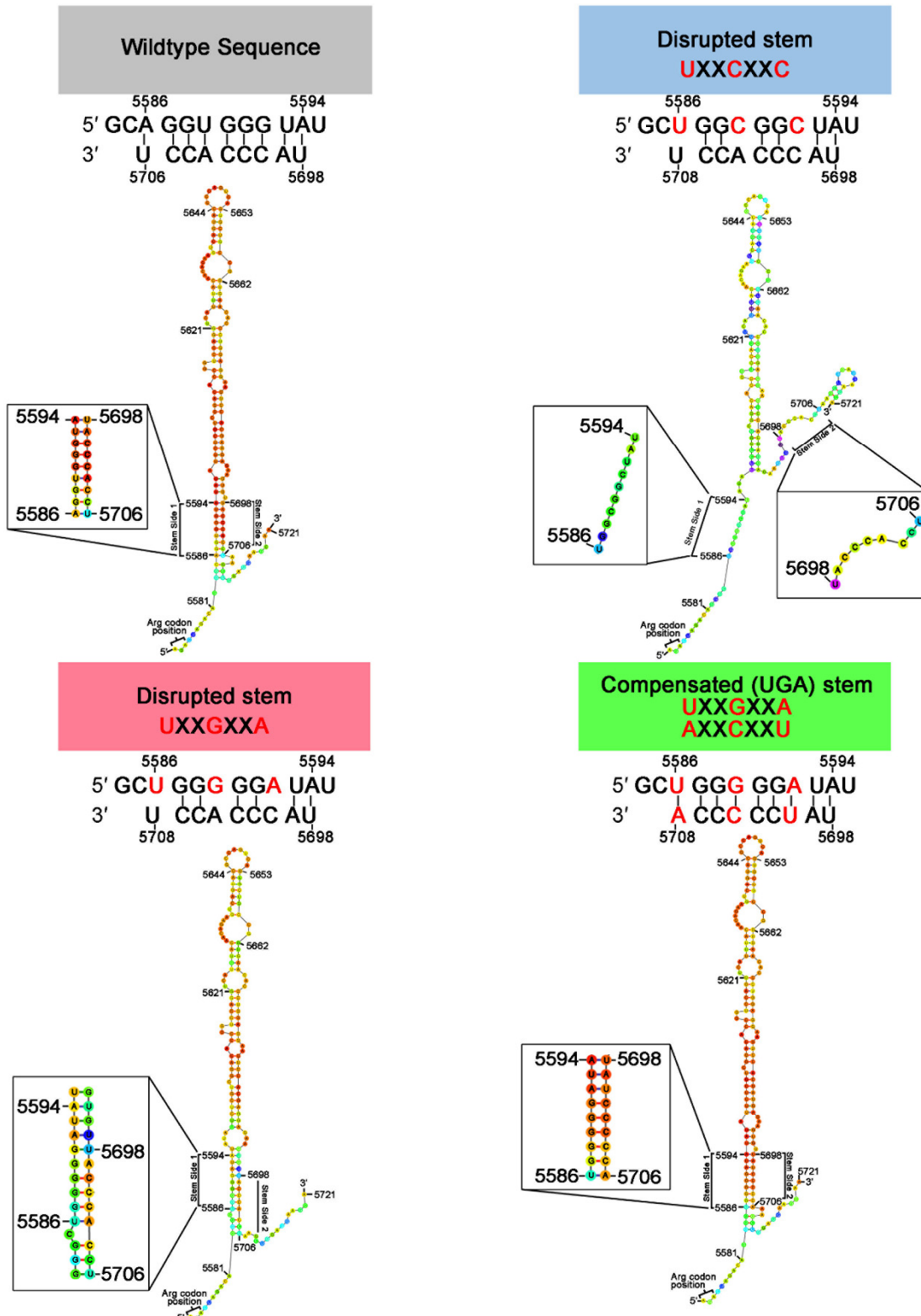
The original stop codon mutants, UCC and UGA disrupted stem and compensated (UGA) stem mutants of ICRES CHIKV were transfected into BHK cells for ICA and plaque assay of the collected virus stocks on BHK cells. N=1.

The ICA and viral titres of the disrupted and compensated mutant ICRES CHIKV clones in BHK cells is shown in Figure 34. The WT UCC disrupted mutant revealed that disrupting the RNA structure had no effect on the RNA infectivity ( $10^3$  PFU/ $\mu$ g of RNA) or resultant viral titre ( $10^5$  PFU/ml) as the mutant produced original WT ICRES CHIKV levels. This was as predicted as the read-through signal would be expected to be less important for this clone as the sense codon arginine does not require recoding. Disrupting the RNA structure with either the UCC or UGA mutations resulted in a severe decrease in RNA infectivity of all three stop codons. No viral titre was detected for the ochre stop codon CHIKV when the RNA structure was disrupted with either the UCC or UGA mutations. Interestingly the opal UCC and amber UGA disrupted stem viruses resulted in the production of viral titres of  $10^5$  and  $10^4$  PFU/ml but no RNA infectivity of either virus was detected. Compensating the UGA mutations to recover the RNA structure resulted in the recovery of the amber mutant RNA infectivity to detectable levels ( $10^2$  PFU/ $\mu$ g of RNA) but not for the opal compensated mutant. It is possible that the RNA infectivity of these mutants was undetected as they were below detectable limits at this time but could also suggest potential reversions. All three stop codon compensated (UGA) mutants did generate detectable titres of virus: opal ( $10^5$  PFU/ml), amber ( $10^5$  PFU/ml) and ochre ( $10^3$  PFU/ml).

Disrupting the RNA structure had the biggest impact on the ochre stop codon mutant as both UCC and UGA disrupted stem mutations resulted in no detectable virus, mutations restoring the stem resulted in the production of viral titres akin to the original ochre ICRES CHIKV. The opal UGA and amber UCC disrupted stem mutants did not produce detectable virus either. It is possible that disrupting RNA structure results in less efficient read-through of the stop codons which reduces the amount of nsP4 expressed during infection, this would have the biggest effect on the ochre stop codon mutants which have been shown to display already reduced read-through efficiency in mammalian cells.

#### **4.2.4 Disrupting the stop codon read-through recoding RNA structure does not affect the production of CHIKV in the C6/36 mosquito cells**

The phenotype of these disrupted and compensated RNA structure mutations in the context of a mosquito cell line was investigated as the read-through RNA structure was shown to form at 28°C. The mutants were screened using the C6/36 mosquito cells, the function of the RNA structure in mosquito cells has not yet been investigated. Mfold simulated folding of the disrupted and compensated stem loop mutants at 28°C was performed to investigate the changes in the structure which would form at the mosquito body temperature (**Figure 35**).



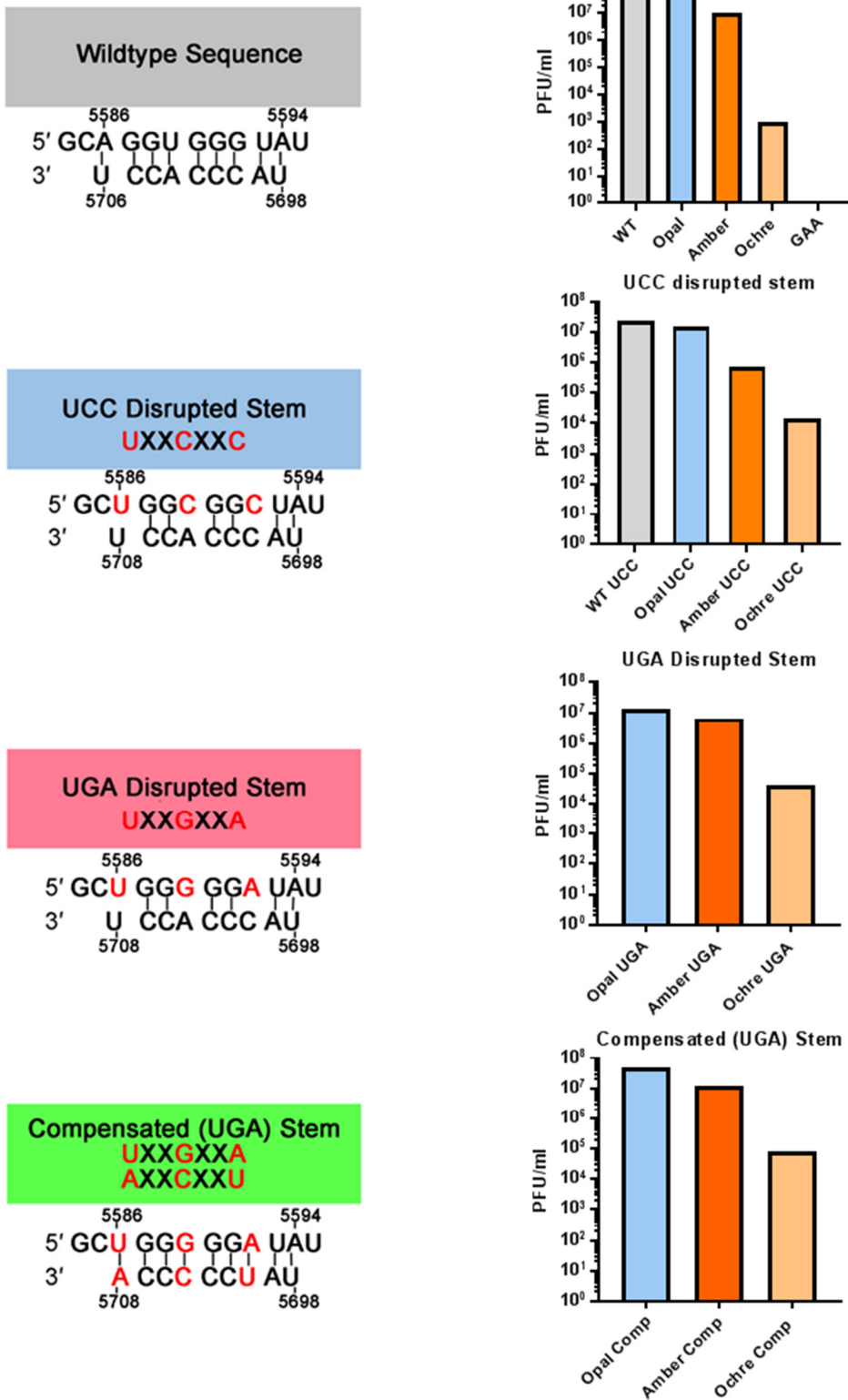
**Figure 35: Mfold predicted structures of disrupted and compensated RNA stem loop in the context of the wildtype ICRES CHIKV when folded at 28°C.**

The structure was disrupted by removing Watson-Crick base pairing using the UCC and UGA disrupted stem mutations which still maintain the correct amino acid sequence of this region. The compensated (UGA) stem recovers the base pairing of the second disrupted mutant. These structures were simulated for folding at 28°C. The boxes show a zoomed in view of the bottom stem.



The introduced disruption and compensation mutations caused changes in the RNA structures which were simulated at 28°C by Mfold (**Figure 35**). The UCC disrupted stem mutations caused the bottom stem of the structure to no longer base pair and caused the P-num values of the rest of the RNA stem loop to decrease. The UGA disrupted mutations decreased the P-num values of the bottom stem to decrease and caused the bottom stem to rearrange as base pairing in this region was altered. The wildtype structure of the RNA stem for the UGA mutant was recovered in the compensated mutant.

The disrupted and compensated structures folded at 28°C resemble the altered structures predicted when these mutants were simulated for 37°C (**Figure 33**). This consistency suggests that the bottom stem is important for forming the RNA structure at both temperatures. These mutants were transfected into C6/36 cells to investigate the effect on CHIKV replication in mosquito cells, virus stocks were harvested after 48 hours post transfection and plaque assayed on BHK cells (**Figure 36**).



**Figure 36: Disrupting the RNA structure did not affect CHIKV replication in C6/36 cells.**

The disrupted and compensated RNA structure mutants were transfected into C6/36 cells. The virus stock was collected 48 hours and plaque assayed using BHK cells. N=1.

Graphs showing the viral titres of the disrupted and compensated RNA structure CHIKV mutants released by transfected C6/36 cells are presented in Figure 36. The WT UXXCXXC mutant produced wildtype levels of virus displaying a titre of  $10^7$  PFU/ml. Disrupting the RNA structure with the UXXCXXC or UXXGXXA mutations had no effect on any mutant virus produced in transfected C6/36 cells, the titres for the opal and amber disrupted mutants were both at original clone titres of  $10^7$  PFU/ml. The ochre UXXCXXC and UXXGXXA mutants resulted in 10-fold higher detected virus. The compensated mutants also displayed wildtype levels of virus titre. These results suggest that disrupting the RNA structure at this bottom stem did not reduce the infectivity of any stop codon mutant in the C6/36 cells. This result contrasts to the generally decreased virus titre seen when the same mutants were transfected into BHK cells (**Figure 34**), potentially suggesting different requirements for stop codon read-through within the two hosts.

### 4.3 Discussion

The stop codon read-through signal RNA structure was shown to form at both 37°C and at 28°C, these temperatures correspond to the body temperatures of the human and mosquito host (**Figure 32**). Some slight changes to the structure was seen at these temperatures suggesting that the external temperature might influence the folding of the RNA structure. The RNA structure model presented in this study was similar to the model proposed by (Kendra et al., 2018) for the CHIKV segment inserted into the Dluc plasmid. This validates the accuracy of the QuSHAPE method and consequent predicted RNA structure proposed here, additionally it appears that context of the full-length RNA genome did not change the folding of the RNA structure.

Different phenotypes were seen when the disrupted RNA structure mutants were transfected into the BHK and C6/36 cells. These preliminary results demonstrate potential phenotypes to investigate in future studies but should be taken with some scepticism until further repeats are performed. The disrupted RNA structure did not reduce the titre of the arginine coding CHIKV produced in BHK cells but the opal UCC, amber UGA and ochre UCC and UGA disrupted stem mutants did not replicate. Compensating the UGA disrupted stem resulted in all three stop

codon mutants replicating to original virus levels (**Figure 34**). In C6/36 cells the disrupted RNA stem loop did not disrupt the replication of any stop codon containing virus as the opal, amber and ochre UCC, UGA disrupted and UGA compensated mutants replicated at original mutant levels (**Figure 36**).

Abundant read-through during insect cell translation has been identified previously, the *A. gambiae* mosquitoes were shown to have 600 functional read-through sites although the exact mechanism of read-through in insects is unknown (Jungreis et al., 2016). The conditions which promote read-through in mosquitoes could perhaps influence the recoding of the nsP3 stop codon hence increasing the read-through of the stop codons seen in this study.

A proposed mechanism of stop codon read-through is the insertion of amino acids by nc-tRNAs which can incorporate different amino acids instead of terminating translation depending on the stop codon: UGA (arginine, cysteine or tryptophan); UAG (glutamine, tyrosine, leucine, lysine or tryptophan) and UAA (glutamine or tyrosine) (Chittum et al., 1998). The amino acid at this position is inserted by nc-tRNA which are normal tRNA molecules that form unconventional base pairing between the stop codon to the anti-codon that results in the recoding of the stop codon. This near cognate base pairing is promoted by favourable read-through signals around the stop codon, this allows nc-tRNA to function essentially as suppressor tRNA (Beier and Grimm, 2001). The exact mechanism of the stop codon recoding RNA structure is unknown; however, possible functions include stalling the ribosome to allow time for near-cognate recognition, or as a physical barrier during translation to block the release factor. The RNA structure presented in this chapter could possibly synergise with the read-through promoting cytidine residue downstream of the stop codon to facilitate the binding of suppressor tRNA and hence promote programmed read-through.

Programmed read-through is determined by the translation factor eIF3 which is dependent on the third nucleotide “wobble” position of the stop codon, eIF3 blocks the release factor eRF1 decoding the stop codon while also allowing the nc-tRNA incorporation (Beznoskova et al., 2015). These suppressor tRNAs are incorporated at different read-through rates, the conservation of one particular stop codon at a read-through site was therefore suggested to maintain the

desired incorporated amino acids to form a functional protein and to maintain efficient read-through (Jungreis et al., 2011). This could suggest a possible reason for the inefficient read-through seen for the ochre but not opal or amber stop codons in this study.

This panel of disrupted and compensated RNA structure mutants can be used for future research investigating the mechanism of read-through recoding. SHAPE reactions on RNA generated from the disrupted and compensated mutants would confirm the predicted phenotypes by Mfold at 37°C and 28°C folding temperatures (**Figure 33 and Figure 35**). Investigating how the RNA structure forms in vivo during CHIKV infection of human and mosquito cells would be of great interest due to differences in replication seen in the two hosts. In-Cell SHAPE (Diaz-Toledano et al., 2017) results could identify any regions of RNA which fold differently in the presence of the host cell factors which may provide further insights into the differences in read-through seen in this study.

The RNA structure may facilitate amber and ochre read-through less efficiently than the opal stop codon, this could increase the impact the effect of disrupted stem loop. To study this a sequence of the CHIKV genome containing the stop codon and the RNA structure could be cloned into a bicistronic dual luciferase reporter designed to investigate the efficiency of read-through. To do this the panel of mutants (original, disrupted and compensated) could be inserted between the Rluc and Fluc reporters of the pStopGoDluc plasmid to measure percentage read-through based on the ratio of successful Fluc expression/Rluc expression. This plasmid contains a “StopGo” site located between the two reporters which is cleaved to release the expressed reporters to prevent undesired effect on stability and activity (Loughran et al., 2017). This experiment would reveal differences in read-through efficiency between the mutants and transfecting these plasmids into mosquito or human cells could reveal host specific phenotypes.

## Chapter 5: Developing an affinity tagged nsP3 pulldown system

### 5.1 Introduction

NsP3 is a multifunctional protein which is essential for multiple processes during the virus lifecycle. The exact function of the protein is not fully understood, host and viral factors are recruited by nsP3 to facilitate virus replication during the virus lifecycle. In this chapter I describe the development of an affinity tagged nsP3 pulldown system to allow the identification of interactions with the viral protein which could provide insights into novel functions.

Some functions for CHIKV nsP3 can be extrapolated from related alphavirus research investigating host and viral interactions of nsP3. Purification of a GFP tagged nsP3 expressed by an nsP3-GFP SINV infectious clone revealed a strong interaction with G3BP1/2 which was present at early and late replication stages (Cristea et al., 2006). In host cells G3BP functions in stress granule assembly where it colocalises with other stress granule markers (Irvine et al., 2004). Two conserved FGDF motifs in the nsP3 hypervariable domain bound to the G3BP to block the formation of stress granules during virus infection (Schulte et al., 2016), the protein is also important for the production of negative sense RNA during early stage CHIKV replication (Scholte et al., 2015). The mosquito homologue of G3BP named Rasputin was shown to interact with the nsP3 hypervariable domain and was shown to be important for *in vivo* infection of *Ae. albopictus* (Fros et al., 2015). The CHIKV nsP3 C-terminal contains a proline-rich motif which binds to the Src homology 3 (SH3) domain of amphiphysin-2 (BIN1) to recruit the host protein to the site of virus replication, BIN1 is hypothesised to facilitate replication by potentially stabilising the infectious CPV-I (Tossavainen et al., 2016).

X-ray crystallography was used to determine the structure of the macro domain of CHIKV nsP3 in complex with ADP-ribose and with RNA (Malet et al., 2009). Mutant infectious clones of CHIKV with macro domains incapable of ADP-ribose hydrolysis were unable to replicate (McPherson et al., 2017), additionally the binding of ADP-ribose was shown to function during the initiation of infection as the binding was important for the formation of the replicase (Abraham et al.,

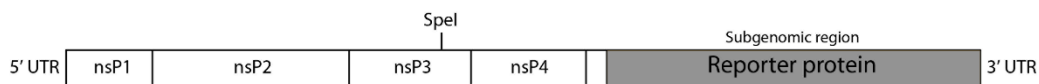
2018). A structure of a fragment of SINV nsP2-nsP3 identified a zinc-binding domain which was predicted as a site for RNA binding in the alphavirus unique domain (AUD) (Shin et al., 2012). The CHIKV AUD was shown to bind to fragments of RNA which corresponded to the (-) 5' UTR, (-) SG promoter and the (+) sense 3' UTR of the CHIKV genome, mutagenesis of the domain revealed the region was essential for CHIKV RNA synthesis (Gao et al., 2019).

In this project in frame affinity tags were inserted into hypervariable domain of nsP3 in the context of CHIKV SGR, a noncytotoxic (NCT) replicon and an infectious clone of CHIKV to develop a system to affinity purify the protein to investigate interactions formed during replication.

## 5.2 Results

### 5.2.1 Tagged nsP3 was successfully expressed using CHIKV SGR

A selection of tags were incorporated into the reading frame of the hypervariable domain of nsP3 at the *SpeI* restriction site. Previous studies have used this site to insert large proteins into the sequence of an infectious clone of SINV, this site was used to insert an Rluc as it was shown to tolerate insertions without attenuating virus replication (Bick et al., 2003). Large inserts are also tolerated at this position, a GFP was inserted into an infectious clone of SINV at the same site (Frolova et al., 2006). A diagram of this insertion site is shown below (**Figure 37**).



**Figure 37: Affinity tags were inserted into the SG Gluc CHIKV SGR using the unique *SpeI* restriction site in the nsP3 HVD.**

The *SpeI* restriction site has been used previously in our lab to successfully incorporate and tolerate a SNAP-tag (NEB) into the coding sequence of nsP3 of a CHIKV SGR replicon. The SNAP-tag is a  $O^6$ -alkylguanine-DNA alkyltransferase that reacts with fluorophores conjugated to benzylguanine, this reaction covalently bonds the fluorophore to the SNAP-tag therefore specifically labelling nsP3 (Remenyi et al., 2017) (Remenyi et al., 2019). The affinity tags chosen in the project were inserted into the nsP3 reading frame of a CHIKV SGR with a Gaussia luciferase (Gluc) reporter under control of the SG promoter, the Gluc signal was measured for these constructs to analyse the effect of these affinity tags on replication. The Gluc is secreted into the cell supernatant which allows for detection of CHIKV replicon replication (Gläsker et al., 2013). A strategy consisting of two sets of tags which follow the Tandem affinity purification (TAP) protocol were initially implemented, double affinity purification results in high specificity.

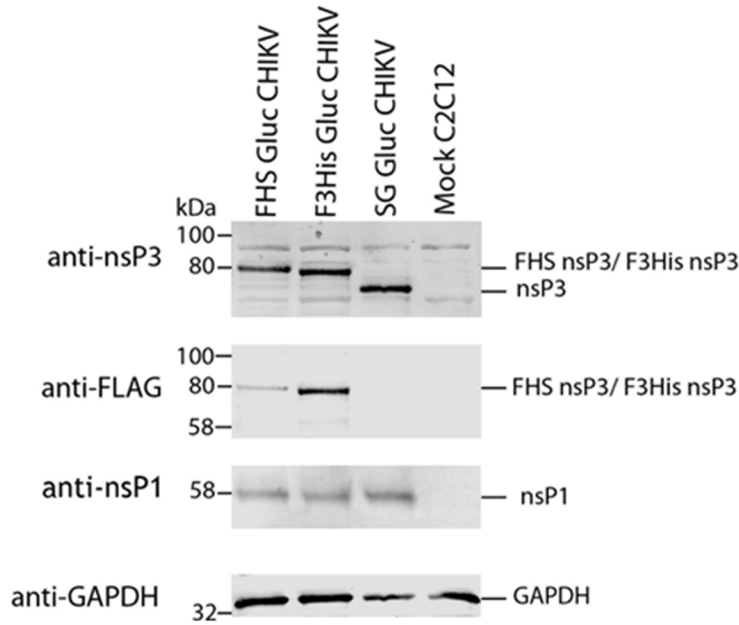


The FLAG, hemagglutinin (HA) and streptavidin binding protein (SBP) were chosen as affinity tag as they have been utilised previously to improve pull down yields. The procedure consists of first purifying the protein using streptavidin, eluting with biotin and then subsequently using an anti-HA antibody column with an acidic pH elution step (Glatter et al., 2009). The FLAG tag was incorporated to provide an additional purification step to increase the purification stringency using an anti-Flag antibody column, free FLAGX3 peptides can be used to elute the protein from the column. An alternative triple FLAG tandem repeats (FLAGX3) and HIS TAP tag was also chosen, the proteins can be purified using the anti-FLAG antibody column and FLAGX3 peptides with subsequent purification with a nickel column (Yang et al., 2006). These affinity tags were inserted into the SG Gluc CHIKV replicon. These will be referred to as FHS or F3His Gluc CHIKV from here onwards (**Figure 38a**). The FHS and F3His Gluc CHIKV constructs were transfected into C2C12 cells, the protein lysates were lysed using Glasgow lysis buffer (GLB) after 24 h.p.t. The samples were Western blotted using antibodies specific for nsP3, FLAG tag and GAPDH (**Figure 38b**). The supernatant was collected for the Gluc assay for these samples (**Figure 38c**).

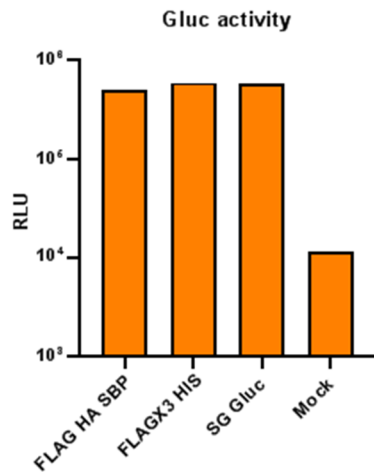
a)



b)



c)



**Figure 38: Tagged nsP3 was successfully expressed when incorporated into the coding sequence of nsP3 in the SG CHIKV Gluc replicon.**

(a) A diagram depicting the insertion of the FHS and F3His tags into the reading frame of nsP3 for the SG Gluc CHIKV replicon. (b) Western blots of protein lysates of C2C12 cells transfected with the CHIKV SGRs using antibodies specific for nsP3, Flag, nsP1 and GAPDH. (c) Gluc activity of the supernatant of the transfected C2C12 cells collected 24 h.p.t was measured. N= 1.

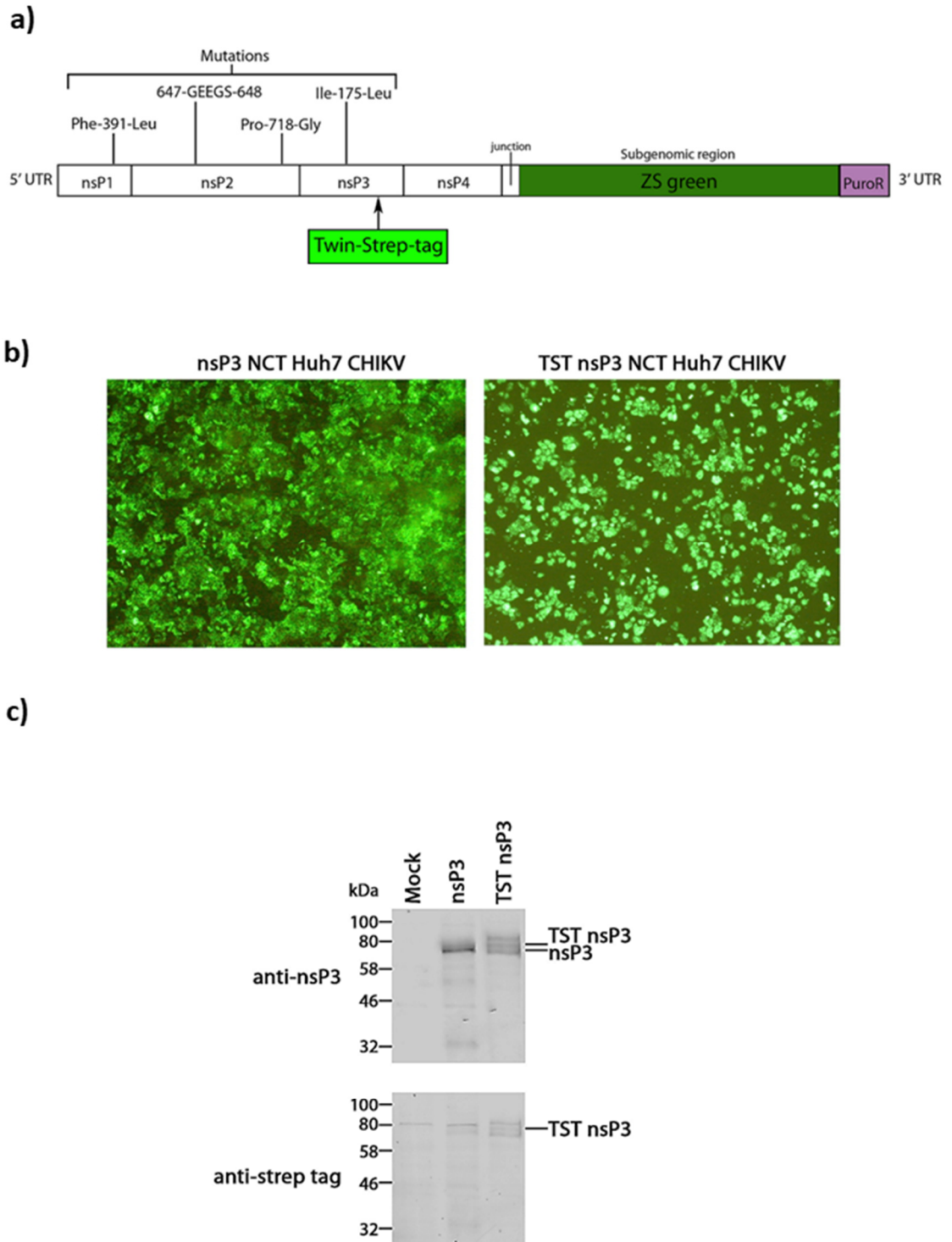
Tagged nsP3 was successfully expressed in C2C12 transfected with the CHIKV SGRs. Western blots confirmed that incorporating these tags resulted in increased molecular weight of nsP3 as seen by the apparent molecular weight shift of nsP3, the FLAG antibody signal for the F3His nsP3 was more prominent than the singly tagged FHS nsP3 (**Figure 38b**). Consistent levels of nsP1 was expressed by the three different CHIKV SGRs, this suggests that the tags did not reduce the expression of the other non-structural proteins. The Gluc activity of the FHS and F3 His Gluc supernatants was at wildtype SG Gluc levels, this suggests that the incorporated tags did not reduce the replication of the replicon (**Figure 38c**). These relatively large insertions into the nsP3 coding sequence were well tolerated, this result suggested that different tags could be inserted into the nsP3 at this region without reducing CHIKV replication. Alternative strategies for nsP3 expression and different affinity tags were tested to optimise the expression protocol.

### **5.2.2 Untagged and TST nsP3 was expressed using a non-cytotoxic replicon in Huh7 cells**

Stable cell lines containing alphavirus replicons which can replicate without inducing cytotoxicity in different cell lines have been developed as a means of studying the nsPs. Two mutations in nsP2 of the SINV replicon result in a non-cytotoxic (NCT) replication phenotype in BHK cells, these were Leu 726 to Pro or a Lys 779 to Asp. These mutations were hypothesised to reduce the amount of toxic viral components in a cell by altering the levels of protein synthesis and by increasing the turnover of these products (Frolov et al., 1999).

A CHIKV NCT replicon containing a Rluc-nsP3, a ZS green reporter and a puromycin acetyltransferase which persistently replicated in Huh7 cells was previously developed, the mutations which permitted this were the Phe 391 to Leu in nsP1, a Gly-Glu-Glu-Gly-Ser insert between position 647 and 648 of nsP2, Pro 718 to Gly in nsP2 and an Ile 175 to Leu in nsP3. These mutations reduce RNA replication and could potentially stabilise viral replication complexes (Utt et al., 2015).

The streptag is an 8 amino acid long peptide which binds to the biotin binding pocket of streptavidin, in this study a tandem repeat of the Streptag named the Twin-Streptag (TST) was used. This TST was incorporated into the NCT CHIKV replicon to express the tagged viral protein in large amount, the TST was chosen as use of this affinity tag was optimised in the lab for purification of NS5A of HCV (Ross-Thriepland and Harris, 2014). An untagged nsP3 variant of the NCT CHIKV replicon was also generated. Diagrams of these nsP3 and TST nsP3 NCT CHIKV replicons can be seen below (**Figure 39a**). The NCT CHIKV constructs were transfected into Huh7 cells, these cells were selected using puromycin until all cells were expressing ZS green. The ZS green reporter was used as a marker for Huh7 cells expressing these constructs (**Figure 39b**). These cells were lysed using GLB, the protein lysates were analysed using Western blots with antibodies targeting the nsP3 and strep tag (**Figure 39c**).

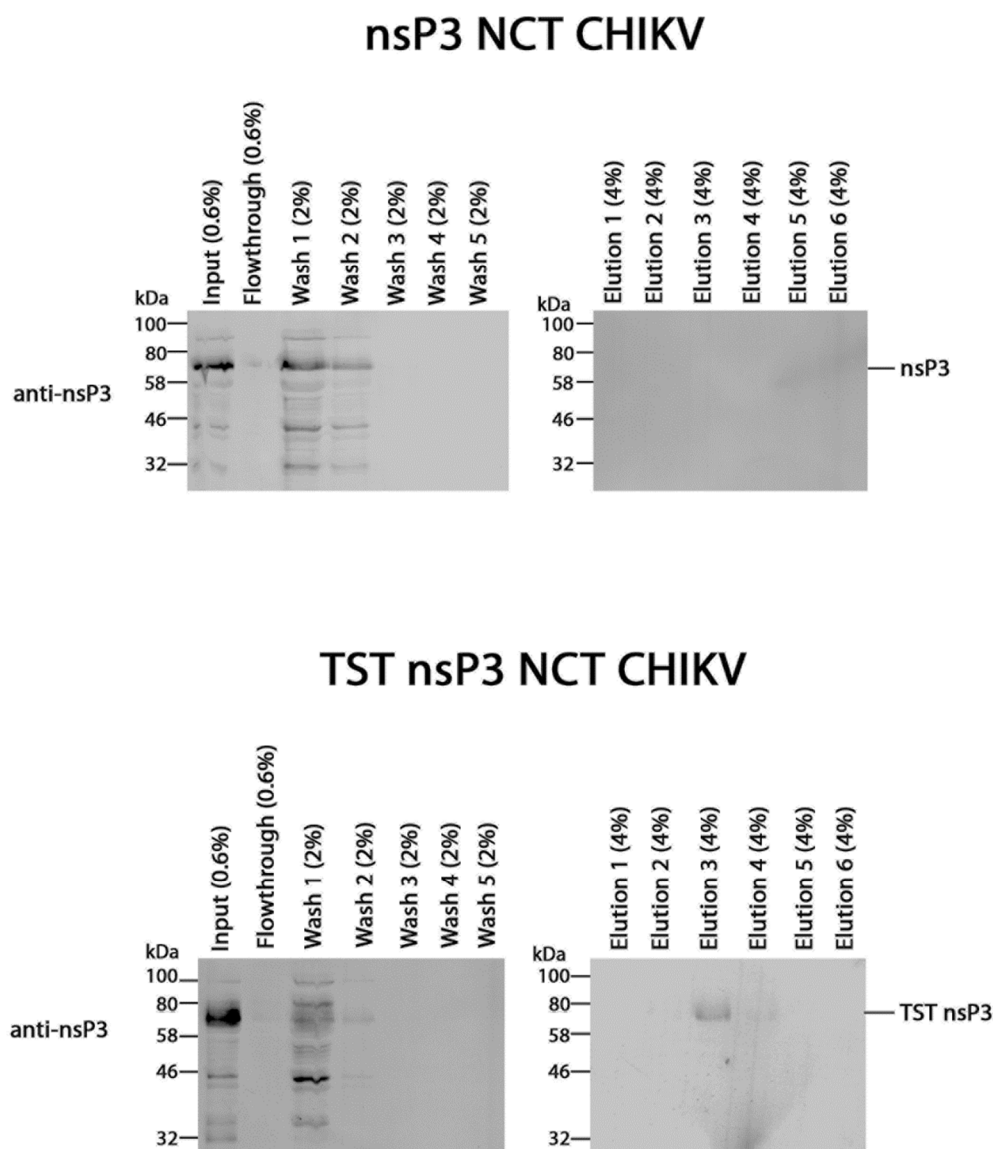


**Figure 39: Stable cell line expressed CHIKV replicons in Huh7 cells produce large amounts of nsP3.**

(a) Diagram showing the untagged and TST nsP3 NCT CHIKV replicon. (b) ZS green signal of Huh7 expressing the NCT CHIKV replicons was detected to confirm successful expression. (c) Western blot of the nsP3 and TST nsP3 using antibodies specific to nsP3 and strep tag. N=2.

The NCT CHIKV replicons contain mutations to facilitate persistent replication in Huh7 cells, replicons containing the untagged and TST nsP3 were generated (**Figure 39a**). A large population of the NCT CHIKV replicon Huh7 cells expressed ZS green, this suggests that the replicon was expressed by most of the cell population (**Figure 39b**). The resultant protein lysates contained large amounts of nsP3 as expected for a stable cell line, bands between 58 and 32 kDa which may correspond to degraded nsP3 can be seen in both stable cell line samples (**Figure 39c**). The highest band for the TST nsP3 would correspond to the increased weight caused by the additional affinity tag, the two lower bands may be degraded forms of the tagged protein.

To validate the use of the TST to purify nsP3 these samples were column purified using Strep-tactin gravity flow columns, 1 mg of clarified protein lysates in a total volume made up to 1 ml using lysis buffer was added to the affinity column. The flowthrough was collected and the column was washed five times using the commercial wash buffer. The samples were eluted using six applications of 500 µl elution buffer which contains 2.5 mM desthiobiotin to compete for the same binding sites as the TST. A Western blot of these samples probed using antibodies targeting nsP3 can be seen below (**Figure 40**).



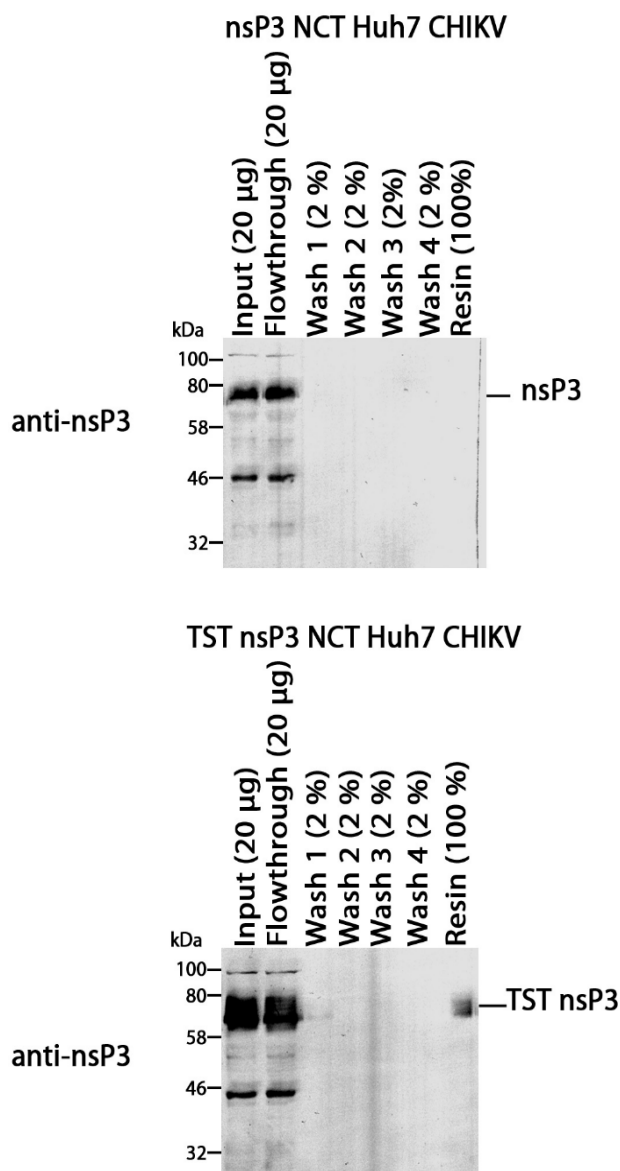
**Figure 40: TST tagged nsP3 can be affinity purified from cell lysates.**

Gravity flow streptactin columns were used to purify the protein lysates isolated from the untagged and TST nsP3 NCT CHIKV Huh7 cell lines. Equivalent amounts of the input and flowthrough were analysed. A proportion of each sample was analysed using Western blot with antibodies specific to nsP3. N=1.

The TST-nsP3 expressed in the NCT CHIKV stable cell lines was specifically purified using the streptactin columns (**Figure 40**). TST nsP3 was released from the column in elution step 3 and 4 but the untagged nsP3 was not detected in any elution step, this suggests that the TST nsP3 binding to the column was specific and could be eluted by dissociating the affinity tag from the binding sites. A batch purification pulldown method was previously developed in the lab to purify TST-

NS5A to identify protein partners using liquid chromatography-tandem mass spectrometry (Goonawardane et al., 2017). The same pulldown method was applied to the TST-nsP3 to establish similar procedures to investigate protein interactions with the viral protein, this method involves incubating the protein lysates directly with the 100  $\mu$ l of 50% streptactin sepharose resin (IBA lifesciences). The resin was prepared by washing the resin twice with 1 ml wash buffer and once with the lysis buffer, 1 mg of protein lysates in a total volume of 500  $\mu$ l of lysis buffer were incubated with resin overnight at 4°C. The flowthrough was collected and resin was washed four times with 1 ml of wash buffer. Proteins were released from resin by boiling the sample in SDS PAGE running buffer for analysis using western blot. The samples were blotted using antibodies specific to nsP3 (**Figure 41**).





**Figure 41: TST nsP3 was successfully purified using the batch purification protocol.**

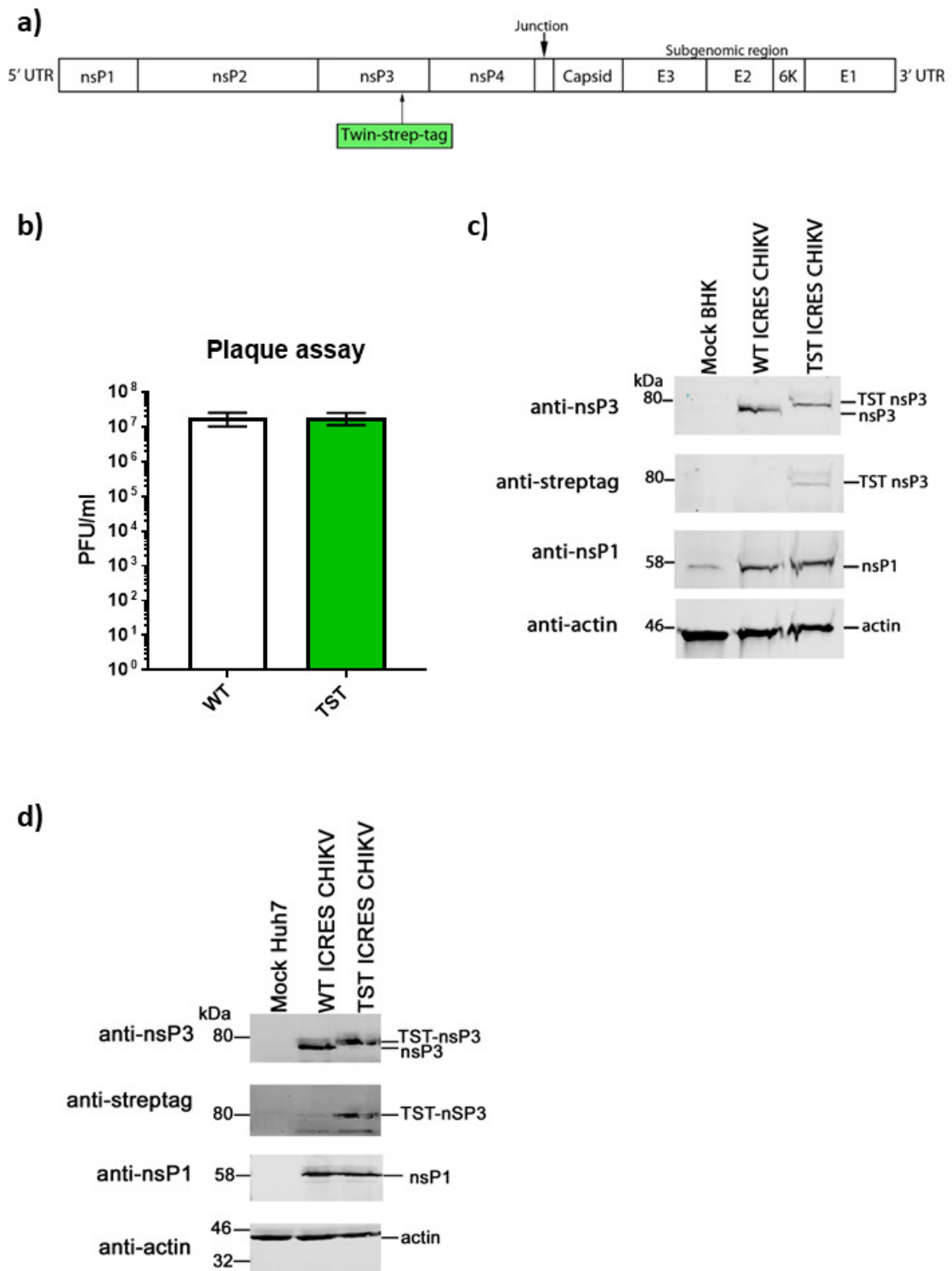
Streptactin sepharose resin was used to purify the untagged and TST nsP3 NCT CHIKV Huh7 protein lysates. An equal amount of the input and flowthrough was analysed. A proportion of each sample was loaded, the samples were western blotted using antibodies specific to nsP3. N=2.

Batch purification using the streptactin resin successfully isolated the TST nsP3 from the NCT CHIKV stable cell line (**Figure 41**). The untagged nsP3 was not seen in the boiled resin sample but the TST nsP3 was released. This suggests that boiling the resin released the protein from the streptactin binding sites. The expression of TST nsP3 in the NCT CHIKV Huh7 cells and successful column

and batch purification validated that the tag could be used to purify the nsP3. The TST was next incorporated into the reading frame of an infectious clone of CHIKV.

### **5.2.3 The TST was successfully incorporated into the reading frame of nsP3 of an infectious clone of CHIKV**

So far, the affinity tags were inserted into nsP3 for CHIKV SGRs but we also wanted to develop an infectious clone containing this tag. Although the SGRs are a useful tool to study CHIKV replication they lack the structural proteins so any nsP3 to structural protein interactions would be unavailable in this system. An infectious clone of CHIKV containing the TST tagged nsP3 was developed to investigate potential interactions formed with the viral protein during the full virus lifecycle. In this context the full set of viral proteins, the non-structural and structural, would be able to interact with nsP3. The TST was inserted into the ICRES CHIKV infectious clone, the affinity tag was inserted into nsP3 reading frame at the same position as the one used for the CHIKV SGR (**Figure 42a**). Virus stocks of the WT ICRES CHIKV and the TST ICRES CHIKV was produced by transfecting BHK cells using *in vitro* transcribed RNA. The supernatant was isolated at 48 h.p.t and titred by plaque assay on BHK cells (**Figure 42b**). The transfected BHK cells were lysed using GLB, these samples were analysed with western blot using antibodies targeting nsP3, streptag, nsP1 and actin (**Figure 42c**). One advantage of creating an infectious clone of CHIKV containing a tagged nsP3 is that it allows researchers to use the virus to infect a range of cell types. The virus stocks produced by the transfected BHK cells was used to infect the Huh7 human cell line at a multiplicity of infection (MOI) of 10, the infected cells were lysed using GLB after 24 hours post infection. These samples were western blotted using antibodies specific to nsP3, streptag, nsP1 and actin (**Figure 42d**).

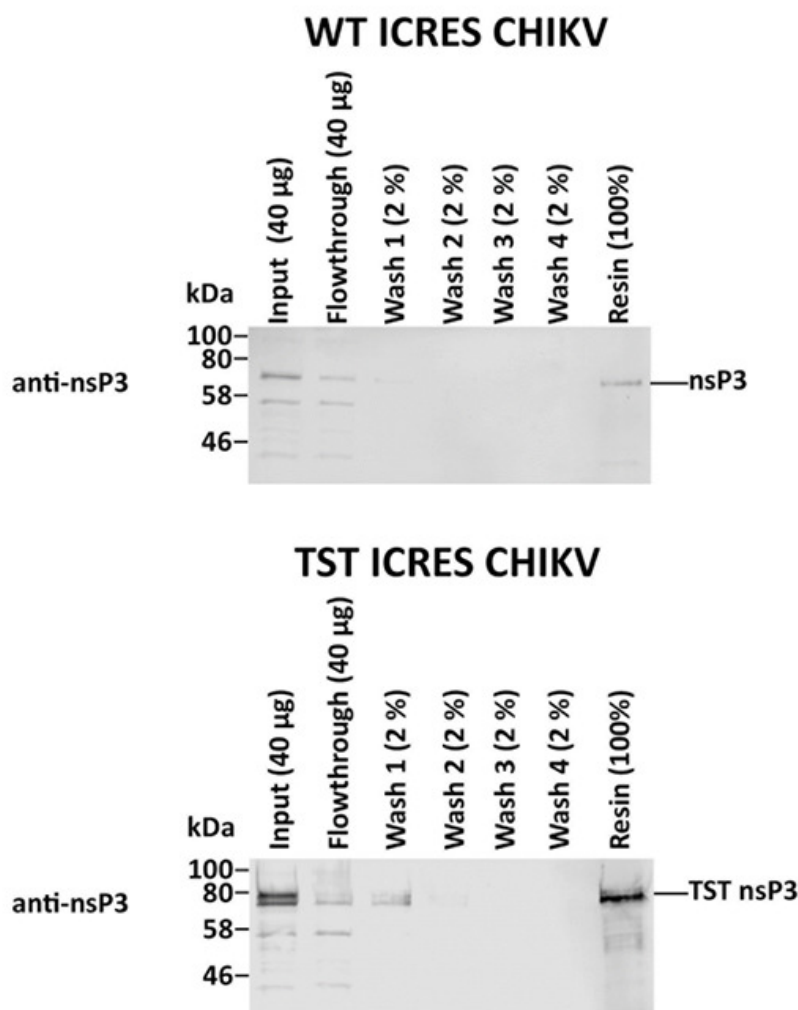


**Figure 42: An ICRES CHIKV expressing a TST nsP3 was successfully generated.**

(a) A schematic of the TST ICRES CHIKV infectious clone. (b) Virus stocks were produced in transfected BHK cells, these samples were harvested 48 hours post transfection. These viruses were plaque assayed on BHK cells. N=2. (c) Western blot of transfected BHK cell lysates with antibodies targeting nsP3, streptag, nsP1 and actin. N=1. (d) Huh7 cells were infected at an MOI of 10, these samples were isolated after 24 hours post infection for Western blot using antibodies targeting nsP3, streptag, nsP1 and actin. N=2.

The TST was successfully inserted into the hypervariable domain of nsP3 of the ICRES CHIKV infectious clone, the viral titre and Western blot images of the TST ICRES CHIKV and WT ICRES CHIKV can be seen in Figure 42. Inserting the TST into nsP3 did not reduce infectivity of the virus, the TST ICRES CHIKV had WT levels of viral titre after harvesting from transfected BHK cells (**Figure 42b**). The western blots of the BHK protein lysates demonstrated that the TST nsP3 could be detected in the protein lysates and that the level of nsP1 expression was the same for both TST and WT ICRES CHIKV (**Figure 42c**). The TST-nsP3 ICRES CHIKV virus was successfully used to infect Huh7 cells, the infected cells produced the tagged nsP3 and nsP1 in the human cell line (**Figure 42d**).

The protein extracts harvested from infected Huh7 cells were used for the batch purification protocol described previously for Figure 41. The streptactin resin was used to purify 1 mg of the WT ICRES CHIKV and TST ICRES CHIKV lysates as described before. The purified protein was released from the resin by boiling with SDS PAGE running buffer. These samples were western blotted with antibodies specific to nsP3 (**Figure 43**).



**Figure 43: TST nsP3 extracted from infected Huh7 cells was purified by batch purification using streptactin resin.**

Streptactin resin was used to purify the protein lysates harvested from infected Huh7 cells, an equivalent amount of protein for the input and flowthrough was analysed. 2 % of each wash and 100 % of the boiled sample was western blotted using antibodies specific to nsP3. N=1.

TST nsP3 expressed in Huh7 cells infected by the TST ICRES CHIKV was enriched during batch purification using streptactin resin (**Figure 43**). A faint band on untagged nsP3 in the resin sample for the WT ICRES CHIKV was seen. This suggests that a small amount of nsP3 had bound to the resin non-specifically but a stronger enrichment of TST nsP3 was seen in the corresponding resin sample. These results presented demonstrated that the TST was successfully inserted into the coding region of nsP3 of the ICRES CHIKV infectious clone. The resultant affinity tagged nsP3 expressed in a human cell line was specifically purified using the batch purification protocol utilised in this research, this result suggests that the ICRES CHIKV will be a valuable tool for studying the interactions formed with nsP3 in different cell types.

### 5.3 Discussion

The developed affinity tagged nsP3 tools will be useful for future studies undertaken by the lab. The TST affinity tag was taken forward due to the small molecular weight of the insert and because previous studies in the lab had optimised the associated purification protocols. Both the NCT cell line containing the TST nsP3 (**Figure 39**) and the TST ICRES CHIKV infectious clone (**Figure 42**) expressed large amounts of tagged nsP3 which could be purified using batch purification (**Figure 41 and Figure 43**).

An infectious clone of CHIKV which contains an affinity tagged nsP3 is particularly useful as this virus can be used to study the viral protein in the context of the full virus lifecycle. Inserting the TST into the reading frame of nsP3 did not reduce virus replication as the TST ICRES CHIKV was shown to produce wildtype levels of virus post electroporation into BHK cells (**Figure 42b**). The TST ICRES CHIKV was then used to successfully infect the human Huh7 cell line (**Figure 42d**), this is one of the model cell lines chosen during previous work in our lab defining recommended systems for studying CHIKV (Roberts et al., 2017). The TST ICRES CHIKV can be used to infect different cell types to investigate cell type specific host protein interactions with nsP3, it would be of interest to investigate the interactions formed in mosquito cells. The mosquito protein Rasputin which is a homologue of the mammalian G3BP was detected previously (Fros et al., 2015), similar essential host protein interactions may identified for CHIKV.

The TST ICRES CHIKV has been used in work published by our lab; the RNA binding activity of nsP3 and nsP3 mutants with AUD mutations was investigated by batch purification of the tagged and untagged nsP3 followed by qRT-PCR to demonstrate differences in RNA binding activity caused by disrupting the AUD (Gao et al., 2019). Chloride channel function was found to promote virus replication, a pulldown experiment of TST nsP3 was used to confirm an interaction between the viral protein and the chloride intracellular channels 1 and 4 (CLIC1 and CLIC4) (Müller et al., 2019). Using the TST nsP3 ICRES CHIKV in future studies to investigate other host and viral interactions with the protein can provide insights into novel functions of the protein.

## **Chapter 6: Conclusion and future perspectives**

The alphavirus CHIKV causes a variety of debilitating symptoms in humans such as muscle pain, joint pain and fever, currently there are no licensed antivirals or vaccines to treat the disease. In depth knowledge of the CHIKV lifecycle is essential for developing new antivirals and vaccines to target the infection as this will reveal drug target candidates. Many viruses utilise stop codon read-through mechanisms to expand the coding potential of the relatively small viral genomes, most alphaviruses possess a stop codon in the nsP3 gene to modulate the expression of the nsP4 RdRp. The alphavirus nsP3 is an essential component of the replication complex but many aspects of the protein's function are still unknown, this project aimed to investigate the opal stop codon found at the C-terminal of the nsP3 gene by determining the phenotype of the stop codon mutations in different cell types and by experimentally determining the conserved stop codon recoding RNA structure in the context of the full virus genome at temperatures corresponding to both human and mosquito body temperature.

### **6.1 Replication of stop codon containing CHIKV mutants displayed a host/vector specific phenotype**

The opal, ochre and amber stop codons were introduced into a D1uc CHIKV SGR derived from an ECSA strain of the virus, this strain naturally possesses an arginine codon in place of the stop codon. These constructs were screened using a panel of mammalian cells and the C6/36 *Ae. albopictus* mosquito cell line to investigate cell type specific CHIKV replication phenotypes. The opal stop codon D1uc CHIKV SGR replicated to wildtype levels in the C2C12 and Huh7 cell types by 24 h.p.t but showed a slightly decreased replication in the BHK, SVG-A and RD cells. Replacing the stop codon with the amber stop codon resulted in the same levels of replication as the opal stop codon in the different mammalian cell types by 24 h.p.t. In contrast the presence of the ochre stop codon caused a significant reduction in replication at 24 h.p.t in all mammalian cell types tested in this study. The ochre stop codon has the highest termination fidelity, amber has similar efficiency, but the opal stop codon has the lowest fidelity as a terminator. This is thought to be due to differences in eRF1 recognition, the interaction between the opal stop codon and eRF1 forms more slowly upon GTP hydrolysis compared to the other stop codons, this allows increased nc-tRNA incorporation (Cridge et al., 2018). In addition to this eukaryotic genes preferentially terminate

with a ochre stop codon followed by a purine (Brown et al., 1990), perhaps the ochre stop codon interacts with the highest affinity with the eRF1 and so is less conducive for nc-tRNA recoding. As the ochre stop codon is the most effective termination signal this could explain why the mutation results in the most drastic effect on CHIKV replication; inefficient read-through during polyprotein expression would result in less nsP4 which would influence the amount of replication complexes present in a host cell. A certain threshold of required stop codon read-through may need to be achieved for CHIKV replication, this may be more readily achieved for the opal and amber stop codon which are more permissive of the read-through signals.

The stop codon mutations were inserted into the ICRES infectious clone of CHIKV to study how stop codon read-through effected CHIKV replication in the full virus lifecycle. The phenotypes observed for the stop codon ICRES mutants harvested from electroporated BHK cells 24 h.p.t were consistent with the phenotypes seen for the stop codon SGRs in the same cell type; both the opal and amber stop codon ICRES CHIKV clone generated similar slightly decreased titres of virus however the ochre stop codon mutant produced undetectable levels of virus at this timepoint. This result concurs with the hypothesis that the inefficient ochre stop codon read-through would impact the expression of the replication complex which may cause detrimental effects on CHIKV replication. Harvesting the virus stocks 48 h.p.t from BHK cells reveals that both the opal and amber stop codon ICRES mutants reach wildtype levels of virus titre at this point and intriguingly the ochre stop codon mutant recovers to higher than wildtype levels of titre. An ICA for these viruses revealed that the ochre stop codon mutant in particular displayed a severely reduced RNA infectivity but could replicate at wildtype titres, taken together these results suggest that the ochre stop codon could potentially revert to recover viability. Further sequencing experiments of the ochre stop codon mutant will reveal if the stop codon mutates to a different codon to recover. Interestingly substituting the opal stop codon to either amber or the ochre stop codon in an infectious clone of SINV caused the related alphavirus to grow slower than the parental strain, the former ochre stop codon caused the most significant reduction in growth (Li and Rice, 1989). The opal stop codon in this position may be conserved over the other stop codons to ensure the correct level of read-through for virus replication. This hypothesis appears to be the case



for the ochre stop codon however some exceptions may exist, for example the amber stop codon mutant replicated to similar levels as the opal stop codon CHIKV mutants seen in this study. A theory behind the bias to preserving the opal stop codon over the other stop codons amongst the alphavirus genomes could however originate from the efficiency of switching between the arginine codon to the opal codon; one base substitution is needed to go from Arg (AGA or CGA) to opal stop codon (UGA) but to change to the amber stop codon would require three mutations (UAG).

The C6/36 mosquito cell line appeared to tolerate the presence of all three stop codons during CHIKV replication. This was seen for both the DLuc SGR containing the mutations as all three stop codons replicated to wildtype levels at all timepoints post transfection and for all three stop codon ICRES mutants. The stop codon ICRES CHIKV mutants produced increased titres in the transfected C6/36 *Ae. albopictus* mosquito cell line. The opal stop codon produced wildtype levels of virus, the amber stop codon produced just below wildtype levels and the ochre stop codon produced a detectable level of virus which was still lower than the other three viruses. This increased virus titre in the mosquito cells suggests that the presence of the opal stop codon in CHIKV promotes virus replication in the vector but not the mammalian host. This result corresponds with findings from previous studies where the presence of the opal stop codon in CHIKV strains enhanced replication in C6/36 cells above the levels seen for the arginine coding viruses (Jones et al., 2017) (Mounce et al., 2017). It would be beneficial to evaluate the differences seen between electroporation and lipofectamine 2000 transfection, to do this BHK and C6/36 cells should be transfected with the infectious clone ICRES RNA with the corresponding techniques and the harvest of virus stocks performed at the same timepoint to compare viral titres. These results would confirm that the two transfection methods did not influence virus replication differently during this study

A similar enhanced vector specific replication facilitated by the opal stop codon was seen for the related ONNV alphavirus. When the opal or amber containing ONNV was transfected into BHK cells this resulted in replication at lower levels than the arginine strain, the ochre stop codon ONNV mutant produced undetectable titres of the virus however in C6/36 cells the opal and amber stop codons demonstrated increased replication and the ochre mutant produced a

small amount of detectable virus. It was also found that the presence of the opal stop codon but not the amber stop codon enhanced virus infectivity *in vivo* in *A. gambiae* mosquitoes compared to the arginine encoding virus (Myles et al., 2006).

How the opal stop codon enhances vector infectivity of alphaviruses is currently unknown, the increased replication of all three stop codons in the mosquito cell line could suggest that read-through is enhanced in the host. The translation machinery in insects is similar to that found in other eukaryotic hosts, the *Drosophila melanogaster* genome was found to express homologs of all the eukaryotic translation factors found in the canonical translation pathway such as the eRF1 and eIF3 which are involved during stop codon read-through (Lasko, 2000) (Marygold et al., 2017). The similarities between the insect and human translation machinery suggests that the read-through mechanism is likely the same in the different species, the differences in read-through potential observed in the different hosts may depend on currently unidentified factors.

## **6.2 The alphavirus stop codon recoding RNA structure folds at temperatures corresponding with the host and vector body temperature**

Stop codon read-through in alphaviruses is promoted by the cytosine nucleotide immediately downstream of the opal stop codon and a conserved downstream large RNA structure, the CHIKV RNA structure was recently experimentally determined in the context of a Dluc bicistronic replicon using SHAPE with radioactive primers (Kendra et al., 2018). This study builds upon this result by using QuSHAPE to experimentally validate the stop codon recoding RNA structure using the full length RNA genome to identify if the presence of other sections of the sequence could influence the structure. Uncapped *in vitro* transcribed RNA which did not possess a poly-(A) tail was used during these shape experiments due to cost however it could be useful to determine the effect that these features could have on RNA structure as they may be involved in RNA folding. It would be interesting to perform QuSHAPE experiments on capped and poly adenylated RNA samples as a comparison to the structures presented in this study.

The stop codon recoding RNA structure experimentally determined for the RNA folded at 37°C was similar to the structure proposed by (Kendra et al., 2018) therefore confirming the accuracy of the QuSHAPE experimental procedure performed in this study. The conserved region matching the bottom stem of the structure proposed by (Firth et al., 2011) was unreactive to NMIA (base-paired) as expected, the majority of the SHAPE data for the remainder of the RNA structure matched the Mfold predicted model generated for this region. These findings suggest that presence of the full CHIKV genome did not alter the expected structure of the RNA stem.

In addition to this, the formation of the RNA structure at 28°C was investigated, this temperature matched the lower body temperature of the invertebrate host. It was hypothesised that the difference in temperature could alter the folding of the RNA structure within the two hosts, differences in stop codon read-through structure could explain the vertebrate and invertebrate specific replication phenotypes seen in this study. Folding the RNA at 28°C caused some small changes but the overall structure closely resembled the structure seen for the 37°C folded RNA, this result suggests that the stop codon recoding structure can form at the mosquito body temperature. Little is known about the effect of the RNA structure on CHIKV replication in both the human host and the mosquito vector, to investigate this a panel of mutants containing substitutions disrupting the bottom RNA structure without changing the amino acid sequence were generated. Some preliminary experiments investigating the phenotype of these mutants in BHK cells and C6/36 cells were performed in this study. In BHK cells it was found that the disrupted stem mutants caused a loss of virus titre and RNA infectivity depending on the stop codon, compensating the stem recovered replication for all three stop codons. In contrast, in C6/36 cells all disrupted and compensated mutants replicated to levels comparable to the original stop codon mutants. This result implied that changing this RNA structure did not influence stop codon read-through in mosquito cells. It should be noted that the limited repeats of the mutated RNA structure experiments mean that these results should be taken as only a guide to potentially interesting phenotypes to investigate in future studies. The mechanism behind this phenomenon is currently unclear but it is possible that the RNA structure is more important for facilitating read-through in the mammalian cell type but not the mosquito. Perhaps the 3' cytosine

downstream of the opal stop codon is sufficient for read-through in mosquito cells but in mammalian cells both the cytosine and the RNA structure are necessary for stop codon read-through. Alternatively stop codon read-through may be enhanced in the invertebrate mosquito cells, it has been previously found that insect genomes have higher incidences of stop codon read-through events during host protein expression, these insects include *Drosophila* and *A. gambiae* which indirectly suggests that the translation machinery may be more permissive to read-through (Jungreis et al., 2011).

This study is the first to experimentally determine the stop codon recoding RNA structure using QuSHAPE experiments on full CHIKV genomes and the first to investigate the formation of the structure at temperatures corresponding to the mosquito body temperature. This optimised procedure can be used in future projects to investigate this how the disruption and compensatory mutations can influence the RNA structure to confirm the Mfold predicted structures at the two different temperatures. Another experiment that would be of interest would be to investigate the effect of the stop codon and RNA stem mutations on read-through efficiency, to do this a short CHIKV sequence containing the stop codon position and the RNA recoding structure will be inserted between the two luciferase reporters of the StopGo bicistronic reporter plasmid (Loughran et al., 2017). The ratio of read-through for each of these constructs could correspond with the results seen for the SGR and infectious clones of CHIKV containing these mutations. Cell type specific phenotypes of read-through, such as the hypothesised permissive insect cell read-through, could be investigated by screening different cell lines with these constructs.

The mechanism of stop codon recoding RNA structures is currently unknown, it is hypothesised that these large structures could directly interact with the ribosome to stall termination, it could directly interfere with the eRF1 or could interfere with other proteins associated with translation termination (Firth and Brierley, 2012). It is possible that the RNA structure could possess additional unidentified functions to promote stop codon read-through in alphaviruses. RdRp expression for the plant Tombusvirus is controlled by stop codon read-through, a novel long range interaction between the RNA recoding stem loop and RNA structures in the 3' UTR promote stop codon read-through (Cimino et al., 2011). An RNA structure promoting read-through of an opal stop codon in the Colorado

tick fever virus structural protein VP9 was shown to not promote ribosome pausing but this may be unique to the virus (Naphthine et al., 2012). Stalling the ribosome would allow time for the nc-tRNA to compete with eRF1 for the A site of the ribosome, alternatively the RNA structures may sequester the eRF1 to bias read-through over termination at the stop codon.

Additional experiments which measure the effect of the stop codon on translation of the nsP would be beneficial. It would be of great interest to perform ribosome profiling experiments using both the different stop codons and the RNA structure mutants to determine how these components directly influence translation. During translation the ribosome protects a discrete footprint of around 30 nucleotides on the mRNA sequence dependent on the position along the reading frame, analysing these protected sequences using ribosome profiling with deep sequencing will give a quantitative measure of the levels of translation rate and ribosome pausing at certain sections of RNA (Ingolia et al., 2009) (Ingolia et al., 2012). In the context of this project it could reveal potential mechanisms for the phenotypic differences seen between the stop codon mutants. If the different stop codon or RNA structure mutants display altered levels of ribosome pausing at the site of stop codon read-through this may cause differences in the nsP expression which could be influencing virus replication.

### **6.3 An infectious clone of CHIKV possessing an affinity tagged TST-nsP3 was successfully generated**

This is the first study to produce an affinity tagged nsP3 containing infectious clone of CHIKV. The infectious clone of CHIKV containing the TST affinity tagged nsP3 generated in this study will be a useful tool for studying unknown functions of the enigmatic protein. The HVD of alphavirus nsP3 was chosen as the insertion site for the affinity tag as insertions into the SINV genome at this position were well tolerated, these insertions include Rluc (Bick et al., 2003) and GFP (Frolova et al., 2006). Inserting the TST into the nsP3 HVD did not attenuate CHIKV replication, the mutant virus produced wildtype level of titres after 48 h.p.t and induced similar levels of viral protein expression in the different infected cell lines. There are multiple benefits in using the TST-nsP3 ICRES CHIKV to investigate nsP3. The nsP3 is expressed in the context of the full virus lifecycle so potential interactions with structural proteins which would be missed when using an SGR based system can be identified. It is more time efficient and more

cost effective to produce large amounts of virus expressing tagged nsP3 and experiments using the virus can be performed at consistent MOIs. Future studies utilising this tool will include performing pulldown experiments to identify protein and RNA interacting partners of the viral protein, identifying novel interactions will provide insights into previously unknown functions and can generate candidates for antiviral or vaccine development.

#### **6.4 Conclusion and future prospects**

In summary, this study has identified novel findings for replication of CHIKV strains containing different stop codons in both the vertebrate and invertebrate contexts. The results presented here have consolidated the conserved preference of the opal stop codon over the other stop codons during CHIKV replication, the hierarchy of stop codon tolerance during CHIKV replication was opal > amber > ochre. Intriguingly all three stop codons, including the ochre stop codon, were better tolerated in the mosquito cell line. Future studies investigating this apparent increased stop codon read-through potential in insect cells may reveal insights previously observed vertebrate specific preference for strains of alphaviruses containing the opal stop codon. The mechanism which promotes stop codon read-through of the opal stop codon must be able to promote read-through for the other two stop codons during CHIKV replication, replication would be impossible if no nsP4 was expressed due to significantly inefficient read-through. Further research into the deficiencies seen for CHIKV replication for the amber and especially the ochre stop codon containing mutants will be of interest, determining if the levels of CHIKV replication match the results seen for future read-through efficiency studies could provide a mechanism behind the discrepancies.

The stop codon recoding RNA structure was confirmed to form when the full CHIKV genome was folded and this study was the first to confirm that this structure could form at temperatures equivalent to the mosquito body temperature. Folding the RNA at 37°C and 28°C results in a very similar overall structure, this suggests that the RNA stem loop should theoretically function to promote stop codon read-through in both organisms. Intriguingly, the preliminary experiments to disrupt this structure was shown to affect the replication of the stop codon mutants in mammalian cells but not the mosquito cells. These mutants can be used in future experiments to further characterise the role of the

RNA structure for stop codon read-through in the vertebrate and invertebrate host. It is possible that the structure is important for stop codon read-through during translation in mammalian systems but not the mosquito due to differences in read-through mechanism.

In conclusion this study has created a panel of valuable tools and procedures to investigate stop codon read-through and nsP3 protein interacting partners. The structural and functional analysis on stop codon read-through during CHIKV replication has identified a mosquito specific enhancement of all three stop codons, built up further evidence supporting the type 2 stop codon read-through structure conserved across the alphaviruses and established the foundation for future work investigating species specific phenotypes of the RNA structure.

## References

- Abraham, R., Hauer, D., McPherson, R.L., Utt, A., Kirby, I.T., Cohen, M.S., Merits, A., Leung, A.K.L. and Griffin, D.E. 2018. ADP-ribosyl-binding and hydrolase activities of the alphavirus nsP3 macrodomain are critical for initiation of virus replication. *Proceedings of the National Academy of Sciences*. **115**(44), pp.E10457–E10466.
- Ahola, T. and Kääriäinen, L. 1995. Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. *Proceedings of the National Academy of Sciences of the United States of America*. **92**(2), pp.507–511.
- Alam, S.L., Wills, N.M., Ingram, J.A., Atkins, J.F. and Gesteland, R.F. 1999. Structural studies of the RNA pseudoknot required for readthrough of the gag-termination codon of murine leukemia virus. *Journal of Molecular Biology*. **288**(5), pp.837–852.
- Alkalaeva, E.Z., Pisarev, A. V., Frolova, L.Y., Kisselev, L.L. and Pestova, T. V. 2006. In Vitro Reconstitution of Eukaryotic Translation Reveals Cooperativity between Release Factors eRF1 and eRF3. *Cell*. **125**(6), pp.1125–1136.
- Atkins, G.J. 2013. The Pathogenesis of Alphaviruses. *International Scholarly Research Notices Virology*. **2013**, p.22.
- Auld, D.S., Thorne, N., Maguire, W.F. and Inglese, J. 2009. Mechanism of PTC124 activity in cell-based luciferase assays of nonsense codon suppression. *Proceedings of the National Academy of Sciences of the United States of America*. **106**(9), pp.3585–90.
- Barry, G., Fragkoudis, R., Ferguson, M.C., Lulla, A., Merits, A., Kohl, A. and Fazakerley, J.K. 2010. Semliki Forest Virus-Induced Endoplasmic Reticulum Stress Accelerates Apoptotic Death of Mammalian Cells. *Journal of Virology*. **84**(14), pp.7369–7377.
- Basore, K., Kim, A.S., Nelson, C.A., Zhang, R., Smith, B.K., Uranga, C., Vang, L., Cheng, M., Gross, M.L., Smith, J., Diamond, M.S. and Fremont, D.H. 2019. Cryo-EM Structure of Chikungunya Virus in Complex with the Mxra8 Receptor. *Cell*. **177**(7), pp.1725–1737.e16.
- Beier, H. 2001. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic Acids Research*. **29**(23), pp.4767–4782.
- Beier, H. and Grimm, M. 2001. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic acids research*. **29**(23), pp.4767–82.
- Beznosková, P., Gunišová, S. and Valášek, L.S. 2016. Rules of UGA-N decoding by near-cognate tRNAs and analysis of readthrough on short uORFs in yeast. *RNA (New York, N.Y.)*. **22**(3), pp.456–66.
- Beznoskova, P., Wagner, S., Jansen, M.E., von der Haar, T. and Vala ek, L.S. 2015. Translation initiation factor eIF3 promotes programmed stop codon readthrough. *Nucleic Acids Research*. **43**(10), pp.5099–5111.
- Bick, M.J., Carroll, J.-W.N., Gao, G., Goff, S.P., Rice, C.M. and MacDonald, M.R. 2003. Expression of the zinc-finger antiviral protein inhibits alphavirus replication. *Journal of virology*. **77**(21), pp.11555–62.
- Braack, L., Gouveia de Almeida, A.P., Cornel, A.J., Swanepoel, R. and de Jager, C. 2018.



- Mosquito-borne arboviruses of African origin: review of key viruses and vectors. *Parasites & Vectors*. **11**(1), p.29.
- Brackney, D.E., Scott, J.C., Sagawa, F., Woodward, J.E., Miller, N.A., Schilkey, F.D., Mudge, J., Wilusz, J., Olson, K.E., Blair, C.D. and Ebel, G.D. 2010. C6/36 *Aedes albopictus* Cells Have a Dysfunctional Antiviral RNA Interference Response S. L. O'Neill, ed. *PLoS Neglected Tropical Diseases*. **4**(10), p.e856.
- Broeckel, R., Haese, N., Messaoudi, I., Streblow, D., Broeckel, R., Haese, N., Messaoudi, I. and Streblow, D.N. 2015. Nonhuman Primate Models of Chikungunya Virus Infection and Disease (CHIKV NHP Model). *Pathogens*. **4**(3), pp.662–681.
- Brown, A., Shao, S., Murray, J., Hegde, R.S. and Ramakrishnan, V. 2015. Structural basis for stop codon recognition in eukaryotes. *Nature*. **524**(7566), pp.493–496.
- Brown, C.M., Stockwell, P.A., Trotman, C.N. and Tate, W.P. 1990. Sequence analysis suggests that tetra-nucleotides signal the termination of protein synthesis in eukaryotes. *Nucleic acids research*. **18**(21), pp.6339–45.
- Brown, R.S., Wan, J.J. and Kielian, M. 2018. The Alphavirus Exit Pathway: What We Know and What We Wish We Knew. *Viruses*. **10**(2), p.89.
- Bulygin, K.N., Repkova, M.N., Ven'yaminova, A.G., Graifer, D.M., Karpova, G.G., Frolova, L.Y. and Kisselev, L.L. 2002. Positioning of the mRNA stop signal with respect to polypeptide chain release factors and ribosomal proteins in 80S ribosomes. *FEBS Letters*. **514**(1), pp.96–101.
- CDC 2018. Chikungunya virus Geographic Distribution. *Centers for Disease Control and Prevention*. [Online]. [Accessed 18 June 2019]. Available from: <https://www.cdc.gov/chikungunya/geo/index.html>.
- Chen, K.C., Kam, Y.-W., Lin, R.T.P., Ng, M.M.-L., Ng, L.F. and Chu, J.J.H. 2013. Comparative analysis of the genome sequences and replication profiles of chikungunya virus isolates within the East, Central and South African (ECSA) lineage. *Virology journal*. **10**(1), p.169.
- Chen, R., Wang, E., Tsetsarkin, K.A. and Weaver, S.C. 2013. Chikungunya Virus 3' Untranslated Region: Adaptation to Mosquitoes and a Population Bottleneck as Major Evolutionary Forces. *PLoS Pathogens*. **9**(8), p.e1003591.
- Chittum, H.S., Lane, W.S., Carlson, B.A., Roller, P.P., Lung, F.D.T., Lee, B.J. and Hatfield, D.L. 1998. Rabbit  $\beta$ -globin is extended beyond its UGA stop codon by multiple suppressions and translational reading gaps. *Biochemistry*. **37**(31), pp.10866–10870.
- Choi, H.-K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M.G. and Wengler, G. 1991. Structure of Sindbis virus core protein reveals a chymotrypsin-like serine proteinase and the organization of the virion. *Nature*. **354**(6348), pp.37–43.
- Cimino, P.A., Nicholson, B.L., Wu, B., Xu, W. and White, K.A. 2011. Multifaceted Regulation of Translational Readthrough by RNA Replication Elements in a Tombusvirus D. Gallie, ed. *PLoS Pathogens*. **7**(12), p.e1002423.
- Coffey, L., Failloux, A.-B., Weaver, S., Coffey, L.L., Failloux, A.-B. and Weaver, S.C. 2014. Chikungunya Virus–Vector Interactions. *Viruses*. **6**(11), pp.4628–4663.

- Cridge, A.G., Crowe-McAuliffe, C., Mathew, S.F. and Tate, W.P. 2018. Eukaryotic translational termination efficiency is influenced by the 3' nucleotides within the ribosomal mRNA channel. *Nucleic Acids Research*. **46**(4), pp.1927–1944.
- Cristea, I.M., Carroll, J.-W.N., Rout, M.P., Rice, C.M., Chait, B.T. and MacDonald, M.R. 2006. Tracking and Elucidating Alphavirus-Host Protein Interactions. *Journal of Biological Chemistry*. **281**(40), pp.30269–30278.
- Csibra, E., Brierley, I. and Irigoyen, N. 2014. Modulation of stop codon read-through efficiency and its effect on the replication of murine leukemia virus. *Journal of virology*. **88**(18), pp.10364–76.
- Dabrowski, M., Bukowy-Bieryllo, Z. and Zietkiewicz, E. 2015. Translational readthrough potential of natural termination codons in eucaryotes – The impact of RNA sequence. *RNA Biology*. **12**(9), pp.950–958.
- Darty, K., Denise, A. and Ponty, Y. 2009. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*. **25**(15), pp.1974–1975.
- Davis, J.L., Hodge, H.M., Campbell, W.E. and Jr. 1971. Growth of chikungunya virus in baby hamster kidney cell (BHK-21-clone 13) suspension cultures. *Applied microbiology*. **21**(2), pp.338–41.
- 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 of the United States of America*. **106**(1), pp.97–102.
- Dever, T.E. and Green, R. 2012. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harbor perspectives in biology*. **4**(7), p.a013706.
- Diaz-Toledano, R., Lozano, G. and Martinez-Salas, E. 2017. In-cell SHAPE uncovers dynamic interactions between the untranslated regions of the foot-and-mouth disease virus RNA. *Nucleic acids research*. **45**(3), pp.1416–1432.
- Dinman, J.D. 2012. Mechanisms and implications of programmed translational frameshifting. *Wiley Interdisciplinary Reviews: RNA*. **3**(5), pp.661–673.
- Dupuis-Maguiraga, L., Noret, M., Brun, S., Le Grand, R., Gras, G. and Roques, P. 2012. Chikungunya Disease: Infection-Associated Markers from the Acute to the Chronic Phase of Arbovirus-Induced Arthralgia. *PLoS Neglected Tropical Diseases*. **6**(3), p.e1446.
- Eckei, L., Krieg, S., Bütepage, M., Lehmann, A., Gross, A., Lippok, B., Grimm, A.R., Kümmerer, B.M., Rossetti, G., Lüscher, B. and Verheugd, P. 2017. The conserved macrodomains of the non-structural proteins of Chikungunya virus and other pathogenic positive strand RNA viruses function as mono-ADP-ribosylhydrolases. *Scientific Reports*. **7**, p.41746.
- Fayzulin, R. and Frolov, I. 2004. Changes of the Secondary Structure of the 5' End of the Sindbis Virus Genome Inhibit Virus Growth in Mosquito Cells and Lead to Accumulation of Adaptive Mutations. *J Virol*. **78**(10), pp.4953–4964.
- Firth, A.E. and Brierley, I. 2012. Non-canonical translation in RNA viruses. *Journal of General Virology*. **93**(Pt\_7), pp.1385–1409.

- Firth, A.E., Chung, B.Y., Fleeton, M.N. and Atkins, J.F. 2008. Discovery of frameshifting in Alphavirus 6K resolves a 20-year enigma. *Virology Journal*. **5**(1), p.108.
- Firth, A.E., Wills, N.M., Gesteland, R.F. and Atkins, J.F. 2011. Stimulation of stop codon readthrough: frequent presence of an extended 3' RNA structural element. *Nucleic Acids Research*. **39**(15), pp.6679–6691.
- Foy, N.J., Akhrymuk, M., Akhrymuk, I., Atasheva, S., Bopda-Waffo, A., Frolov, I. and Frolova, E.I. 2013. Hypervariable Domains of nsP3 Proteins of New World and Old World Alphaviruses Mediate Formation of Distinct, Virus-Specific Protein Complexes. *J Virol*. **87**(4), pp.1997–2010.
- Frolov, I., Agapov, E., Hoffman, T.A., Pragai, B.M., Lippa, M., Schlesinger, S. and Rice, C.M. 1999. Selection of RNA Replicons Capable of Persistent Noncytopathic Replication in Mammalian Cells. *J. Virol*. **73**(5), pp.3854–3865.
- Frolova, E., Gorchakov, R., Garmashova, N., Atasheva, S., Vergara, L.A. and Frolov, I. 2006. Formation of nsP3-Specific Protein Complexes during Sindbis Virus Replication. *J Virol*. **80**(8), pp.4122–4134.
- Fros, J.J., Geertsema, C., Zouache, K., Baggen, J., Domeradzka, N., van Leeuwen, D.M., Flipse, J., Vlak, J.M., Failloux, A.-B. and Pijlman, G.P. 2015. Mosquito Rasputin interacts with chikungunya virus nsP3 and determines the infection rate in *Aedes albopictus*. *Parasites & Vectors*. **8**, p.464.
- Gao, Y., Goonawardane, N., Ward, J., Tuplin, A. and Harris, M. 2019. Multiple roles of the non-structural protein 3 (nsP3) alphavirus unique domain (AUD) during Chikungunya virus genome replication and transcription R. J. Kuhn, ed. *PLOS Pathogens*. **15**(1), p.e1007239.
- Gardner, J., Anraku, I., Le, T.T., Larcher, T., Major, L., Roques, P., Schroder, W.A., Higgs, S. and Suhrbier, A. 2010. Chikungunya virus arthritis in adult wild-type mice. *Journal of virology*. **84**(16), pp.8021–32.
- Geigenmüller-Gnirke, U., Nitschko, H. and Schlesinger, S. 1993. Deletion analysis of the capsid protein of Sindbis virus: identification of the RNA binding region. *Journal of virology*. **67**(3), pp.1620–6.
- Des Georges, A., Hashem, Y., Unbehaun, A., Grassucci, R.A., Taylor, D., Hellen, C.U.T., Pestova, T. V. and Frank, J. 2014. Structure of the mammalian ribosomal pre-termination complex associated with eRF1•eRF3•GDPNP. *Nucleic Acids Research*. **42**(5), pp.3409–3418.
- Gibbons, D.L., Vaney, M.-C., Rousset, A., Vigouroux, A., Reilly, B., Lepault, J., Kielian, M. and Rey, F.A. 2004. Conformational change and protein–protein interactions of the fusion protein of Semliki Forest virus. *Nature*. **427**(6972), pp.320–325.
- Gläser, S., Lulla, A., Lulla, V., Couderc, T., Drexler, J.F., Liljeström, P., Lecuit, M., Drosten, C., Merits, A. and Kümmerer, B.M. 2013. Virus replicon particle based Chikungunya virus neutralization assay using *Gussia luciferase* as readout. *Virology Journal*. **10**, p.235.
- Glatzer, T., Wepf, A., Aebersold, R. and Gstaiger, M. 2009. An integrated workflow for charting the human interaction proteome: insights into the PP2A system. *Molecular*

*systems biology*. **5**, p.237.

- Goff, S.P. and Green, L. 2015. Translational readthrough-promoting drugs enhance pseudoknot-mediated suppression of the stop codon at the Moloney murine leukemia virus gag-pol junction. *Journal of General Virology*. **96**(11), pp.3411–3421.
- Gomez de Cedrón, M., Ehsani, N., Mikkola, M.L., García, J.A. and Kääriäinen, L. 1999. RNA helicase activity of Semliki Forest virus replicase protein NSP2. *Federation of European Biochemical Societies Letters*. **448**(1), pp.19–22.
- Goonawardane, N., Gebhardt, A., Bartlett, C., Pichlmair, A. and Harris, M. 2017. Phosphorylation of Serine 225 in Hepatitis C Virus NS5A Regulates Protein-Protein Interactions. *Journal of Virology*. **91**(17), pp.e00805-17.
- Götte, B., Liu, L. and McInerney, G. 2018. The Enigmatic Alphavirus Non-Structural Protein 3 (nsP3) Revealing Its Secrets at Last. *Viruses*. **10**(3), p.105.
- Goupil, B.A. and Mores, C.N. 2016. A Review of Chikungunya Virus-induced Arthralgia: Clinical Manifestations, Therapeutics, and Pathogenesis. *The open rheumatology journal*. **10**, pp.129–140.
- De Groot, R.J., Rumenapf, T., Kuhn, R.J., Strauss, E.G. and Strauss, J.H. 1991. Sindbis virus RNA polymerase is degraded by the N-end rule pathway. *Proceedings of the National Academy of Sciences of the United States of America*. **88**(20), pp.8967–8971.
- Hardy, R.W. and Rice, C.M. 2005. Requirements at the 3' End of the Sindbis Virus Genome for Efficient Synthesis of Minus-Strand RNA. *Journal of Virology*. **79**(8), pp.4630–4639.
- Hawman, D.W., Stoermer, K.A., Montgomery, S.A., Pal, P., Oko, L., Diamond, M.S. and Morrison, T.E. 2013. Chronic joint disease caused by persistent Chikungunya virus infection is controlled by the adaptive immune response. *Journal of virology*. **87**(24), pp.13878–13888.
- Helenius, A. 1995. *Alphavirus and Flavivirus Glycoproteins: Structures and Functions Minireview* [Online]. [Accessed 25 June 2019]. Available from: [https://www.cell.com/cell/pdf/0092-8674\(95\)90523-5.pdf](https://www.cell.com/cell/pdf/0092-8674(95)90523-5.pdf).
- Hinson, A.R.P., Jones, R., Crose, L.E.S., Belyea, B.C., Barr, F.G. and Linardic, C.M. 2013. Human rhabdomyosarcoma cell lines for rhabdomyosarcoma research: utility and pitfalls. *Frontiers in oncology*. **3**, p.183.
- Hong, E.M., Perera, R. and Kuhn, R.J. 2006. Alphavirus Capsid Protein Helix I Controls a Checkpoint in Nucleocapsid Core Assembly. *JOURNAL OF VIROLOGY*. **80**(18), pp.8848–8855.
- Hwang Kim, K., Rumenapf, T., Strauss, E.G. and Strauss, J.H. 2004. Regulation of Semliki Forest virus RNA replication: a model for the control of alphavirus pathogenesis in invertebrate hosts. *Virology*. **323**(1), pp.153–163.
- Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M. and Weissman, J.S. 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols*. **7**(8), pp.1534–1550.

- Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S. and Weissman, J.S. 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science*. **324**(5924), pp.218–223.
- Irvine, K., Stirling, R., Hume, D. and Kennedy, D. 2004. Rasputin, more promiscuous than ever: a review of G3BP. *The International journal of developmental biology*. **48**(10), pp.1065–77.
- Ivanov, A., Mikhailova, T., Eliseev, B., Yeramala, L., Sokolova, E., Susorov, D., Shuvalov, A., Schaffitzel, C. and Alkalaeva, E. 2016. PABP enhances release factor recruitment and stop codon recognition during translation termination. *Nucleic Acids Research*. **44**(16), pp.7766–7776.
- Jackson, R.J., Hellen, C.U. and Pestova, T. V 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*. **11**.
- Jaffar-Bandjee, M.C., Das, T., Hoarau, J.J., Krejbich Trotot, P., Denizot, M., Ribera, A., Roques, P. and Gasque, P. 2009. Chikungunya virus takes centre stage in virally induced arthritis: possible cellular and molecular mechanisms to pathogenesis. *Microbes and Infection*. **11**(14–15), pp.1206–1218.
- Jones, J.E., Long, K.M., Whitmore, A.C., Sanders, W., Thurlow, L.R., Brown, J.A., Morrison, C.R., Vincent, H., Peck, K.M., Browning, C., Moorman, N., Lim, J.K. and Heise, M.T. 2017. Disruption of the Opal Stop Codon Attenuates Chikungunya Virus-Induced Arthritis and Pathology. *mBio*. **8**(6), pp.e01456-17.
- Jose, J., Snyder, J.E. and Kuhn, R.J. 2009. A structural and functional perspective of alphavirus replication and assembly. *Future Microbiol*. **4**(7), pp.837–856.
- Jose, J., Taylor, A.B. and Kuhn, R.J. 2017. Spatial and Temporal Analysis of Alphavirus Replication and Assembly in Mammalian and Mosquito Cells. *mBio*. **8**(1), pp.e02294-16.
- Jungreis, I., Chan, C.S., Waterhouse, R.M., Fields, G., Lin, M.F. and Kellis, M. 2016. Evolutionary Dynamics of Abundant Stop Codon Readthrough. *Molecular Biology and Evolution*. **33**(12), pp.3108–3132.
- Jungreis, I., Lin, M.F., Spokony, R., Chan, C.S., Negre, N., Victorsen, A., White, K.P. and Kellis, M. 2011. Evidence of abundant stop codon readthrough in Drosophila and other metazoa. *Genome Research*. **21**(12), pp.2096–2113.
- 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.
- Katoh, K., Rozewicki, J. and Yamada, K.D. 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*.
- Keeling, K.M., Xue, X., Gunn, G. and Bedwell, D.M. 2014. Therapeutics Based on Stop Codon Readthrough. *Annual Review of Genomics and Human Genetics*. **15**(1), pp.371–394.
- Kendra, J.A., Advani, V.M., Chen, B., Briggs, J.W., Zhu, J., Bress, H.J., Pathy, S.M. and

- Dinman, J.D. 2018. Functional and structural characterization of the chikungunya virus translational recoding signals. *The Journal of biological chemistry*. **293**(45), pp.17536–17545.
- Kim, D.Y., Reynaud, J.M., Rasalouslykaya, A., Akhrymuk, I., Mobley, J.A., Frolov, I. and Frolova, E.I. 2016. New World and Old World Alphaviruses Have Evolved to Exploit Different Components of Stress Granules, FXR and G3BP Proteins, for Assembly of Viral Replication Complexes M. T. Heise, ed. *PLOS Pathogens*. **12**(8), p.e1005810.
- Kryuchkova, P., Grishin, A., Eliseev, B., Karyagina, A., Frolova, L. and Alkalaeva, E. 2013. Two-step model of stop codon recognition by eukaryotic release factor eRF1. *Nucleic acids research*. **41**(8), pp.4573–86.
- Kumar, S., Mamidi, P., Kumar, A., Basantray, I., Bramha, U., Dixit, A., Maiti, P.K., Singh, S., Suryawanshi, A.R., Chattopadhyay, Subhasis and Chattopadhyay, Soma 2015. Development of novel antibodies against non-structural proteins nsP1, nsP3 and nsP4 of chikungunya virus: potential use in basic research. *Archives of Virology*. **160**(11), pp.2749–2761.
- Laakkonen, P., Ahola, T. and Kaariainen, L. 1996. The Effects of Palmitoylation on Membrane Association of Semliki Forest Virus RNA Capping Enzyme. *Journal of Biological Chemistry*. **271**(45), pp.28567–28571.
- Labadie, K., Larcher, T., Joubert, C., Mannioui, A., Delache, B., Brochard, P., Guigand, L., Dubreil, L., Lebon, P., Verrier, B., de Lamballerie, X., Suhrbier, A., Cherel, Y., Le Grand, R. and Roques, P. 2010. Chikungunya disease in nonhuman primates involves long-term viral persistence in macrophages. *The Journal of Clinical Investigation*. **120**(3), pp.894–906.
- Lambert, N., Strebel, P., Orenstein, W., Icenogle, J. and Poland, G.A. 2015. Rubella. *Lancet (London, England)*. **385**(9984), pp.2297–307.
- Lanciotti, R.S., Ludwig, M.L., Rwaguma, E.B., Lutwama, J.J., Kram, T.M., Karabatsos, N., Cropp, B.C. and Miller, B.R. 1998. Emergence of Epidemic O'nyong-nyong Fever in Uganda after a 35-Year Absence: Genetic Characterization of the Virus. *Virology*. **252**(1), pp.258–268.
- Lasko, P. 2000. The drosophila melanogaster genome: translation factors and RNA binding proteins. *The Journal of cell biology*. **150**(2), pp.F51-6.
- LaStarza, M.W., Lemm, J.A. and Rice, C.M. 1994. Genetic analysis of the nsP3 region of Sindbis virus: evidence for roles in minus-strand and subgenomic RNA synthesis. *Journal of Virology*. **68**(9), pp.5781–5791.
- Leung, J.Y.-S., Ng, M.M.-L. and Chu, J.J.H. 2011. Replication of Alphaviruses: A Review on the Entry Process of Alphaviruses into Cells. *Advances in Virology*. **2011**, pp.1–9.
- Levis, R., Schlesinger, S. and Huang, H. V 1990. Promoter for Sindbis virus RNA-dependent subgenomic RNA transcription. *Journal of virology*. **64**(4), pp.1726–1733.
- Li, G. and Rice, C.M. 1993. The signal for translational readthrough of a UGA codon in Sindbis virus RNA involves a single cytidine residue immediately downstream of the termination codon. *J Virol*. **67**(8), pp.5062–5067.

- Li, G.P. and Rice, C.M. 1989. Mutagenesis of the in-frame opal termination codon preceding nsP4 of Sindbis virus: studies of translational readthrough and its effect on virus replication. *J Virol.* **63**(3), pp.1326–1337.
- Li, Y.-G., Siripanyaphinyo, U., Tumkosit, U., Noranate, N., A-nuegoonpipat, A., Tao, R., Kurosu, T., Ikuta, K., Takeda, N. and Anantapreecha, S. 2013. Chikungunya Virus Induces a More Moderate Cytopathic Effect in Mosquito Cells than in Mammalian Cells. *Intervirology.* **56**(1), pp.6–12.
- Loughran, G., Howard, M.T., Firth, A.E. and Atkins, J.F. 2017. Avoidance of reporter assay distortions from fused dual reporters. *RNA (New York, N.Y.).* **23**(8), pp.1285–1289.
- Low, J.T. and Weeks, K.M. 2010. SHAPE-directed RNA secondary structure prediction. *Methods.* **52**(2), pp.150–158.
- Lukavsky, P.J. 2009. Structure and function of HCV IRES domains. *Virus Research.* **139**(2), pp.166–171.
- Lum, F.-M. and Ng, L.F.P. 2015. Cellular and molecular mechanisms of chikungunya pathogenesis. *Antiviral Research.* **120**, pp.165–174.
- Malet, H., Coutard, B., Jamal, S., Dutartre, H., Papageorgiou, N., Neuvonen, M., Ahola, T., Forrester, N., Gould, E.A., Lafitte, D., Ferron, F., Lescar, J., Gorbalenya, A.E., de Lamballerie, X. and Canard, B. 2009. The Crystal Structures of Chikungunya and Venezuelan Equine Encephalitis Virus nsP3 Macro Domains Define a Conserved Adenosine Binding Pocket. *J Virol.* **83**(13), pp.6534–6545.
- Manuvakhova, M., Keeling, K. and Bedwell, D.M. 2000. Aminoglycoside antibiotics mediate context-dependent suppression of termination codons in a mammalian translation system. *RNA (New York, N.Y.).* **6**(7), pp.1044–55.
- Marygold, S.J., Attrill, H. and Lasko, P. 2017. The translation factors of *Drosophila melanogaster*. *Fly.* **11**(1), pp.65–74.
- McPherson, R.L., Abraham, R., Sreekumar, E., Ong, S.-E., Cheng, S.-J., Baxter, V.K., Kistemaker, H.A. V, Filippov, D. V, Griffin, D.E. and Leung, A.K.L. 2017. ADP-ribosylhydrolase activity of Chikungunya virus macrodomain is critical for virus replication and virulence. *Proceedings of the National Academy of Sciences of the United States of America.* **114**(7), pp.1666–1671.
- Mehta, R., Gerardin, P., de Brito, C.A.A., Soares, C.N., Ferreira, M.L.B. and Solomon, T. 2018. The neurological complications of chikungunya virus: A systematic review. *Reviews in medical virology.* **28**(3), p.e1978.
- Melton, J. V, Ewart, G.D., Weir, R.C., Board, P.G., Lee, E. and Gage, P.W. 2002. Alphavirus 6K Proteins Form Ion Channels. *Journal of Biological Chemistry.* **277**(49), pp.46923–46931.
- Mortimer, S.A., Kidwell, M.A. and Doudna, J.A. 2014. Insights into RNA structure and function from genome-wide studies. *Nature Reviews Genetics.* **15**(7), pp.469–479.
- Mounce, B.C., Cesaro, T., Vlajnić, L., Vidiņa, A., Vallet, T., Weger-Lucarelli, J., Passoni, G., Stapleford, K.A., Levraud, J.-P. and Vignuzzi, M. 2017. Chikungunya Virus Overcomes Polyamine Depletion by Mutation of nsP1 and the Opal Stop Codon To Confer Enhanced Replication and Fitness. *Journal of Virology.* **91**(15), pp.e00344-

17.

- Mudiganti, U., Hernandez, R. and Brown, D.T. 2010. Insect response to alphavirus infection—Establishment of alphavirus persistence in insect cells involves inhibition of viral polyprotein cleavage. *Virus Research*. **150**(1), pp.73–84.
- Müller, M., Slivinski, N., Todd, E.J.A.A., Khalid, H., Li, R., Karwatka, M., Merits, A., Mankouri, J. and Tuplin, A. 2019. Chikungunya virus requires cellular chloride channels for efficient genome replication. *PLoS Neglected Tropical Diseases*. **13**(9), p.e0007703.
- Mutso, M., Morro, A.M., Smedberg, C., Kasvandik, S., Aquilimeba, M., Teppor, M., Tarve, L., Lulla, A., Lulla, V., Saul, S., Thaa, B., McInerney, G.M., Merits, A. and Varjak, M. 2018. Mutation of CD2AP and SH3KBP1 Binding Motif in Alphavirus nsP3 Hypervariable Domain Results in Attenuated Virus. *Viruses*. **10**(5).
- Myles, K.M., Kelly, C.L.H., Ledermann, J.P. and Powers, A.M. 2006. Effects of an opal termination codon preceding the nsP4 gene sequence in the O’Nyong-Nyong virus genome on *Anopheles gambiae* infectivity. *Journal of virology*. **80**(10), pp.4992–7.
- Myles, K.M., Morazzani, E.M. and Adelman, Z.N. 2009. Origins of alphavirus-derived small RNAs in mosquitoes. *RNA Biology*. **6**(4), pp.387–391.
- Napthine, S., Yek, C., Powell, M.L., Brown, T.D.K. and Brierley, I. 2012. Characterization of the stop codon readthrough signal of Colorado tick fever virus segment 9 RNA. *RNA*. **18**(2), pp.241–252.
- Naresh Kumar, C.V.M. and Sai Gopal, D.V.R. 2010. Reemergence of Chikungunya virus in Indian Subcontinent. *Indian journal of virology: an official organ of Indian Virological Society*. **21**(1), pp.8–17.
- NCBI Resource Coordinators, N.R. 2018. Database resources of the National Center for Biotechnology Information. *Nucleic acids research*. **46**(D1), pp.D8–D13.
- Niesters, H.G. and Strauss, J.H. 1990. Mutagenesis of the conserved 51-nucleotide region of Sindbis virus. *J Virol*. **64**(4), pp.1639–1647.
- Parashar, D. and Cherian, S. 2014. Antiviral perspectives for chikungunya virus. *Biomed Res Int*. **2014**, p.631642.
- Pardigon, N. 2009. The biology of chikungunya: A brief review of what we still do not know. *Pathologie Biologie*. **57**(2), pp.127–132.
- Parker, F., White, K., Phillips, S. and Peckham, M. 2016. Promoting differentiation of cultured myoblasts using biomimetic surfaces that present alpha-laminin-2 peptides. *Cytotechnology*. **68**(5), pp.2159–2169.
- Pisarev, A. V, Hellen, C.U.T. and Pestova, T. V 2007. Recycling of eukaryotic posttermination ribosomal complexes. *Cell*. **131**(2), pp.286–99.
- Pohjala, L., Utt, A., Varjak, M., Lulla, A., Merits, A., Ahola, T. and Tammela, P. 2011. Inhibitors of Alphavirus Entry and Replication Identified with a Stable Chikungunya Replicon Cell Line and Virus-Based Assays. *PLoS ONE*. **6**(12), p.e28923.
- Powers, A.M. and Logue, C.H. 2007. Changing patterns of chikungunya virus: re-emergence of a zoonotic arbovirus. *Journal of General Virology*. **88**(9), pp.2363–



2377.

- Remenyi, R., Li, R. and Harris, M. 2019. On-demand Labeling of SNAP-tagged Viral Protein for Pulse-Chase Imaging, Quench-Pulse-Chase Imaging, and Nanoscopy-based Inspection of Cell Lysates. *BIO-PROTOCOL*. **9**(4), p.e3177.
- Remenyi, R., Roberts, G.C., Zothner, C., Merits, A. and Harris, M. 2017. SNAP-tagged Chikungunya Virus Replicons Improve Visualisation of Non-Structural Protein 3 by Fluorescence Microscopy. *Scientific Reports*. **7**(1), p.5682.
- Reuter, J.S. and Mathews, D.H. 2010. *RNAstructure: software for RNA secondary structure prediction and analysis* [Online]. [Accessed 24 April 2019]. Available from: <http://www.biomedcentral.com/1471-2105/11/129>.
- Roberts, G.C., Zothner, C., Remenyi, R., Merits, A., Stonehouse, N.J. and Harris, M. 2017. Evaluation of a range of mammalian and mosquito cell lines for use in Chikungunya virus research. *Scientific Reports*. **7**(1), p.14641.
- Ross-Thriepland, D. and Harris, M. 2014. Insights into the Complexity and Functionality of Hepatitis C Virus NS5A Phosphorylation. *J Virol*. **88**(3), pp.1421–1432.
- Rozanov, M.N., Koonin, E. V. and Gorbalenya, A.E. 1992. Conservation of the putative methyltransferase domain: a hallmark of the 'Sindbis-like' supergroup of positive-strand RNA viruses. *Journal of General Virology*. **73**(8), pp.2129–2134.
- Rubach, J.K., Wasik, B.R., Rupp, J.C., Kuhn, R.J., Hardy, R.W. and Smith, J.L. 2009. Characterization of purified Sindbis virus nsP4 RNA-dependent RNA polymerase activity in vitro. *Virology*. **384**(1), pp.201–208.
- Rupp, J.C., Sokoloski, K.J., Gebhart, N.N. and Hardy, R.W. 2015. Alphavirus RNA synthesis and non-structural protein functions. *Journal of General Virology*. **96**(9), pp.2483–2500.
- Russo, A.T., White, M.A. and Watowich, S.J. 2006. The Crystal Structure of the Venezuelan Equine Encephalitis Alphavirus nsP2 Protease. *Structure*. **14**(9), pp.1449–1458.
- Sanz, M.A., Rejas, M.T. and Carrasco, L. 2003. Individual Expression of Sindbis Virus Glycoproteins. E1 Alone Promotes Cell Fusion. *Virology*. **305**(2), pp.463–472.
- Schlesinger, S. and Dubensky, T.W. 1999. Alphavirus vectors for gene expression and vaccines. *Current Opinion in Biotechnology*. **10**(5), pp.434–439.
- Scholte, F.E.M., Tas, A., Albulescu, I.C., Žusinaite, E., Merits, A., Snijder, E.J. and Van Hemert, M.J. 2015. Stress granule components G3BP1 and G3BP2 play a proviral role early in chikungunya virus replication. *J Virol*. **89**(8), pp.4457–4469.
- Schuller, A.P. and Green, R. 2018. Roadblocks and resolutions in eukaryotic translation. *Nature Reviews Molecular Cell Biology*. **19**(8), pp.526–541.
- Schulte, T., Liu, L., Panas, M.D., Thaa, B., Dickson, N., Götte, B., Achour, A. and McInerney, G.M. 2016. Combined structural, biochemical and cellular evidence demonstrates that both FGDF motifs in alphavirus nsP3 are required for efficient replication. *Open biology*. **6**(7), pp.160078-.
- Schwartz, O. and Albert, M.L. 2010. Biology and pathogenesis of chikungunya virus.

*Nature Reviews Microbiology*. **8**(7), pp.491–500.

- Schweighardt, B., Shieh, J.T. and Atwood, W.J. 2001. CD4/CXCR4-independent infection of human astrocytes by a T-tropic strain of HIV-1. *Journal of NeuroVirology*. **7**(2), pp.155–162.
- Sharma, R., Kesari, P., Kumar, P. and Tomar, S. 2018. Structure-function insights into chikungunya virus capsid protein: Small molecules targeting capsid hydrophobic pocket. *Virology*. **515**, pp.223–234.
- Sherf, B.A., Navarro, S.L., Hannah, R.R. and Wood, K. V 1996. *Dual-Luciferase TM Reporter Assay: An Advanced Co-Reporter Technology Integrating Firefly and Renilla Luciferase Assays* [Online]. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.652.7565&rep=rep1&type=pdf>.
- Shin, G., Yost, S.A., Miller, M.T., Elrod, E.J., Grakoui, A. and Marcotrigiano, J. 2012. Structural and functional insights into alphavirus polyprotein processing and pathogenesis. *Proceedings of the National Academy of Sciences*. **109**(41), pp.16534–16539.
- Siddiqui, N. and Sonenberg, N. 2016. Proposing a mechanism of action for ataluren. *Proceedings of the National Academy of Sciences of the United States of America*. **113**(44), pp.12353–12355.
- Singh, S.K. and Unni, S.K. 2011. Chikungunya virus: host pathogen interaction. *Reviews in Medical Virology*. **21**(2), pp.78–88.
- Siu, R.W.C., Fragkoudis, R., Simmonds, P., Donald, C.L., Chase-Topping, M.E., Barry, G., Attarzadeh-Yazdi, G., Rodriguez-Andres, J., Nash, A.A., Merits, A., Fazakerley, J.K. and Kohl, A. 2011. Antiviral RNA interference responses induced by Semliki Forest virus infection of mosquito cells: characterization, origin, and frequency-dependent functions of virus-derived small interfering RNAs. *Journal of virology*. **85**(6), pp.2907–17.
- Skuzeski, J.M., Nichols, L.M., Gesteland, R.F. and Atkins, J.F. 1991. The signal for a leaky UAG stop codon in several plant viruses includes the two downstream codons. *Journal of Molecular Biology*. **218**(2), pp.365–373.
- Snyder, A.J. and Mukhopadhyay, S. 2012. The Alphavirus E3 Glycoprotein Functions in a Clade-Specific Manner. *J Virol*. **86**(24), pp.13609–13620.
- Sokoloski, K.J., Hayes, C.A., Dunn, M.P., Balke, J.L., Hardy, R.W. and Mukhopadhyay, S. 2012. Sindbis Virus Infectivity Improves During the Course of Infection in Both Mammalian and Mosquito Cells. *Virus Research*. **167**(1), p.26.
- Solignat, M., Gay, B., Higgs, S., Briant, L. and Devaux, C. 2009. Replication cycle of chikungunya: a re-emerging arbovirus. *Virology*. **393**(2), pp.183–197.
- Song, H., Zhao, Z., Chai, Y., Jin, X., Li, C., Yuan, F., Liu, S., Gao, Z., Wang, H., Song, J., Vazquez, L., Zhang, Y., Tan, S., Morel, C.M., Yan, J., Shi, Y., Qi, J., Gao, F. and Gao, G.F. 2019. Molecular Basis of Arthritogenic Alphavirus Receptor MXRA8 Binding to Chikungunya Virus Envelope Protein. *Cell*. **177**(7), pp.1714-1724.e12.
- Spuul, P., Salonen, A., Merits, A., Jokitalo, E., Kääriäinen, L. and Ahola, T. 2007. Role of

- the amphipathic peptide of Semliki forest virus replicase protein nsP1 in membrane association and virus replication. *Journal of virology*. **81**(2), pp.872–83.
- Strauss, J.H. and Strauss, E.G. 1994. The alphaviruses: gene expression, replication, and evolution. *Microbiological Reviews*. **58**(3), pp.491–562.
- Suthar, M.S., Shabman, R., Madric, K., Lambeth, C. and Heise, M.T. 2005. Identification of Adult Mouse Neurovirulence Determinants of the Sindbis Virus Strain AR86. *Journal of Virology*. **79**(7), pp.4219–4228.
- Tanabe, I.S.B., Tanabe, E.L.L., Santos, E.C., Martins, W. V., Araújo, I.M.T.C., Cavalcante, M.C.A., Lima, A.R. V., Câmara, N.O.S., Anderson, L., Yunusov, D. and Bassi, Ê.J. 2018. Cellular and Molecular Immune Response to Chikungunya Virus Infection. *Frontiers in Cellular and Infection Microbiology*. **8**, p.345.
- Teo, T.-H., Lum, F.-M., Claser, C., Lulla, V., Lulla, A., Merits, A., Rénia, L. and Ng, L.F.P. 2013. A pathogenic role for CD4+ T cells during Chikungunya virus infection in mice. *Journal of immunology (Baltimore, Md. : 1950)*. **190**(1), pp.259–69.
- Tomar, S., Hardy, R.W., Smith, J.L. and Kuhn, R.J. 2006. Catalytic Core of Alphavirus Nonstructural Protein nsP4 Possesses Terminal Adenylyltransferase Activity. *Journal of Virology*. **80**(20), pp.9962–9969.
- Tossavainen, H., Aitio, O., Hellman, M., Saksela, K. and Permi, P. 2016. Structural Basis of the High Affinity Interaction between the Alphavirus Nonstructural Protein-3 (nsP3) and the SH3 Domain of Amphiphysin-2. *The Journal of biological chemistry*. **291**(31), pp.16307–17.
- Tsetsarkin, K., Higgs, S., McGee, C.E., Lamballerie, X. De, Charrel, R.N. and Vanlandingham, D.L. 2006. Infectious Clones of Chikungunya Virus (La Réunion Isolate) for Vector Competence Studies. *Vector-Borne and Zoonotic Diseases*. **6**(4), pp.325–337.
- Tsetsarkin, K.A., Vanlandingham, D.L., McGee, C.E. and Higgs, S. 2007. A Single Mutation in Chikungunya Virus Affects Vector Specificity and Epidemic Potential. *PLoS Pathogens*. **3**(12), p.e201.
- Tuittila, M. and Hinkkanen, A.E. 2003. Amino acid mutations in the replicase protein nsP3 of Semliki Forest virus cumulatively affect neurovirulence. *Journal of General Virology*. **84**(6), pp.1525–1533.
- Utt, A., Das, P.K., Varjak, M., Lulla, V., Lulla, A. and Merits, A. 2015. Mutations Conferring a Noncytotoxic Phenotype on Chikungunya Virus Replicons Compromise Enzymatic Properties of Nonstructural Protein 2. *J Virol*. **89**(6), pp.3145–3162.
- Varjak, M., Žusinaite, E. and Merits, A. 2010. Novel Functions of the Alphavirus Nonstructural Protein nsP3 C-Terminal Region. *Journal of Virology*. **84**(5), pp.2352–2364.
- Vasiljeva, L., Merits, A., Auvinen, P. and Kääriäinen, L. 2000. Identification of a Novel Function of the Alphavirus Capping Apparatus: RNA 5'-TRIPHOSPHATASE ACTIVITY OF Nsp2. *Journal of Biological Chemistry*. **275**(23), pp.17281–17287.
- Villa, N., Do, A., Hershey, J.W.B. and Fraser, C.S. 2013. Human eukaryotic initiation factor 4G (eIF4G) protein binds to eIF3c, -d, and -e to promote mRNA recruitment to the

ribosome. *Journal of Biological Chemistry*. **288**(46), pp.32932–32940.

- Voss, J.E., Vaney, M.-C., Duquerroy, S., Vonrhein, C., Girard-Blanc, C., Crublet, E., Thompson, A., Bricogne, G. and Rey, F.A. 2010. Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature*. **468**(7324), pp.709–712.
- Wang, Y.F., Sawicki, S.G. and Sawicki, D.L. 1994. Alphavirus nsP3 functions to form replication complexes transcribing negative-strand RNA. *Journal of virology*. **68**(10), pp.6466–75.
- Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J. 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. **25**(9), pp.1189–1191.
- Weaver, S.C. 2014. Arrival of Chikungunya Virus in the New World: Prospects for Spread and Impact on Public Health. *PLoS Neglected Tropical Diseases*. **8**(6), p.e2921.
- Weaver, S.C. and Forrester, N.L. 2015. Chikungunya: Evolutionary history and recent epidemic spread. *Antiviral Research*. **120**, pp.32–39.
- Weger-Lucarelli, J., Aliota, M.T., Wlodarchak, N., Kamlangdee, A., Swanson, R. and Osorio, J.E. 2016. Dissecting the Role of E2 Protein Domains in Alphavirus Pathogenicity. *Journal of Virology*. **90**(5), pp.2418–2433.
- Welch, E.M., Barton, E.R., Zhuo, J., Tomizawa, Y., Friesen, W.J., Trifillis, P., Paushkin, S., Patel, M., Trotta, C.R., Hwang, S., Wilde, R.G., Karp, G., Takasugi, J., Chen, G., Jones, S., Ren, H., Moon, Y.-C., Corson, D., Turpoff, A.A., Campbell, J.A., Conn, M.M., Khan, A., Almstead, N.G., Hedrick, J., Mollin, A., Risher, N., Weetall, M., Yeh, S., Branstrom, A.A., Colacino, J.M., Babiak, J., Ju, W.D., Hirawat, S., Northcutt, V.J., Miller, L.L., Spatrick, P., He, F., Kawana, M., Feng, H., Jacobson, A., Peltz, S.W. and Sweeney, H.L. 2007. PTC124 targets genetic disorders caused by nonsense mutations. *Nature*. **447**(7140), pp.87–91.
- Wielgosz, M.M., Raju, R. and Huang, H. V 2001. Sequence Requirements for Sindbis Virus Subgenomic mRNA Promoter Function in Cultured Cells. *JOURNAL OF VIROLOGY*. **75**(8), pp.3509–3519.
- Wills, N.M., Gesteland, R.F. and Atkins, J.F. 1991. Evidence that a downstream pseudoknot is required for translational read-through of the Moloney murine leukemia virus gag stop codon. *Proceedings of the National Academy of Sciences of the United States of America*. **88**(16), pp.6991–5.
- Yaffe, D. and Saxel, O. 1977. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. *Nature*. **270**(5639), pp.725–7.
- Yang, P., Sampson, H.M. and Krause, H.M. 2006. A modified tandem affinity purification strategy identifies cofactors of the Drosophila nuclear receptor dHNF4. *Proteomics*. **6**(3), pp.927–935.
- Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*. **31**(13), pp.3406–3415.
- Zuker, M. and Jacobson, A.B. 1998. Using reliability information to annotate RNA secondary structures. *RNA*. **4**(6), pp.669–679.

## Appendix

**Table 1: List of primers used during this project.**

Primer Name	Direction	Sequence (5' to 3')	Function
Rev_New_Q5_Opal_6588	Reverse	TCCGTGTCTGAGCACGTG	Q5 mutagenesis of Dluc SGR CHIKV for opal, ochre and amber mutant
For_Q5_Opal_6589	Forward	CGACGAGTTATGACTAGACAGG	Q5 mutagenesis of Dluc SGR CHIKV for opal amber mutant
For_Q5_Ochre_6589	Forward	CGACGAGTTATAACTAGACAGGCAG	Q5 mutagenesis of Dluc SGR CHIKV for ochre mutant
For_Q5_Amber_6589	Forward	CGACGAGTTATAGCTAGACAGGCAGGTG	Q5 mutagenesis of Dluc SGR CHIKV for amber mutant
For_Seq_opal_CHIKV_6487	Forward	AATCGAAAGCTTGTCTTCTGAGCTACTAAC	Forward sequencing primer for the opal, ochre and amber mutant
FP_Fluc_Spel_5119	Forward	CGTTGATGGCGAGATACTGCCCGTC	Forward sequencing primer for the nsP3 HVD insertion site
Rev_Agel_CHIKV_Opal_6765	Reverse	AGTAATAGTTGTCCTTTGCTTCATCC	Reverse sequencing primer for the nsP3 HVD insertion site
FP_CHIKV_Capsid_7742	Forward	AGAAGCCACGCAGGAATCGGAAG	Forward sequencing primer for the ICRES CHIKV capsid protein
RP_CHIKV_8224	Reverse	CGAAGATCGGTCTGCCGCTG	Reverse sequencing primer for the ICRES CHIKV capsid protein
RP_CHIKV_E3_8392	Reverse	ACGTGGTGTGGCCAACAGG	Reverse sequencing primer for the ICRES CHIKV E3 protein
RP_CHIKV_8738	Reverse	GTCCAATCGTGGCTGTCATC	Reverse sequencing primer for the ICRES CHIKV E2 protein
RP_CHIKV_9253	Reverse	ACTGCCACTTTTTGTGATTG	Reverse sequencing primer for the ICRES CHIKV E2 protein

RP_CHIKV_9769	Reverse	GGACGGTAGCTCCTGGTGTC	Reverse sequencing primer for the ICRES CHIKV E2 protein
FP_FL_Mut_CHIKV_1172	Forward	ATAGTGGTTAACGGCAGAACGCAAC	Forward sequencing primer for NCT CHIKV stable replicon mutation FL
RP_FL_Mut_CHIKV_1328	Reverse	AGCAGGTCAGTGTCTTTCTCTGAC	Reverse sequencing primer for NCT CHIKV stable replicon mutation FL
FP_5A_Mut_CHIKV_3545	Forward	GGAAAGAATGGAATGGCTGGTTAACAAG	Forward sequencing primer for NCT CHIKV stable replicon mutation 5A
RP_5A_Mut_CHIKV_3713	Reverse	GCAGACCCAACTCTAGGTTGTATGTG	Reverse sequencing primer for NCT CHIKV stable replicon mutation 5A
FP_P718G_Mut_3774	Forward	ACCATTACCAACAGTGCCTCGAC	Forward sequencing primer for NCT CHIKV stable replicon mutation P718G
RP_P718G_Mut_CHIKV_3930	Reverse	AACTTGCGTCCCAATACGCAGATGAC	Reverse sequencing primer for NCT CHIKV stable replicon mutation P718G
FP_IL_Mut_CHIKV_4534	Forward	GGAGAAGAAAATATCTGAGGCCATAC	Forward sequencing primer for NCT CHIKV stable replicon mutation IL
RP_IL_CHIKV_4691	Reverse	GATATGAGTACAGTGCCTTCC	Reverse sequencing primer for NCT CHIKV stable replicon mutation IL
RP_SHP23_CHIKV_5835 (FAM or HEX)	Reverse	GCACTCTCCTGGAGTTTCTTAAG	SHAPE primer (1)
RP_SHP24_CHIKV_5855 (FAM or HEX)	Reverse	ACCTGCTTCTGTTGGCCATGGATG	SHAPE primer (2)
RP_SHP23_CHIKV_5835 (FAM or HEX)	Reverse	GACTGATACCTGCTTCTGTTGGCC	SHAPE primer (3)
RSDM_AGA_CHIKV_5659	Reverse	CCAGCCCTGTCTAGTCTTAACTC	Q5 mutagenesis of Arg ICRES CHIKV RNA structure stem (side 1)
RSDM_UGA_CHIKV_5659	Reverse	CCAGCCCTGTCTAGTCATAACTC	Q5 mutagenesis of opal ICRES CHIKV RNA structure stem (side 1)

RSDM_UAA_CHIKV_5659	Reverse	<b>CCAGCCCTGTCTAGTTATAACTC</b>	Q5 mutagenesis of ochre ICRES CHIKV RNA structure stem (side 1)
RSDM_UAG_CHIKV_5659	Reverse	<b>CCAGCCCTGTCTAGCTATAACTC</b>	Q5 mutagenesis of amber ICRES CHIKV RNA structure stem (side 1)
FSDM_Dis_CHIKV_5669 (UCC mutant)	Forward	<b>CGGCTATATATTCTCGTCGGAC</b>	Q5 mutagenesis of above ICRES CHIKV RNA structure stem (side 1) to create UCC disrupted mutant
FSDM_Com_CHIKV_5669 (UGA mutant)	Forward	<b>GGGATATATATTCTCGTCGGAC</b>	Q5 mutagenesis of above ICRES CHIKV RNA structure stem (side 1) to create UGA disrupted mutant
RSDM_ComS2_CHIKV_5773 (Stem side 2 UCA)	Reverse	<b>GGATAACACTTCTCCTCGTGGAC</b>	Q5 mutagenesis of UGA disrupt ICRES CHIKV RNA structure stem (side 2) to compensate the second side
FSDM_ComS2_CHIKV_5783 (Stem side 2 UCA)	Forward	<b>CCCAAAGCTGGATGAAGCAAAG</b>	Q5 mutagenesis of UGA disrupt ICRES CHIKV RNA structure stem (side 2) to compensate the second side

**Table 2: List of NMIA reactivities of the 37°C folded RNA SHAPE reactions.**

37°C folded RNA Nucleotide position	Base	NMIA reactivity				Averaged NMIA reactivity
		1	2	3	4	
5571	A	1.05	1.18	1.17	0.11	0.88
5572	A	1.14	1.49	1.54	0.66	1.21
5573	G	0.33	0.73	0.51	1.50	0.77
5574	A	0.29	0.83	0.17	0.03	0.33
5575	C	0.65	0.70	0.68	0.67	0.68
5576	U	1.05	1.13	1.10	0.00	0.82
5577	A	0.96	0.49	1.09	0.07	0.65
5578	G	0.08	0.61	0.49	0.49	0.42
5579	A	0.34	0.07	0.00	0.40	0.20
5580	C	0.00	0.03	0.21	0.23	0.12
5581	A	0.22	0.59	0.55	0.66	0.51
5582	G	0.68	0.36	0.48	0.05	0.39
5583	G	0.31	0.14	0.27	0.00	0.18
5584	G	0.29	0.25	0.32	0.26	0.28
5585	C	0.00	0.00	0.00	0.00	0.00
5586	A	0.20	0.00	0.12	0.42	0.19
5587	G	0.09	0.00	0.03	0.04	0.04
5588	G	0.01	0.00	0.01	0.00	0.01
5589	U	0.00	0.00	0.00	0.00	0.00
5590	G	0.22	0.00	0.00	0.03	0.06
5591	G	0.05	0.00	0.07	0.08	0.05
5592	G	0.00	0.00	0.00	0.20	0.05
5593	U	0.07	0.00	0.06	0.61	0.19
5594	A	0.61	0.57	0.71	0.32	0.55
5595	U	0.17	0.40	0.50	1.21	0.57
5596	A	1.03	0.89	1.07	0.66	0.91
5597	U	0.32	0.47	0.57	1.17	0.63
5598	A	0.96	0.81	0.75	0.47	0.75
5599	U	0.39	0.08	0.35	0.38	0.30
5600	U	0.20	0.23	0.22	0.00	0.16
5601	C	0.00	0.00	0.03	0.00	0.01
5602	U	0.00	0.00	0.12	0.00	0.03
5603	C	0.00	0.00	0.00	0.33	0.08
5604	G	0.23	0.00	0.13	0.02	0.10
5605	U	0.17	0.45	0.64	0.28	0.39
5606	C	0.45	0.24	0.22	0.20	0.28
5607	G	0.31	0.29	0.00	0.21	0.20
5608	G	0.27	0.07	0.11	0.00	0.11
5609	A	0.10	0.00	0.02	0.22	0.09
5610	C	0.18	0.21	0.13	0.33	0.21
5611	A	0.06	0.00	0.15	0.00	0.05



5612	C	0.13	0.00	0.04	0.00	0.04
5613	C	0.50	0.17	0.02	0.11	0.20
5614	G	1.16	0.81	0.77	0.78	0.88
5615	G	0.32	0.27	1.23	0.14	0.49
5616	U	0.06	0.00	0.34	0.00	0.10
5617	C	0.00	0.00	0.01	0.00	0.00
5618	C	0.29	0.00	0.00	0.31	0.15
5619	A	0.37	0.09	0.18	0.50	0.29
5620	G	0.03	0.15	0.29	0.00	0.12
5621	G	0.01	0.00	0.03	0.02	0.02
5622	U	0.12	0.00	0.10	0.00	0.06
5623	C	0.36	0.00	0.21	0.44	0.25
5624	A	0.33	0.08	0.46	0.48	0.34
5625	U	0.63	0.01	0.57	0.58	0.45
5626	U	0.81	0.41	0.67	0.63	0.63
5627	U	0.74	0.55	0.53	0.68	0.63
5628	A	0.23	0.64	0.00	0.00	0.22
5629	C	0.61	0.00	0.08	0.55	0.31
5630	A	0.81	0.66	0.50	0.77	0.69
5631	A	0.34	1.09	0.76	0.00	0.55
5632	C	0.00	0.16	0.33	0.26	0.19
5633	A	2.18	0.47	0.96	1.06	1.17
5634	G	1.04	1.51	1.29	1.54	1.35
5635	A	1.18	1.81	1.26	1.31	1.39
5636	A	0.05	0.11	1.63	1.66	0.86
5637	G	0.11	0.42	0.43	0.35	0.33
5638	U	0.00	0.00	0.38	0.30	0.17
5639	C	0.00	0.00	0.00	0.00	0.00
5640	A	0.13	0.29	0.00	0.11	0.13
5641	G	0.12	0.28	0.00	0.19	0.15
5642	U	0.20	0.62	0.11	0.42	0.34
5643	A	0.35	0.00	0.45	0.67	0.37
5644	C	0.22	0.37	0.19	0.21	0.25
5645	G	0.04	0.00	0.45	0.31	0.20
5646	C	0.17	0.10	0.14	0.05	0.12
5647	C	0.00	0.00	0.03	0.00	0.01
5648	A	0.49	0.64	0.53	0.39	0.51
5649	G	0.48	0.00	0.82	0.66	0.49
5650	U	0.46	0.82	0.81	0.50	0.65
5651	C	0.11	0.26	0.26	0.17	0.20
5652	A	0.76	1.19	0.83	0.89	0.92
5653	G	0.94	0.96	0.68	1.15	0.93
5654	U	0.54	0.56	0.42	0.73	0.56
5655	G	0.10	0.25	0.17	0.32	0.21
5656	C	0.01	0.02	0.04	0.00	0.02
5657	U	0.02	0.01	0.02	0.06	0.03

5658	G	0.00	0.18	0.00	0.04	0.06
5659	C	0.00	0.29	0.06	0.00	0.09
5660	C	0.40	0.25	0.05	0.00	0.18
5661	G	0.11	0.60	0.51	0.38	0.40
5662	G	0.00	0.94	0.17	0.40	0.38
5663	U	0.38	0.48	0.66	0.72	0.56
5664	G	1.40	0.83	0.68	0.66	0.89
5665	A	0.37	0.02	0.38	0.34	0.28
5666	A	0.51	0.54	0.08	0.04	0.29
5667	C	0.00	0.00	0.00	0.00	0.00
5668	A	0.00	0.00	0.21	0.18	0.10
5669	C	0.18	0.05	0.00	0.05	0.07
5670	C	0.00	0.00	0.14	0.00	0.04
5671	C	0.00	0.00	0.00	0.00	0.00
5672	U	0.01	0.00	0.00	0.00	0.00
5673	G	0.00	0.00	0.03	0.07	0.03
5674	G	0.00	0.00	0.00	0.00	0.00
5675	A	0.07	0.00	0.00	0.07	0.04
5676	G	0.24	0.00	0.06	0.14	0.11
5677	G	0.30	0.51	0.20	0.62	0.41
5678	A	0.05	1.38	0.77	0.63	0.71
5679	A	0.65	1.24	0.88	0.19	0.74
5680	G	0.20	0.24	0.09	0.00	0.13
5681	U	0.00	0.15	0.03	0.00	0.05
5682	C	0.03	0.00	0.01	0.00	0.01
5683	C	0.00	0.00	0.00	0.00	0.00
5684	A	0.11	0.12	0.17	0.28	0.17
5685	C	0.00	0.00	0.07	0.03	0.03
5686	G	0.09	0.00	0.03	0.00	0.03
5687	A	0.10	0.01	0.00	0.26	0.09
5688	G	0.12	0.00	0.00	2.74	0.72
5689	G	0.00	0.00	0.00	0.20	0.05
5690	A	0.00	0.10	0.13	0.27	0.13
5691	G	0.13	0.87	0.31	0.26	0.39
5692	A	0.72	0.78	0.51	0.56	0.64
5693	A	0.79	0.49	0.24	1.05	0.64
5694	G	0.90	0.70	1.61	0.77	1.00
5695	U	1.06	0.32	0.81	0.45	0.66
5696	G	0.63	0.57	0.65	0.39	0.56
5697	U	0.36	0.25	0.20	0.84	0.41
5698	U	0.42	0.00	0.15	0.24	0.20
5699	A	0.00	0.00	0.00	0.00	0.00
5700	C	0.00	0.00	0.00	0.00	0.00
5701	C	0.00	0.00	0.00	0.04	0.01
5702	C	0.00	0.01	0.00	0.07	0.02
5703	A	0.00	0.00	0.00	0.08	0.02

5704	C	0.00	0.00	0.07	0.12	0.05
5705	C	0.00	0.07	0.02	0.04	0.03
5706	U	0.15	0.00	0.42	0.50	0.27
5707	A	0.62	0.21	0.94	0.83	0.65
5708	A	0.71	0.37	0.85	0.82	0.69
5709	G	0.35	0.00	0.45	0.36	0.29
5710	C	0.07	0.00	0.08	0.15	0.08
5711	U	0.16	0.00	0.07	0.04	0.07
5712	G	0.12	0.00	0.00	0.00	0.03
5713	G	0.19	0.00	0.05	0.13	0.09
5714	A	0.34	0.11	0.09	0.28	0.21
5715	U	0.40	0.22	0.39	0.43	0.36
5716	G	0.24	0.03	0.57	0.63	0.37
5717	A	0.06	0.00	0.58	0.53	0.29
5718	A	0.04		0.47	0.42	0.31
5719	G	0.29		0.09	0.06	0.15
5720	C			0.00	0.51	0.26
5721	A			0.39	0.00	0.20

**Table 3: List of NMIA reactivities of the 28°C folded RNA SHAPE reactions.**

28°C folded RNA		NMIA reactivity		
Nucleotide position	Base	1	2	Averaged NMIA reactivity
5571	A	0		0.73
5572	A	0.86		1.03
5573	G	0	0.27	0.68
5574	A	0.35	0.39	0.00
5575	C	0	0	0.12
5576	U	0	0.1	0.02
5577	A	0	0.49	0.45
5578	G	0.26	0	0.69
5579	A	0	0.05	0.34
5580	C	0	0.09	0.13
5581	A	0	0	0.15
5582	G	0.34	0	0.68
5583	G	0	0	0.76
5584	G	0	0	0.97
5585	C	0	0	1.57
5586	A	0	0	1.10
5587	G	0.01	0	0.44
5588	G	0	0	0.97
5589	U	0	0	1.25
5590	G	0	0	0.99
5591	G	0	0.9	0.76
5592	G	0.33	0.71	0.72
5593	U	0.97	1.45	0.25
5594	A	0.48	0.8	0.23
5595	U	1.67	1.1	0.03
5596	A	0.8	0.77	0.03
5597	U	1.24	0.6	0.10
5598	A	0.58	0	0.16
5599	U	0.58	0.07	0.11
5600	U	0	0	0.07
5601	C	0	0.21	0.35
5602	U	0	0	0.30
5603	C	0.42	0.46	0.42
5604	G	0	0	0.51
5605	U	0.35	0.21	0.52
5606	C	0.08	0.15	0.14
5607	G	0.23	0	0.04
5608	G	0	0.13	0.00
5609	A	0	0	0.00
5610	C	0.06	0.06	0.22
5611	A	0	0.2	1.20

5612	C	0.04	0.11	0.00
5613	C	0	0.13	0.34
5614	G	0	0.1	0.99
5615	G	0.52	0.25	0.30
5616	U	0.19	0	0.03
5617	C	0	0	0.00
5618	C	0	0.46	0.15
5619	A	0.29	0.59	0.23
5620	G	0.43	0	0.09
5621	G	0.02	0	0.01
5622	U	0	0.05	0.06
5623	C	0	0.29	0.18
5624	A	0.64	0.49	0.21
5625	U	0.66	0.51	0.32
5626	U	0.75	0.41	0.61
5627	U	0.73	0.77	0.65
5628	A	0	0.18	0.44
5629	C	0.48	0	0.31
5630	A	0.85	0.4	0.74
5631	A	0	0.37	0.72
5632	C	0.08	0.11	0.08
5633	A	0.92	0.69	1.33
5634	G	0.76	0.38	1.28
5635	A	1.19	1	1.50
5636	A	1.75	1.3	0.08
5637	G	0.09	0	0.27
5638	U	0.14	0.02	0.00
5639	C	0	0	0.00
5640	A	0.03	0.46	0.21
5641	G	0	0.27	0.20
5642	U	0.43	0.59	0.41
5643	A	0.42	0.6	0.18
5644	C	0.18	0.26	0.30
5645	G	0.4	0	0.02
5646	C	0.04	0	0.14
5647	C	0	0	0.00
5648	A	0.27	0.07	0.57
5649	G	0	0	0.24
5650	U	0.3	0.18	0.64
5651	C	0	0.05	0.19
5652	A	0.6	0.72	0.98
5653	G	1.19	1.27	0.95
5654	U	0.51	0.69	0.55
5655	G	0.13	0.17	0.18
5656	C	0	0.05	0.02
5657	U	0	0	0.02









































5658	G	0	0	0.09
5659	C	0	0	0.15
5660	C	0	0	0.33
5661	G	0.18	0	0.36
5662	G	0.5	0.07	0.47
5663	U	0	0.21	0.43
5664	G	1.08	0.5	1.12
5665	A	0.51	0.33	0.20
5666	A	0.31	0.03	0.53
5667	C	0	0.17	0.00
5668	A	0.16	0.07	0.00
5669	C	0.01	0	0.12
5670	C	0	0	0.00
5671	C	0	0	0.00
5672	U	0	0	0.01
5673	G	0.16	0	0.00
5674	G	0.06	0	0.00
5675	A	0.03	0	0.04
5676	G	0.1	0	0.12
5677	G	0.12	0.39	0.41
5678	A	0.49	0.42	0.72
5679	A	0.77	0	0.95
5680	G	0.17	0.01	0.22
5681	U	0	0	0.08
5682	C	0	0	0.02
5683	C	0	0.28	0.00
5684	A	0.2	0	0.12
5685	C	0.02	0	0.00
5686	G	0.04	0	0.05
5687	A	0.25	0	0.06
5688	G	2.88	0	0.06
5689	G	0.23	0	0.00
5690	A	0.2	0	0.05
5691	G	0.11	0.23	0.50
5692	A	0.6	0.45	0.75
5693	A	2.07	0.45	0.64
5694	G	0	0.39	0.80
5695	U	0.37	0.58	0.69
5696	G	0.4	0.52	0.60
5697	U	0.6	0.76	0.31
5698	U	0.53	0.4	0.21
5699	A	0	0.08	0.00
5700	C	0	0.05	0.00
5701	C	0.19	0	0.00
5702	C	0.2	0.21	0.01
5703	A	0	0.16	0.00

5704	C	0	0.08	0.00
5705	C	0.43	0.07	0.04
5706	U	0.89	0.51	0.08
5707	A	0.91	0.82	0.42
5708	A	0.45	0.87	0.54
5709	G	0.11	0.42	0.18
5710	C	0.01	0.15	0.04
5711	U	0.26	0.19	0.08
5712	G	0.4	0.06	0.06
5713	G	0.22	0.18	0.10
5714	A	0.33	0.34	0.23
5715	U	0.56	0.25	0.31
5716	G	0.53	0.42	0.14
5717	A	0.41	0.82	0.03
5718	A	0.02	0.72	0.04
5719	G	0.49	0.51	0.29
5720	C	0.12	0.36	0.00
5721	A	0.67	0.79	0.00

**Table 4: Mfold P-num values.**

This table depicts the colour scheme for the P-num values predicted by the Mfold software when predicting RNA structures. This table was obtained from the Mfold software.

**Probability annotation is not (yet) available on the mfold web server.**

Color	Hex	% P-num % ss-count	Probability	Color	Hex	% P-num % ss-count	Probability
	ff0000	0.0-2.5	0.999		00ffff	50.0-52.5	0.500
	ff1f00	2.5-5.0	0.998		00bfff	52.5-55.0	0.366
	ff3f00	5.0-7.5	0.997		007fff	55.0-57.5	0.269
	ff5f00	7.5-10.0	0.997		003fff	57.5-60.0	0.197
	ff7f00	10.0-12.5	0.995		0000ff	60.0-62.5	0.144
	ff9f00	12.5-15.0	0.994		1f00ff	62.5-65.0	0.106
	ffbf00	15.0-17.5	0.991		3f00ff	65.0-67.5	0.077
	ffd000	17.5-20.0	0.988		5f00ff	67.5-70.0	0.057
	ffff00	20.0-22.5	0.984		7f00ff	70.0-72.5	0.042
	dfff00	22.5-25.0	0.978		9f00ff	72.5-75.0	0.031
	bfff00	25.0-27.5	0.969		bf00ff	75.0-77.5	0.022
	9fff00	27.5-30.0	0.958		df00ff	77.5-80.0	0.016
	7fff00	30.0-32.5	0.943		af00cf	80.0-82.5	0.012
	5fff00	32.5-35.0	0.923		7f009f	82.5-85.0	0.009
	3fff00	35.0-37.5	0.894		5f007f	85.0-87.5	0.006
	1fff00	37.5-40.0	0.856		3f005f	87.5-90.0	0.005
	00ff00	40.0-42.5	0.803		1f003f	90.0-92.5	0.003
	00ff3f	42.5-45.0	0.731		09001f	92.5-95.0	0.003
	00ff7f	45.0-47.5	0.634		040009	95.0-97.5	0.002
	00ffbf	47.5-50.0	0.500		000000	97.5-100.0	0.001



**Table 5: List of primary antibodies.**

Primary antibodies	Supplier	Dilution
Rabbit anti nsP3 (Whole protein)	Gift from Prof Andres Merits	1:1000
Rabbit anti nsP1	Gift from Prof Andres Merits	1:1000
Monoclonal Mouse anti Anti- $\beta$ -Actin	SIGMA (A1978- 200UL)	1:10,000
Monoclonal Mouse anti GAPDH (6C5)	AB-CAM (ab8245)	1:10,000
Monoclonal Mouse anti-FLAG® M2	SIGMA (F2165- 2MG)	1:200
StrepMAB-Classic Monoclonal mouse anti-strep-tag	IBA (2-1507-001)	1:1000